Evidence Synthesis

Number 238

Screening for Osteoporosis to Prevent Fractures: An Evidence Review for the U.S. Preventive Services Task Force

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
https://www.ahrq.gov

Contract No. HHSA-75Q80120D00007, Task Order No. 75Q80122F32003

Prepared by:
RTI International–University of North Carolina at Chapel Hill Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Leila C. Kahwati, MD, MPH
Christine E. Kistler, MD, MASc
Graham Booth, BSPH
Nila Sathe, MA, MLIS
Rachel Gordon, MD
Alexia Cox, BA
Ebiere Okah, MD, MS
Roberta C. Wines, MPH
Meera Viswanathan, PhD

AHRQ Publication No. 23-05312-EF-1
January 2025

This report is based on research conducted by the RTI International–University of North Carolina Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHSA-75Q80120D00007, Task Order No. 75Q80122F32003). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators has any affiliation or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Howard Tracer, MD, AHRQ Medical Officer and Brandy Peaker, MD, MPH, former AHRQ Medical Officer; Tracy Wolff, MD, MPH, AHRQ Scientific Director for the USPSTF; current and former members of the U.S. Preventive Services Task Force; expert peer reviewers Carolyn Crandall, MD, MS, MACP, Kristine Ensrud, MD, MPH, Guylene Theriault, MD, CCFP, and one reviewer who wishes to remain anonymous. In addition, we thank several federal partner reviewers including two from the Centers for Disease Control and Prevention National Center for Injury Prevention and Control Division of Injury Prevention, one from the National Institute on Aging, and three from the National Institute of Arthritis and Musculoskeletal and Skin Diseases. We also thank Megan Van Noord, MSIS, librarian; Sharon Barrell, MA, and Mary Gendron, BA, editors; Teyonna Downing and Michelle Bogus, publications specialists; and Kayla Giger, research assistant. The authors also acknowledge the contributions of Corrine V. Evans, MPP, and Eric S. Johnson, PhD, to the Prediction model study Risk Of Bias Assessment Tool (PROBAST) ratings for the Fracture Risk Assessment Tool (FRAX®) and QFracture Development Cohorts that appear in Appendix G of this report.

Suggested Citation

Kahwati LC, Kistler CE, Booth G, Sathe N, Gordon R, Cox A, Okah E, Wines RC, Viswanathan M. Screening for Osteoporosis to Prevent Fractures: An Evidence Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 238. AHRQ Publication No. 23-05312-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2025.

Structured Abstract

Purpose: To review the evidence on screening for osteoporosis to prevent fractures in community-dwelling adults in primary care settings.

Data Sources: MEDLINE, Embase, the Cochrane Library, and trial registries through January 9, 2024; bibliographies from retrieved articles, outside experts, and surveillance of the literature through July 31, 2024.

Study Selection: Two reviewers independently selected English-language studies. We included trials or systematic reviews (SRs) that evaluated the benefits or harms of screening for osteoporosis or fracture risk in adults without known osteoporosis or medical conditions associated with bone metabolism compared with no screening or usual care and that reported fracture, mortality, or harm outcomes. We included studies or recent SRs that reported on the accuracy of risk assessment instruments or bone mineral density (BMD) for predicting fracture or the diagnostic accuracy of risk assessment instruments for identifying osteoporosis. We included randomized, controlled trials (RCTs) that reported on U.S. Food and Drug Administration (FDA)-approved bisphosphonates or denosumab for the treatment of osteoporosis among participants without secondary osteoporosis or prior fragility fracture. Except for studies of predictive accuracy, we excluded studies with poor methodological quality.

Data Extraction: One reviewer extracted data and a second checked accuracy. Two reviewers independently rated methodological quality for all included studies using predefined criteria. When more than one similar study was available, we conducted meta-analyses.

Data Synthesis: We included 145 studies (in 195 publications). Three RCTs and three SRs reported on the direct benefits of screening in European women (median ages, 71 to 76 years); one of the trials and one of the SRs also reported on the direct harms of screening. Two-staged screening interventions were used by two trials, which included a Fracture Risk Assessment Tool [FRAX®] risk estimate followed by BMD testing if the estimated risk was above a specified threshold; the third trial, which required participants to have at least one clinical risk factor, performed BMD testing, vertebral fracture assessment, falls risk assessment, and laboratory measures related to bone metabolism. Across trials, screening was associated with a reduced risk of hip fractures (pooled relative risk [RR], 0.83 [95% confidence interval {CI}, 0.73 to 0.93]; 3 RCTs; 42,009 participants) and major osteoporotic fractures (MOFs) (pooled RR, 0.94 [95% CI, 0.88 to 0.99]; 3 RCTs; 42,009 participants) compared with usual care. The absolute risk differences corresponding to these estimates are 5 (hip) to 6 (MOF) fewer fractures per 1,000 participants screened. One RCT reported no difference in anxiety between screened and unscreened participants. One SR estimated the risk for overdiagnosis as between 11.8 and 24.1 percent.

For predicting fracture, six SRs and 30 unique cohorts reported on the accuracy of 11 risk assessment instruments, and 22 unique cohorts reported on the accuracy of BMD alone. Calibration outcomes were limited. For risk assessment instruments, discrimination as measured by area under the curve (AUC) ranged from 0.52 to 0.93 and varied by instrument, inclusion of BMD as an input, and fracture type. The AUC of BMD alone for predicting MOF or hip fracture ranged from 0.60 to 0.86. Forty-three unique cohorts reported on the diagnostic accuracy of risk assessment instruments for identifying osteoporosis. In women, AUCs ranged from 0.32 to 0.87 across 11 instruments. In men, AUCs ranged from 0.62 to 0.94 across 12 instruments. Five studies reported information relevant to screening intervals that suggested no additional predictive accuracy for repeat BMD testing at an interval of 4 to 8 years.

Twenty-seven RCTs reported on the benefits of treatment, and 40 RCTs and three cohort studies reported on the harms of treatment. Compared with placebo, bisphosphonates (pooled RR, 0.67 [95% CI, 0.45 to 1.00]; 6 RCTs; 12,055 participants) and denosumab (RR, 0.60 [95% CI, 0.37 to 0.97] from the largest RCT of 7,808 participants) were associated with a reduction in hip fractures; these drugs were also associated with reductions in vertebral fractures and nonvertebral fractures. The absolute risk difference across fracture types and medications ranged from 3 fewer to 44 fewer per 1,000 participants treated compared with placebo. For mortality, the pooled RR for bisphosphonates was 0.71 (95% CI, 0.49 to 1.05; 6 RCTs; 3,714 participants) and the pooled RR for denosumab was 0.79 (95% CI, 0.58 to 1.07; 5 RCTs; 8,828 participants). Compared with placebo, no statistically significant associations were observed for discontinuation due to adverse events, serious adverse events, or gastrointestinal adverse events (pooled RRs ranging from 0.97 to 2.18).

Limitations: Direct evidence for BMD screening alone was not available. Direct evidence was available for screening in older European women that included country-specific fracture risk estimations, but this evidence was limited by modest adherence in intervention groups and contamination in control groups. Limited direct evidence for harms was identified. Predictive and diagnostic accuracy were limited by heterogeneity in populations evaluated, analytic methods used, and insufficient reporting of calibration. For treatment, populations exclusively comprising persons with prior fragility fracture or secondary osteoporosis or in long-term care were not included. Only FDA-approved bisphosphonates for prevention or treatment of osteoporosis and denosumab were included, and comparative effectiveness and harms were not addressed. Few studies of treatment in men were eligible. Treatment studies enrolled persons with osteoporosis based on BMD rather than fracture risk, and sample sizes and treatment durations may not have been adequate for the detection of rare harms such as osteonecrosis of the jaw and atypical femur fractures.

Conclusions: Screening in older, higher-risk women was associated with a small absolute risk reduction in hip and MOF fractures compared with usual care. Screening strategies varied and no direct evidence evaluated screening using dual-energy X-ray absorptiometry alone or screening in women younger than age 65 years or in men. Risk assessment instruments, BMD at the hip or spine has poor to modest discrimination in men and older women for predicting fracture and studies of calibration were limited. For identifying osteoporosis, risk assessment instruments had modest to good accuracy in men and modest accuracy in older women. In women younger than age 65 years, risk assessment instruments had poor predictive (fracture) and diagnostic (osteoporosis) discrimination. Treatment of osteoporosis with FDA-approved bisphosphonates or denosumab was associated with reductions in vertebral, nonvertebral, and hip fractures with no increase in discontinuations due to adverse events or serious adverse events compared with placebo in studies conducted over one to several years' duration; however, data about rare and longer-term harms were limited from the evidence included in this update.

Table of Contents

Figures

Tables

Appendixes

Chapter 1. Introduction

Scope and Purpose

This report will be used by the U.S. Preventive Services Task Force (USPSTF) to update its 2018 recommendations for screening for osteoporosis to prevent fractures.[1] The USPSTF recommended screening for osteoporosis with bone measurement testing to prevent fractures in women age 65 years or older (B recommendation). For postmenopausal women younger than age 65 years, the USPSTF recommended screening with bone measurement testing for those at increased risk of osteoporosis, as determined by a formal clinical risk assessment tool (B recommendation). For men, the evidence was insufficient to assess the balance of benefits and harms (I statement). These recommendations and statements were consistent with the prior recommendation from 2011;[2] the primary difference was that for postmenopausal women younger than age 65 years, the 2018 recommendation updated the threshold to consider bone mineral density (BMD) testing based on fracture risk assessment. The USPSTF suggested that postmenopausal women younger than age 65 years with at least the 10-year risk of major osteoporotic fracture (MOF) from the Fracture Risk Assessment Tool (FRAX®) for a 65-year-old White woman of average weight (73.9 kg) and height (160.3 cm) based on National Health and Nutrition Examination Survey (NHANES) 2011–2014 data (MOF risk of 8.4%) could be used to identify younger women for bone measurement testing.[1, 3]

Condition Definition

Osteoporosis is a disorder of the skeletal system and is characterized by decreased bone mass, microarchitectural deterioration of bone tissue, and a consequent increase in bone fragility and risk of fractures.[4] The ability to measure bone density (related to bone mass) using dual-energy X-ray absorptiometry (DXA) in grams/centimeter^2, also referred to as areal BMD, was available in routine clinical practice by the 1990s. However, differences in DXA machines made by different manufacturers led to widely varying absolute BMD results (in grams/centimeter^2) for a single individual depending on the machine used. This variation led to the use of relative measures to express BMD results, specifically T-scores, to account for variation across DXA machines. In 1994, the World Health Organization (WHO) operationalized the definition of osteoporosis in postmenopausal White women as bone density at the hip or spine that is 2.5 standard deviations (SDs) or lower (T-score ≤-2.5) than the mean BMD measured at the femoral neck (FN) for a reference population of young healthy White women. WHO chose this threshold because the lifetime risk of osteoporotic fracture in women was at least 30 to 40 percent and a T-score of -2.5 (acknowledged by WHO as somewhat arbitrary) would categorize approximately 30 percent of women as having osteoporosis. At the time this threshold was selected, it was not known whether the 30 percent of women identified based on T-score would be the same women who would eventually have a fracture.[5, 6] We now know that although there is some overlap in these populations, they are not the same.

Soon after the WHO definition, DXA machine manufacturers reached consensus on using a specific reference population for FN and total hip (TH) BMD measurements that is still used today. This reference population is White women ages 20 to 29 years from NHANES III (1988–

1994).[7] After the implementation of T-scores to report BMD for women, BMD for men was still being reported in reference to a young male population.[8] However, because males have a higher average BMD than females, the same absolute BMD measurement in grams/centimeter^2 for a male would result in a lower T-score in reference to a young male population than in reference to a young female population.[8] Because fracture risk for males and females is similar at the same absolute BMD (in grams/centimeter^2),[9] the use of sex-specific reference populations for generating T-scores results in more osteoporosis diagnoses and treatment among males compared with females with the same absolute BMD.[8] The sex differences in BMD do not appear to be explained by nutrition, level of activity, weight, or lean mass but may be explained by bone size.[10]

The use of country- or race-specific reference populations to calculate T-scores also leads to different T-scores for the same absolute BMD. To ensure that the same absolute BMD result in grams/centimeter^2 generates the same T-score worldwide, it is necessary for all DXA manufacturers to use the same reference population for all persons (without regard to sex, race, or country of origin). Thus, the International Society for Clinical Densitometry recommended using the Caucasian (non-race-adjusted) young female NHANES III reference standard for calculating FN and TH BMD T-scores for both males and females and for all racial and ethnic groups.[11] Because lumbar spine (LS) BMD was not included in NHANES III data, DXA machines use their own reference data for reporting T-scores at the LS. These are referred to as "local reference populations" and vary by manufacturer.

Osteoporosis and low bone mass (T-score between -1.0 and -2.5, formerly referred to as osteopenia) are asymptomatic risk factors for fragility fractures (also known as "low-energy" or "low-trauma" fractures), which are fractures sustained from a fall from standing height or lower that would not cause a fracture in most healthy persons.[12] Although low-trauma hip and vertebral fractures are usually considered to be fragility fractures, low-trauma fractures at other skeletal sites often depend on the fall circumstances, and there is debate as to whether such fractures should be considered fragility fractures. For example, higher physical activity is associated with an increased risk for wrist fracture but lower risk of proximal humerus fractures. Bone density is one of many risk factors for fragility fractures, and persons with a BMD in the osteoporotic range have a higher relative risk of fragility facture compared with those in the low or normal bone mass range. But the majority of fragility fractures actually occur in persons with low or normal bone mass because these categories of BMD include many more people compared with the category of persons with osteoporosis.[13-16] As a result, some experts have suggested a revision to the operational definition of osteoporosis.[17] Many consider a personal history of a fragility fracture as pathognomonic for osteoporosis, regardless of T-score. The U.K. National Institute for Health and Care Excellence noted that although osteoporosis is defined by a T-score of -2.5 or below on a DXA scan, the diagnosis may be assumed in women age 75 years or older if the responsible clinician considers a DXA scan to be clinically inappropriate or infeasible.[18] The National Bone Health Alliance has proposed that in addition to a T-score of less than or equal to -2.5 at the spine or hip, the identification of a hip fracture; vertebral, proximal humerus, pelvis, or some wrist fractures in persons with low bone mass; or fracture risk assessment scores above prespecified thresholds should confer an osteoporosis diagnosis.[19, 20]

Prevalence and Burden of Disease

An analysis of NHANES data from 2017 to 2018 suggests an age-adjusted prevalence of osteoporosis of 12.6 percent among the noninstitutionalized U.S. population age 50 years or older; the prevalence was higher in women (19.6%) compared with men (4.4%).[21] Prevalence is higher among persons age 65 years or older (Women 27.1%, Men 5.7%) compared with persons ages 50 to 64 years (Women 13.1%, Men 3.3%).[21] Prevalence also varied by race and ethnicity: prevalence was 12.9 percent in non-Hispanic White persons, 18.4 percent in non-Hispanic Asian persons, and 14.7 percent in Hispanic persons; these differences were not statistically significant.[22] The prevalence in non-Hispanic Black persons was 6.8 percent and was significantly different from other racial and ethnic groups.[22] The prevalence of osteoporosis or low bone mass is 51.5 percent in women and 33.5 percent in men.[21]

The most worrisome concern resulting from osteoporosis is a fragility fracture, which can lead to significant morbidity and mortality.[23] These fractures are associated with an increase in excess mortality,[24] risk of subsequent fractures,[25-27] loss of independence,[28, 29] reduced ability to perform activities of daily living,[28, 29] and psychological consequences.[29] Mortality associated with a hip fracture is highest in the first few months immediately after the fracture.[30, 31] Although osteoporosis and fragility fractures are more common in women than men,[32] excess mortality is more common in men.[32-34] Among Medicare beneficiaries in 2016, 40 percent with a new osteoporotic fracture were hospitalized within a week of fracture, and among those with hip fracture, 90 percent were hospitalized.[35] One review found that only between 40 and 60 percent of persons experiencing a hip fracture recovered their prefracture level of mobility and ability to perform instrumental activities of daily living, while only 40 to 70 percent gained their level of independence for basic activities of daily living.[28] The burden associated with hip fractures is more commonly reported than the burden associated with vertebral or other fractures, leading to a concern that the burden from vertebral fractures and other fractures may be underestimated.[23, 36] However, despite excess mortality associated with fractures, trials of fracture prevention have not clearly demonstrated a reduction in mortality.

Based on Medicare fee-for-service and Medicare Advantage data, the number of beneficiaries who experienced a new osteoporotic fracture was 1.8 million in 2016.[35] **Appendix A Table 1** depicts the age-standardized incidence of hip fractures from a cohort of over 1.8 million Medicare Advantage health plan enrollees between 2007 and 2017.[35] Age-standardized incidence rates of fragility fractures decreased between 2007 and 2013.[35, 37] This decline was hypothesized to be because of increasing rates of obesity, increasing use of antiresorptive agents, and birth cohort effects.[38] However, because of the aging of the population, the absolute incidence is increasing. Further, recent studies have suggested that the decline in age-standardized fracture rates may have plateaued in the last 5 to 7 years.[35, 39, 40]

Etiology and Natural History

Fragility fractures can be a consequence of osteoporosis. Although those with osteoporosis have the greatest relative risk of fracture, most fractures occur in those with low bone mass (i.e., T-scores between -1.0 and 2.5) or normal bone density (T-score >-1) because they represent a greater share of the population.[13, 41-45]

Osteoporosis may occur either without a known cause (referred to as primary osteoporosis) or secondary to a medical condition or medications (referred to as secondary osteoporosis).[46] Postmenopausal osteoporosis is considered a type of primary osteoporosis.[46] Secondary osteoporosis is bone loss associated with certain medical conditions: various endocrine conditions of the pituitary, thyroid, parathyroid, or reproductive organs; eating disorders; disorders of the gastrointestinal (GI) or biliary tract; renal disease; bone marrow disorders; and cancer.[46, 47] Secondary osteoporosis can also result after organ transplantation and can arise from chronic use of medications with known deleterious effects on bone mass, such as glucocorticoids, immunosuppressants, antiepileptic medications, heparin, gonadotropin-releasing hormone agonists, and some long-acting progesterone agents used as contraceptives (which may be reversible).[46, 47] The identification and management of secondary osteoporosis is outside of the scope of the USPSTF's current recommendation.

A biological basis for differences in the age of onset and the prevalence of osteoporosis between males and females exists. We note that most of the research in the area of bone metabolism and fractures uses the terms "men" and "women" to refer to biological sex (male and female); we use the terms used by individual study authors in this report, which is typically "men" and "women." Women lose bone mass at a younger age, and the rate of loss is faster than for men.[10] The prevalence of low bone mass in women increases rapidly beginning around age 60 years, and the prevalence of osteoporosis doubles by age 70 years, whereas the prevalence of osteoporosis only doubles by age 80 years for men.[10] Transmen and transwomen who have not undergone any hormonal treatment associated with transitioning likely have the same risks and prevalence as persons assigned female and male sex at birth, respectively.

Data from the Study of Women's Health Across the Nation (SWAN),[48] a multisite longitudinal epidemiologic study in the United States, reported that bone turnover increases about 2 years before the final menstrual period, increases rapidly for the next 4 years with a peak 2 years after the final menstrual period, and subsequently plateaus thereafter. However, the rate of turnover after this plateau is approximately 20 percent higher than premenopausal levels. In SWAN, larger increases in bone turnover were observed for women with body mass index (BMI) less than 25 kg/m^2, and the smallest increases in turnover were observed in women with BMI greater than 30 kg/m^2. Furthermore, higher turnover levels were observed among Japanese Americans, and smaller turnover levels were observed among African Americans, even after adjusting for other variables such as BMI.

Risk Factors

Although bone density is an important risk factor for fragility fractures for both males and females, advancing age is the more critical determinant.[49] Older adults have much higher fracture rates than younger adults with the same BMD because of concurrent increasing risk from declining bone quality and an increasing tendency to fall.[50] **Appendix A Figure 1** demonstrates the impact of age on estimated fracture risk based on the FRAX calibrated to the U.S. population by race (Caucasian, Black, Hispanic, Asian). As seen in this figure, the risk of fracture is higher at age 70 years compared with age 50 years, holding BMD constant for both males and females of all races and ethnicities. Race-neutral estimated fracture risks from FRAX calibrated to the Canadian and U.K. populations are also provided in this figure for comparison.

Bone density may not be as useful a predictor of fracture risk, particularly in younger persons. An Australian case-control study evaluating the relationship between osteoporosis and fragility fractures found that only 20 percent of women ages 50 to 59 years with incident fracture had osteoporosis. In comparison, 45 percent, 60 percent, and 70 percent of those ages 60 to 69 years, 70 to 79 years, and age 80 years or older with incident fractures had osteoporosis.[45] Fractures in younger persons that occur at some sites (e.g., wrist) may be associated with higher physical activity levels and greater risk-taking behaviors, so some experts have suggested they should not be considered fragility fractures.

Aside from medical conditions and medications (e.g., corticosteroids) associated with secondary osteoporosis, additional risk factors include menopausal status in women, previous osteoporotic fracture, low body weight (less than 58 kg [127 lbs]), parental history of hip fracture, cigarette smoking, and excess alcohol consumption.[51, 52] Diabetes treated with insulin (type 1 or type 2) increases the risk of fracture but has a variable relationship with BMD. Type 1 diabetes is associated with a reduction in BMD and an increased risk of fracture. Type 2 diabetes has a variable relationship. Some studies have observed that type 2 diabetes mellitus (DM) is associated with both increased BMD and fracture risk, suggesting BMD may be less useful in predicting fracture risk because bone integrity, not density, may be responsible for fracture in this population.[53] However, two recent large cohort studies suggest negligible contribution of type 2 DM to overall fracture risk. One study among men in the United Kingdom observed no association between type 2 DM and future fracture,[54] while another study among adults in Sweden observed a small increase in relative risk of MOF and in hip fracture for persons with diabetes, but negligible contribution of diabetes to overall fracture risk when all other risks were considered.[55] Further, the association between type 2 DM and fracture risk was absent when competing mortality was considered.

A systematic review (SR) and meta-analysis identified risk factors associated with fragility fractures in men.[56] The review found statistically significant associations between fractures and increasing age, low BMI, excessive alcohol intake (daily intake or greater than 10 servings per week), current smoking, chronic corticosteroid use, history of prior fractures, history of falls within the past year, hypogonadism, history of cerebrovascular accident, and history of diabetes.

Racial differences in both the prevalence of osteoporosis and incidence of osteoporotic fractures are discussed in detail in **Appendix A Contextual Question 3**. Studies reported lower fracture incidence in Asian, Hispanic, and Black populations compared with White populations among both men and women.[57-59] Decreases in BMD are observed with increasing age across all races and ethnicities, but differences in BMD alone are not sufficient to explain racial and ethnic differences in fracture incidence. For example, Asian women have been found to have lower BMD than White women but lower fracture risk.[60-62] Moreover, racial categories are broad, are socially determined, and vary between countries. It is possible that unaccounted for environmental differences between racial and ethnic groups are responsible for differences in fracture incidence or that racial and ethnic differences in fracture incidence may reflect differences in underlying clinical risks in these populations. U.S. racial categories obscure the tremendous diversity that occurs within racial groups.

Rationale for Screening/Screening Strategies

The rationale for screening is to identify persons with osteoporosis or at risk of a fragility fracture and provide treatment to increase bone mass or prevent further losses to minimize the occurrence of fragility fractures and related morbidity.

Bone Measurement Tests

As described earlier, the WHO defines osteoporosis in postmenopausal females and males age 50 years or older as a BMD measurement associated with a T-score of -2.5 or lower obtained through DXA at a central skeletal site (e.g., total hip, FN, or LS). This definition is widely used throughout the world and has remained unchanged for decades. Compared with other imaging modalities, DXA has been correlated to biomechanical bone strength and clinical fracture outcomes and uses low doses of radiation.[63] Further, centrally measured DXA was the test used for diagnosis of osteoporosis among participants enrolled in nearly all trials of bone-conserving pharmacotherapies.[64] Evidence suggests that BMD at any skeletal site can predict fracture risk, but fracture risk at a specific site (e.g., hip or spine) is best predicted by BMD measurement at that site.[65] Further, morbidity of fragility fractures at central sites, particularly the hip, is much higher than morbidity of fragility fractures that occur at other sites.[66-68] For these reasons, and because centrally measured DXA does not require any followup tests to confirm the diagnosis of osteoporosis, it is the test recommended for assessing BMD and is the one that is used most widely.

Other bone measurement tests are available but are not in widespread use for primary screening.[63] These include enhancements to traditional DXA scanning such as vertebral fracture assessment or trabecular bone score, quantitative ultrasound, DXA measured at peripheral sites (e.g., wrist), quantitative computed tomography, and radiograph absorptiometry.[69] However, none of these tests were used to identify participants for randomized, controlled trials (RCTs) of pharmacotherapy for fracture prevention.

Risk Assessment Tools

BMD alone may not be a sensitive enough screening tool for identifying persons at high fracture risk.[69] Some experts recommend a screening approach that assesses for increased fracture risk, rather than identifying osteoporosis, because 1) most fragility fractures occur in persons without osteoporosis, 2) measured bone density only reflects one aspect of bone quality, and 3) nonskeletal factors also contribute to fracture risk.[69] Several risk assessment tools that incorporate age and sex, with or without other risk factors, have been developed to assess the risk for current osteoporosis or to predict the risk for future fragility fracture. **Appendix A Table 2** summarizes tools that were evaluated in the prior review for the USPSTF.[3, 70] These tools were originally developed to either 1) identify osteoporosis or 2) predict fracture risk, but subsequent studies have evaluated the diagnostic or predictive accuracy of many of them with respect to both outcomes. However, some of the risk assessment tools were developed on small cohorts using homogenous populations or have limited published evidence. Three instruments (FRAX,[12, 71] Fracture Risk Calculator [FRC],[72, 73] and the Garvan Fracture Risk Calculator[74, 75]) can be used with or without BMD as a risk factor input. The instruments designed to identify osteoporosis generally require fewer risk inputs than instruments designed to predict future fracture. Additionally, several instruments include risk factors (such as personal history of fragility fracture or medical conditions or medications known to be associated with secondary osteoporosis), suggesting that the population for which these tools were developed includes persons beyond a general primary care screening population.

Some risk assessment instruments incorporate race or ethnicity as a risk factor. These include the Simple Calculated Osteoporosis Risk Estimation (SCORE) to identify a person's current risk for osteoporosis (i.e., T-score <-2.5)[76] and two fracture risk prediction instruments: FRAX, calibrated for use internationally, and QFracture, developed for use in the United Kingdom.[77] Additional information about how race and ethnicity are used in these risk assessment tools is in **Appendix A Contextual Question 2**.

The most studied fracture risk assessment instrument is FRAX; however, its underlying model parameters are not publicly available. It was derived from nine cohorts in Europe, the United States, Japan, and Canada and further validated in an additional 11 cohort studies.[12, 71] Detailed information about FRAX is in **Appendix A Additional Background**. As of release version 4.2, 73 different country-specific versions of FRAX are available that have been calibrated using country-specific fracture incidence and mortality data, which is considered a competing risk in the model.[78] As of 2016, FRAX was incorporated into 120 guidelines worldwide and added into DXA machine software following regulatory approval by the Food and Drug Administration (FDA) and has been incorporated into clinical decision support tools within electronic health record systems.[69] For the United States, four different versions of FRAX are available that have been calibrated based on racial- and ethnic-specific fracture incidence data, including a version for non-Hispanic Caucasians, a version for non-Hispanic Blacks, a version for Hispanics, and a version for non-Hispanic Asians. It is unclear what version of FRAX clinicians should use for persons who are mixed race, of other races, or immigrants from other countries who are now living in the United States.[79] In the wake of recent attention to racial bias in clinical algorithms, some have raised questions regarding the validity of race-specific FRAX calculators, which predict lower rates of fracture for people of color compared with White persons of the same age, BMD, and clinical risk factors.[80, 81] Few countries other than the United States have developed race-specific versions of FRAX, and other countries with similar ethnic diversity as the United States (e.g., United Kingdom, Australia, Canada) use a single version of FRAX with all races and ethnicities.

Interventions

Reducing fracture risk involves addressing underlying modifiable risks through approaches such as smoking cessation, increased physical activity, avoidance of heavy alcohol use, adequate calcium and vitamin D intake, and fall prevention interventions in older persons at increased risk for falls. However, most relevant to the scope of this update is the use of pharmacologic treatment to increase bone mass or prevent further loss of bone mass.

First-line therapy typically includes drugs within the bisphosphonate class. FDA-approved drugs for prevention or treatment include four bisphosphonates (alendronate, zoledronic acid, risedronate, ibandronate), the RANK-ligand inhibitor denosumab, the sclerostin inhibitor romosozumab, recombinant parathyroid hormone agents (teriparatide, abaloparatide), estrogen (with or without progesterone), selective estrogen receptor modulators (raloxifene, bazedoxifene in combination with estrogen), and calcitonin. Although most second-line agents have demonstrated efficacy at reducing loss of bone mass or decreasing fragility fractures, not all have demonstrated efficacy for specifically reducing hip fractures.[82, 83] Off-label treatments (i.e., drugs that do not have an FDA-approved indication for the prevention or treatment of osteoporosis) include testosterone, tamoxifen, and other bisphosphonates (i.e., etidronate, pamidronate, tiludronate). **Appendix A Table 3** provides detailed information related to bisphosphonate drugs with FDA-approved indications and denosumab for the prevention or treatment of osteoporosis in the United States.

For primary prevention of fractures, pharmacotherapy is generally recommended for T-scores of -2.5 or less (osteoporosis). Further, pharmacotherapy may also be warranted based on shared decision making for persons with T-scores between -1.0 and -2.5 (low bone mass) who are at high risk for fracture as determined based on clinical judgment or increasingly based on standardized risk calculators such as FRAX. For primary fracture prevention in the United States, the Bone Health and Osteoporosis Foundation (formerly known as the National Osteoporosis Foundation [NOF]) recommends treatment for individuals with low bone mass who have a 10-year hip fracture risk of at least 3 percent or a 10-year MOF risk of at least 20 percent based on FRAX.[84] The hip fracture risk threshold was selected based on a U.S.-specific economic analysis of cost-effectiveness from a societal perspective sponsored by the NOF and that assumed one-step BMD screening, use of generic bisphosphonates, and a willingness-to-pay threshold of $60,000 per quality-adjusted life-year gained.[80, 85] These treatment thresholds have not been evaluated in trials. The use of absolute fracture risk in addition to BMD increases the number of candidates for pharmacologic therapy in the United States.[86, 87] Some countries have adopted the U.S. thresholds for intervention, while others use age-dependent thresholds or a combination of fixed and age-dependent thresholds.[88] Countries may establish different risk thresholds for initiating treatment based on country-specific epidemiology, competing health priorities, costs, and resource availability.[88] For example, Japan recommends the use of FRAX in persons without a prior fracture with a T-score between -1.8 and -2.7 and recommends treatment for an MOF risk of 15 percent or higher.[89]

Current Clinical Practice

Screening and primary prevention of osteoporosis in asymptomatic adults without known risks for secondary osteoporosis or prior fragility fracture is within the scope of practice for most primary care providers (PCPs). Guidelines developed by various organizations and specialty societies vary widely and provide recommendations based on age, sex, menopausal status, and other characteristics (**Appendix A Table 4**). Many guidelines recommend fracture risk assessment, DXA measurement, or both. Variation with respect to population, timing, and frequency also exists. Some guidelines include recommendations for those with prior fractures or at-risk conditions (e.g., long-term glucocorticoid steroid use), which is beyond the scope of the review for the USPSTF.

In 2023, the Canadian Task Force on Preventive Health Care (CTFPHC) issued updated recommendations for screening to prevent primary fragility fractures.[90] The CTFPHC recommends screening women age 65 years or older with the Canadian FRAX tool to facilitate shared decision making about pharmacotherapy. If pharmacotherapy is a consideration, it then recommends ordering DXA testing to facilitate re-estimation of fracture risk with a BMD input. The CTFPHC recommends against screening in men age 40 years or older and in women younger than age 65 years.

For primary osteoporosis, nearly all guidelines acknowledge that a variety of medications are available and can be effective for treating osteoporosis. Some specifically state that bisphosphonates should be used as first-line therapy. Some also suggest denosumab as initial therapy, particularly for patients who are intolerant of bisphosphonates or because of its proven efficacy for reducing hip fracture.[91] However, as the field has evolved from focusing solely on the treatment of osteoporosis to identification and treatment of high fracture risk, guidelines diverge about when to treat. Some guidelines focus on BMD exclusively when deciding whether to begin treatment, others on predicted fracture risk assessment, often without a specific rationale.

An SR of osteoporotic fracture risk assessment and treatment guidelines identified 120 guidelines[88] recommending the use of FRAX-based fracture risks for conducting DXA testing and considering treatment. Of these, 38 did not provide a rationale for how fracture probabilities derived should be used for decision making. Some guidelines recommend DXA testing or treatment using fixed-probability thresholds (k=58, a group that includes the USPSTF 2011 and 2018 recommendations), while others recommend an age-dependent threshold (k=22) or a combination of the two (k=2). Of the guidelines referencing fixed-probability thresholds for treatment, over half (k=39) reference an absolute fracture risk of 20 percent or greater for MOF as the threshold for treatment in those with low bone mass.

Implementation in Practice

The implementation of screening for osteoporosis in practice is heavily influenced by quality performance measures related to this service. In 2006, the National Committee for Quality Assurance introduced the Healthcare Effectiveness Data and Information Set measure assessing the percentage of women ages 65 to 85 years who report ever having received a bone density test to screen for osteoporosis.[92] The rate of receipt of bone density tests rose in the ensuing decade.[93] In 2006, 64.4 percent of women ages 65 to 85 years in a Medicare health maintenance organization plan and 71.3 percent in a Medicare preferred provider organization reported ever having a bone density test. By 2014, these numbers had risen to 74.2 percent and 78.5 percent, respectively. The Centers for Medicare & Medicaid Services (CMS) Measures Inventory now includes "Screening for Osteoporosis for Women Ages 65–85 Years of Age."[94] Despite these quality measures, a review of the CMS data between 2006 and 2016 found that performance gaps persist in osteoporosis identification and treatment.[92] A study using a sample of U.S. Medicare claims-based data evaluated physician-reported reasons for not providing recommended screening or treatment. In this study, 24 percent of claims documented that care was considered but not provided because of contraindications, other reasons, or patient preference.[95] This suggests that it may be difficult to achieve further improvement on this measure beyond current levels. However, racial differences in screening and treatment exist. Black women are less likely to be screened and treated for osteoporosis than White women.[96, 97] Additional information about differences in receipt of screening and treatment for osteoporosis by race and ethnicity is addressed in **Appendix A Contextual Question 4**.

Although some underuse may exist, some studies have also identified overuse of BMD screening. The Choosing Wisely Campaign, which is endorsed by multiple medical societies, lists bone density testing as a test that should be considered carefully before ordering in women younger than age 65 years and in men younger than age 70 years with no risk factors.[98] The National Physicians Alliance Good Stewardship Working Group defines overuse as DXA screening in women younger than age 65 years or men younger than 70 years with no risk factors.[99] CMS includes a measure to decrease overuse: "Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture (eCQM)."[100] Findings from the National Ambulatory Medical Care Survey indicated that overuse of DXA in primary care accounted for $527 million per year in expenditures.[101] Further, a study in a large regional healthcare system suggested that about one half of women younger than age 65 years without risk factors received DXA screening over a 7-year period.[96]

Poor treatment adherence among those identified with osteoporosis and offered medication potentially limits the beneficial impact of widespread, routine screening. In one study conducted in the United States, nearly 30 percent of persons who were prescribed bisphosphonates filled the prescription, and only half of those who filled the prescription were still taking medication 1 year later.[102] In an analysis of a U.S. commercial insurance database from January 2009 to March 2020, alendronate was the most common medication used for osteoporosis, representing just over 60 percent of prescriptions for bone-directed therapies.[103] Further, denosumab prescriptions increased since 2009 and represented about 20 percent of prescriptions by the end of the study period, outpacing all other medications except alendronate.[103] Over 92 percent of prescriptions were directed to women and 76 percent were to persons older than the age of 65 years.[103]

Chapter 2. Methods

Key Questions and Analytic Framework

The Evidence-based Practice Center (EPC) investigators, USPSTF members, and Agency for Healthcare Research and Quality (AHRQ) Medical Officers developed the scope and key questions (KQs). **Figure 1** shows the analytic framework and KQs that guided the review. Five KQs were developed for this review:

1. Does screening for fracture risk or osteoporosis reduce fractures and fracture-related morbidity and mortality in adults?

2a. What is the predictive accuracy of risk assessment tools for identifying adults who are at increased risk for hip fractures or major osteoporotic fractures?

2b. What is the predictive accuracy of BMD testing with DXA at central skeletal sites for identifying adults who are at increased risk for hip or major osteoporotic fractures?

2c. What is the diagnostic accuracy of risk assessment tools for identifying adults with osteoporosis?

2d. What is the evidence to determine screening intervals, and how do these intervals vary by baseline or current individual fracture risk?

3. What are the harms of screening for fracture risk or osteoporosis?

4. What is the effectiveness of pharmacotherapy with selected FDA-approved medications on fracture incidence and fracture-related morbidity and mortality?

5. What are the harms associated with selected FDA-approved medications?

We also addressed the following contextual questions:

1. What is the evidence from modeling studies about the effectiveness of risk screening strategies that use different ages at which to start and stop screening and different screening intervals?

2. How do various fracture risk assessment tools use race and ethnicity in fracture risk calculations?

3. What is the incidence of fractures among persons of different races and ethnicities in the United States in the last 10 to 15 years, and what factors might explain differences in incidence among different races and ethnicities?

4. What are the differences in rates of screening or treatment initiation among persons of different races and ethnicities, and what might explain these differences?

5. What are the implications of using fixed-fracture risk thresholds for decisions regarding stepwise screening or treatment?

6. What is the evidence for rare harms of bisphosphonate treatment (i.e., osteonecrosis of the jaw, atypical femur fractures) from observational studies that use noneligible control groups or are uncontrolled?

7. What is the evidence for rebound fractures after discontinuation of denosumab?

These contextual questions are not shown in the analytic framework because they were not analyzed using the same systematic methods as the KQs. They were intended to provide additional background or contextual information for interpreting the results of the KQ and were addressed through targeted literature searches to identify the most recent and relevant information to the questions at hand.

Data Sources and Searches

We searched PubMed/MEDLINE, Embase, and the Cochrane Library for English-language articles published through January 9, 2024. We used Medical Subject Headings as search terms and keywords when appropriate to describe relevant populations, tests, interventions, outcomes, and study designs and applied additional limits on the completed search to remove case reports, case series, articles with child in the title, and articles with a type categorized as conference abstract. The complete search strategy for all data sources is detailed in **Appendix B.1**. We conducted targeted searches for unpublished literature by searching ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform. In addition to database searches, we reviewed reference lists of relevant articles, studies suggested by expert reviewers, and comments received during public commenting periods. Since January 2024, we conducted ongoing surveillance through article alerts and targeted searches of journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and the related USPSTF recommendation. The last surveillance was conducted on July 31, 2024.

Study Selection

We developed inclusion and exclusion criteria for populations, interventions, comparators, outcomes, settings, and study designs with input from the USPSTF (**Appendix B.2**). We included good- or fair-quality, English-language studies focused on adults age 40 years or older conducted in countries categorized as *very high* on the Human Development Index.[104] Other criteria were specific to each KQ. For KQ 1 (direct benefits of screening), we included controlled trials of screening vs. no screening or usual care that reported fracture or mortality outcomes among persons not known to have osteoporosis, prior fragility fracture, or medical conditions or medications associated with secondary osteoporosis. Eligible screening strategies comprised risk assessment instruments, DXA measurement of BMD, or both. For KQ 3, we used similar criteria as KQ 1 except that we looked for harms of screening and allowed for controlled cohort studies in addition to trials.

For KQ 2, we included cohort studies or SRs of cohort studies that evaluated the accuracy of risk assessment instruments (KQ 2a) or BMD alone (KQ 2b) for predicting future incident fractures or the diagnostic accuracy of DXA for identifying osteoporosis (KQ 2c). For predictive accuracy, we sought studies reporting calibration or discrimination outcomes and for diagnostic accuracy, we sought studies reporting discrimination outcomes. Calibration measures the extent to which predicted fracture risks are similar to observed risks over time for the population overall and across the spectrum of predicted risks. Discrimination measures the extent to which a risk assessment (KQ 2a) or BMD (KQ 2b) identifies persons who ultimately experience a fracture as higher risk compared with those who do not. For KQ 2c, discrimination measures the extent to which risk assessment instruments identify persons with osteoporosis compared with those without osteoporosis. Discrimination reported with area under the receiver operating characteristics curve (AUC), sensitivity, and specificity were eligible for KQ 2a, 2b, and 2c. For risk assessment instruments, we included only instruments that had been evaluated in at least two independent external cohorts to the development cohort. We allowed risk assessment instruments that had been evaluated in only one external cohort if it was conducted in men because of a more limited pool of evidence for use of such tools in men. We limited primary research studies for KQ 2a and 2b to studies conducted in countries with hip fracture incidence similar to the United States.[105] We did not include any accuracy data from model development cohorts. For KQ 2a and KQ 2b, we included poor-quality studies because of the limited pool of good- or fair-quality predictive accuracy studies. For KQ 2d, we included studies reporting data that would inform conclusions related to screening intervals.

For KQs 4 and 5 (benefits and harms of treatment), we included RCTs or controlled cohort studies (for harms only) that reported on FDA-approved bisphosphonates or denosumab compared with placebo and that reported fracture, mortality, or harm outcomes. The review was limited to these agents for women because this was not a comprehensive review of all treatment options and bisphosphonates and denosumab are the most frequently prescribed agents. We also considered studies of teriparatide, abaloparatide, or romosozumab for men because of the limited pool of treatment studies among men. We excluded studies where the majority of enrolled participants had secondary osteoporosis or prior fragility fractures.

Two team members independently reviewed titles, abstracts, and full-text articles using study selection criteria to determine inclusion in or exclusion from this update. Disagreements were resolved by discussion or review by a third reviewer. We reassessed studies included in the prior 2018 review[3, 70] against the updated study selection criteria for this update. We screened all citations using the DistillerSR platform (DistillerSR, Inc.) and managed citations using EndNote Version 9.2 (Clarivate™).

Data Abstraction and Quality Assessment

One reviewer abstracted relevant information for each included study into a structured form in DistillerSR including design, population, intervention, comparator, outcomes, timing, and setting. A second person reviewed all data abstractions for accuracy. We considered data from the same study population or cohort but reported in separate publications as one study. We contacted study authors to clarify study data when needed.

We assessed the risk of bias for each included study using design-specific risk of bias assessments (RoB 2 for RCTs,[106] ROBINS-I for nonrandomized studies of interventions,[107] QUADAS-2 for diagnostic test accuracy,[108] ROBIS for SRs).[109, 110] For predictive accuracy of risk assessment instruments, we first evaluated the risk of bias of each instrument in its development cohort(s) using the full PROBAST risk of bias instrument adapted to include health equity signaling items.[111, 112] We next evaluated the risk of bias of studies assessing these instruments in external validation cohorts using an adapted version of the PROBAST short form.[110] For all study designs, we translated risk-of-bias ratings from instruments to methodological quality ratings using predefined criteria developed by the USPSTF and adapted for this topic (**Appendix B.3**).[113] Two independent investigators assigned quality ratings for each study. Disagreements were resolved by discussion or review by a third reviewer.

Data Synthesis and Analysis

We synthesized findings for each KQ in tabular and narrative format. To determine whether meta-analyses were appropriate, we assessed the clinical and methodological heterogeneity of the studies following established guidance that includes evaluating the similarities in study population, intervention, dose, and frequency and similarities in timing and specification of outcomes.[114] For KQs 1, 3, 4, and 5, when at least two similar studies were available, quantitative synthesis was conducted with random-effects models using the inverse-variance weighted method (DerSimonian and Laird) to estimate pooled effects.[115] We calculated pooled relative risks (RRs) and 95 percent confidence intervals (CIs) for fracture and mortality outcomes; we then re-expressed the pooled RRs as absolute risk differences (ARDs) per 1,000 persons screened or treated.[116] Statistical significance was assumed when 95 percent CIs of pooled results did not cross the null effect. All testing was two-sided. For all quantitative syntheses, we used the I^2 statistic to assess statistical heterogeneity.[117, 118] An I^2 from 0 to 40 percent might not be important, 30 to 60 percent may represent moderate heterogeneity, 50 to 90 percent may represent substantial heterogeneity, and 75 percent or greater represents considerable heterogeneity.[117, 118] For KQ 4 and KQ 5, data were pooled across dosage groups for studies with more than one active intervention arm and we conducted sensitivity analyses for alternative types of vertebral fractures (clinical vs. vertebral), for drug dosages that were not FDA-approved dosages, and for alternative pooling methods to account for rare or zero events in one or both study arms.[113] We conducted all quantitative analyses using Stata version 17 (StataCorp LLC).

We assessed the overall strength of the body of evidence for each comparison and outcome organized by KQ as high, moderate, low, or insufficient using methods developed for the USPSTF (and the EPC program), based on the quality of studies and limitations, consistency of results between studies, precision of findings, and risk of reporting bias.[113, 119] We also assessed the applicability of the findings to U.S. primary care populations and settings.

Expert Review and Public Comment

A draft research plan for this topic was posted on the USPSTF website for public comment from August 12, 2021, to September 9, 2021. In response, the USPSTF included additional outcomes, added two contextual questions regarding rare but serious harms, and listed special populations of interest. The USPSTF also made several minor additions and wording changes to improve the clarity and specificity of the research approach.

A draft report was reviewed by four content experts, five representatives of federal partners, USPSTF members, and AHRQ Medical Officers and was revised based on comments received. In response to these comments, additional sensitivity analyses were conducted for KQ 1 and KQ 4 and new information related to overdiagnosis was added for KQ 3. Several new studies were added to KQ 2, and results were further stratified by men, women, and younger women. Additional sources of heterogeneity were discussed for KQ 2, and additional limitations were noted for all KQs. In addition, the draft report was posted for public comment from June 11, 2024, to July 8, 2024. Based on public comments, we made revisions to improve the clarity of the report, but no new eligible studies were identified.

USPSTF and AHRQ Involvement

Members of the USPSTF helped develop the scope of work, including the analytic framework and KQs, and review the evidence synthesis. AHRQ staff provided project oversight, conducted reviews of the draft report, and helped facilitate an external review of the evidence synthesis.

Chapter 3. Results

Results of Literature Searches

We included 145 unique studies published in 195 articles for this update review (**Figure 2**). Three RCTs and three SRs (published in 14 articles) reported direct evidence for the benefits of screening (KQ 1).[120-133] One RCT (published in 2 articles)[120, 121] and one SR[131, 132] reported on direct evidence for the harms of screening. Six SRs[131, 134-138] and 30 cohort studies (published in 49 articles)[72, 73, 139-181] reported on the accuracy (discrimination or calibration or both) of various risk assessment instruments for predicting fracture (KQ 2a). Twenty-two unique cohorts (published in 28 articles) reported on the accuracy of BMD for predicting fracture (KQ 2b).[15, 146, 150, 151, 154, 156, 157, 160-162, 166, 172, 175, 176, 179, 182-194] Findings from 43 unique cohorts published in 54 articles reported on the diagnostic accuracy of fracture risk assessment instruments for identifying osteoporosis as defined by a BMD T-score of less than -2.5 (KQ 2c).[141, 143, 159, 195-245] Five studies reported information relevant to the determination of screening intervals (KQ 2d).[246-250] Twenty-seven RCTs (published in 36 articles) reported on the benefits of treatment.[251-286] Lastly, 40 RCTs in 48 articles[251-254, 256-269, 272-280, 284-304] and three controlled cohorts studies[305-307] reported on the harms of treatment. A list of studies for which we reviewed the full-text article but excluded is provided in **Appendix C** along with the reason for exclusion. Note that although studies may have multiple reasons for exclusion, we only recorded one reason.

KQ 1. Does Screening for Fracture Risk or Osteoporosis Reduce Fractures or Fracture-Related Morbidity or Mortality in Adults?

We identified three pragmatic, fair-quality RCTs (published in 10 articles) that evaluated screening compared with no screening in older European women.[120-129] In this section, we provide a summary of the study characteristics and findings from these trials. Detailed study, population, and intervention characteristics are described in **Appendix D Table 1** with additional narrative description in **Appendix E.1** and detailed outcomes in **Appendix D Table 8**. In addition, we identified three good-quality SRs[130-133] that included these three trials. One of these SRs[131, 132] was conducted in support of the CTPHC's 2023 recommendation on screening for primary prevention of fragility fractures.[90] Details about the included SRs are in **Appendix D Table 9** with SR quality ratings in **Appendix D Tables 23–27**.

Study Characteristics

We identified three fair-quality, pragmatic RCTs (**Table 1**): the Risk-stratified Osteoporosis Strategy Evaluation (ROSE) study (N=34,229 randomized population; N=18,605 per protocol 1 analysis population),[126-129] the Screening in the Community to Reduce Fractures in Older Women (SCOOP) study (N=12,483 randomized),[120-123] and the Stichting Artsen Laboratorium en Trombosedienst (SALT) Osteoporosis Study (SOS) (N=11,032 randomized).[124, 125] ROSE and SOS are new to this update. All three RCTs randomized persons to screening vs. usual care (i.e., no systematic screening) and reported clinical fracture outcomes. All three RCTs included older European women (median ages, 71 to 76 years) who we presumed to be predominantly White (exact racial and ethnicity characteristics not reported in two of the three trials). Other inclusion and exclusion criteria varied across studies. Among those enrolled, the mean or median 10-year FRAX-estimated risk of MOF was 19 percent in SCOOP, 20 percent in ROSE, and 24.6 percent in SOS; the 10-year estimated hip fracture risks were 8.5 percent, 6.7 percent, and 11.6 percent, respectively.[120, 124, 126] The proportion of participants with a prior fracture was 12.6 percent in ROSE, 22 percent in SCOOP, and 43 percent in SOS; however, there was significant variability in the definition and reporting of prior fractures between trials.[120, 124, 126]

Two RCTs (SCOOP[120] and ROSE[124]) used a two-step screening intervention consisting of a FRAX risk assessment (without BMD input) on participants randomized to screening, then invited those with a high fracture risk score (≥15% risk for MOF in ROSE; at or above the age-based hip risk threshold in SCOOP) for DXA. In both studies, DXA results and treatment recommendations were shared with the participant and their PCP, who made final decisions about treatment. In contrast, SOS included women already known to have at least one clinical risk factor for osteoporosis and conducted a DXA, vertebral fracture assessment, blood chemistries, falls risk assessment, and FRAX without BMD input on all participants randomized to screening.[124] The FRAX risk was recalculated with BMD information, and results of all tests and recommendations were provided to the participant's PCP. The comparison group in all three studies was routine care as guided by the participant's PCP.

We rated study quality of the three primary research studies as fair largely because of issues related to contamination in the control groups, poor to modest adherence in the intervention groups, and lack of blinding, which was not feasible because of the pragmatic nature of the trials (**Appendix D Tables 18–22**).

In addition to the three primary studies that we included in our analysis, we identified three SRs.[130-133] The SR authored by Merlijn et al[130] included the same three RCTs that we included in our analysis. The SR authored by Gates et al was performed in support of the CTFPHC recommendation and included two additional studies.[131] One of these additional studies[308] was excluded in the previous USPSTF SR on this topic[3] for an ineligible study design as it was a nonconcurrent cohort study. The other additional study was an RCT of population-based screening with DXA compared with usual care in women ages 45 to 54 years (Aberdeen Prospective Osteoporosis Screening Study).[309] Women with DXA results in the lowest quartile of the first 1,000 persons screened were considered "high risk," and their results were shared with their PCPs, who were advised to offer hormone replacement therapy when the woman reached menopause assuming no contraindications.[309] This study was excluded in the previous USPSTF SR on this topic[3] for poor quality. Further, the treatment intervention used in this study is no longer standard practice in the United States. The SR authored by Auais et al included the same three RCTs, plus 11 additional studies focused on cost or qualitative research studies.[133]

Findings

All three included RCTs confirmed fractures through medical records or radiology reports and were powered for evaluating differences in any fracture (SCOOP, SOS) or MOF (ROSE). SOS also reported MOF outcomes and a broader composite of osteoporotic fractures, while SCOOP reported osteoporotic fractures excluding those of the hand, foot, skull, or cervical vertebrae. All three studies reported hip fractures as secondary outcomes.

The impact of screening on fracture and mortality outcomes is depicted in **Figure 3**. For fracture outcomes, we used the per-protocol results from the ROSE trial in our quantitative synthesis because this comparison was the most methodologically comparable with the intention to treat (ITT) analytic results in the SCOOP and SOS trials. Randomization in the SCOOP and SOS trials occurred after collection of the data that were needed to calculate a baseline FRAX score. The ROSE trial randomized participants prior to mailing the questionnaire used to collect information required to calculate baseline FRAX and nearly 40 percent of those randomized did not return the baseline questionnaire and so could not be included in subsequent steps of the screening intervention. The pooled RR for the effect of screening on hip fractures was 0.83 (95% CI, 0.73 to 0.93; 3 RCTs; 42,009 participants; I^2=0.0%), and the pooled RR for MOF was 0.94 (95% CI, 0.88 to 0.99; 3 RCTs; 42,009 participants; I^2=0.0%) (**Figure 3**). When we removed SCOOP from the MOF analysis, the pooled RR estimate was 0.93 (95% CI, 0.86 to 1.00; see note in **Figure 3**). The pooled estimates for "all" fractures or "osteoporotic" fractures favored screening but were not statistically significant (**Figure 3**). The ARDs across these outcomes ranged from 5 to 6 fewer fractures per 1,000 participants for screening compared with usual care. No significant association was observed for all-cause mortality; we calculated a pooled RR of 0.99 (95% CI, 0.95 to 1.04; 3 studies; 57,633 participants; I^2=0%), which corresponds to an absolute effect of 1 fewer death per 1,000 persons screened (95% CI, from 5 fewer to 4 more).[120, 124, 126]

Except for one outcome in the SCOOP trial, the authors did not observe any statistically significant differences in any reported fracture outcomes ("all," osteoporotic, MOF, or hip) or mortality outcomes in the three trials over the years of followup, which ranged from 3.7 to 5 years. The SCOOP trial reported a statistically significant reduction in hip fractures in the screening vs. control group (adjusted hazard ratio [aHR], 0.72 [95% CI, 0.59 to 0.89]), which was a prespecified secondary endpoint.[120] We also conducted a sensitivity analysis (**Appendix E Figure E.1-1)** for osteoporotic, MOF, and hip fracture outcomes using the ROSE ITT analytic sample. In this analysis, the pooled RRs for osteoporotic, MOF, and hip fractures moved closer to the null effect and were no longer statistically significant for MOF and hip fracture compared with our main analysis that used the per-protocol analytic sample.

The fracture results reported in the three SRs[130-133] that we identified were consistent with our pooled findings using the ROSE per-protocol analytic sample (**Appendix D Table 9)**. The pooled estimate for all-cause mortality reported in one of the SRs[131] was also similar to our estimate.

Findings in Specific Populations

All three RCTs conducted subgroup analyses. In the ROSE trial, authors carried out three subgroup analyses by age (65 to 69 years, 70 to 74 years, and 75 years or older) and reported no significant effect modification by age (results not shown by authors).[126] ROSE authors also adjusted for differences in baseline characteristics such as prior fracture and found no significant effect modification. In SOS, authors adjusted analyses for significant differences in baseline characteristics and observed no significant interaction effect with age, history of fracture after age 50 years, or recent fracture for the primary outcome of all fractures (p=0.60, 0.48, and 0.34, respectively).[124] In SCOOP, authors observed a significant interaction effect with baseline FRAX risk (as a continuous measure) for hip fracture (p=0.02), but not for other fracture outcomes, after controlling for baseline characteristics.[120] A related finding was observed in the second per-protocol analysis for ROSE; authors observed that most of the between-group differences in MOF events were driven by differences in the hip fracture component of that composite outcome.

KQ 2. Accuracy of Screening Strategies

KQ 2a. Predictive Accuracy of Osteoporosis and Fracture Risk Assessment Instruments

Summary

Thirty cohort studies (published in 49 articles[72, 73, 139-181, 310-313]) and six SRs[131, 134-138] reported on the accuracy (discrimination or calibration or both) of 11 risk assessment instruments for predicting fracture (KQ 2a). We judged all of the SRs to be good quality; however, authors of SRs generally rated the included primary studies as poor quality, and we also evaluated all of the primary studies we included as poor quality. We relied primarily on findings reported by the SRs, supplemented by results from the primary studies.

Two SRs[131, 132, 137] and 25 cohorts reported in 40 articles[72, 73, 139, 141-158, 160, 161, 163-173, 176, 177, 181, 311-313] reported on calibration outcomes for six risk assessment models (FRAX, Fracture Risk Evaluation Model [FREM], FRC, Garvan, Osteoporosis Self-Assessment Tool [OST], QFracture) for the prediction of MOF, hip fracture, or both. Calibration results were heterogenous with no discernible patterns with respect to instrument, age, or sex.

Six SRs[131, 132, 134-138] and 16 cohorts published in 25 articles[73, 139-143, 146, 159, 160, 163, 165, 172-181, 311, 312] reported on the discriminative accuracy of 11 risk assessment models (EPIC, FRAX, FRC, FREM, Garvan, ORAI, OSIRIS, OST, QFracture, SCORE, Women's Health Initiative [WHI] Prediction Model) to predict MOF or hip fracture or both using primarily AUC. Findings were heterogenous, spanning a range considered *poor* accuracy (AUC, 0.52) to *very good* accuracy (AUC, 0.93); however, most were between 0.60 and 0.80. Sources of heterogeneity in AUC estimates likely include age and source of population evaluated, variation in outcome definitions, and differences in analytic methods used by authors. Discrimination was largely similar in men and women. For risk assessment instruments with the option to include BMD as an input (FRAX, FRC, Garvan), the predictive accuracy was improved when BMD was included compared with when it was not included. Further, some instruments (FRAX, FRC, Garvan, QFracture) had higher accuracy for predicting hip fracture than for predicting MOF. Few studies reported sensitivity or specificity of specific risk thresholds. In one cohort of U.S. women ages 50 to 64 years, a FRAX risk threshold of 9.3 percent had a 26 percent sensitivity and 83 percent specificity to predict MOF.[139]

Study Characteristics

Six good-quality SRs reported on the predictive accuracy (calibration, discrimination, or both) of various risk assessment instruments.[131, 132, 134-138] Five were new to this update.[131, 132, 135-138] Three SRs[134, 135, 138] reported only discrimination outcomes, two SRs[136, 137] reported both discrimination and calibration outcomes, and one SR[131, 132] primarily focused on calibration outcomes and included discrimination outcomes that were reported in the previous USPSTF review on this topic.[3] Some primary research studies included in the three SRs reporting only discrimination outcomes reported calibration outcomes, and we have included these calibration data as new in this update. Detailed study and population information for the primary research studies is in **Appendix D Table 2**, and detailed information about the included SRs is in **Appendix D Table 12**.

We observed substantial overlap of included studies across the included SRs. Marques et al, published in 2015, used search dates through September 2014 and included 45 studies; however, accuracy data were not reported from all studies that were included.[134] Jiang et al, published in 2017, used search dates through July 2016 but only focused on the predictive accuracy of the FRAX instrument, so only seven studies were included.[136] Beaudoin et al, published in 2019, used search dates through August 2017[135]and included 53 studies.[135] Sun et al, published in 2022, used search dates through April 2021 and included 68 studies of 70 prediction models and 138 reports of external validation.[137] The review by Gates et al (in support of the CTFPHC) was published in 2023 and included search dates through June 2021 and included 59 articles from 32 unique cohorts.[131, 132] The most recent review, Adami et al, included search dates through December 2020 and included 43 articles; however, this review was focused exclusively on FRAX and two less commonly used risk assessment instruments (FRA-HA and DeFRA).[138] We note that most of these reviews included some studies from the development cohorts used to develop the risk assessment instruments, were conducted in countries that we excluded from this update as not very highly developed per the United Nations' Human Development Programme Index[104] (e.g., China), or were conducted in countries with a different category of hip fracture incidence than the United States (e.g., Denmark, high incidence).[105] Detailed study quality ratings for the SRs are in **Appendix D Tables 33–37**. We did not evaluate the risk of bias for the primary studies included in these SRs. However, we note that authors of the most recent and comprehensive SRs judged their included studies as nearly all high risk of bias.[131, 137]

In addition to the SRs, we identified 30 cohorts (published in 49 articles[72, 73, 139-181, 310-313]) that reported discrimination, calibration, or both and that (with some exceptions) were either not included in the SRs or were published subsequent to the search dates covered by the SRs. Several of these studies reported data for more than one risk assessment instrument. We assessed nearly all studies as poor quality (i.e., high risk of bias) (**Appendix D Tables 28–32**) for all instruments evaluated because either no fracture risk model development study has been performed (e.g., OST), or the original model development studies were assessed as high risk of bias (**see Appendix G**) and the external validation analyses included in this update had risk of bias related to patient selection bias, missing data for risk factors or outcomes, and deviations in how risks and outcomes were ascertained as compared with the development cohort.

Of the 30 primary research cohorts, 13 articles were published representing five unique U.S. cohorts.[72, 73, 139-143, 145, 146, 152, 153, 165, 311] One of these U.S. cohorts[72, 145, 146] was exclusively among men, one was conducted in a mixed-sex population,[311] and the rest were exclusively in women. The remaining cohorts were from Canada, Australia, New Zealand, Japan, Israel, Belgium, France, Portugal, Poland, and Spain. Most were exclusively women, but 11 cohorts included mixed populations of men and women, and in some cases, results were reported separately for men and women. The mean or median age ranged from approximately 50 years to 75 years. Cohorts were either retrospectively assembled based on clinical or administrative healthcare use data, such as BMD registries, electronic health records, or billing claims data; or data were collected in prospective, population-based epidemiologic studies or clinical trials, sometimes but not always focused on osteoporosis.

Findings

Calibration

Detailed calibration findings from the included SRs and the primary research studies are reported in **Appendix D Table 11**.

The Gates et al SR (in support of the CTFPHC) synthesized calibration findings for FRAX (with and without BMD), Garvan (with and without BMD), QFracture, FRISC, and FRC (with and without BMD).[131, 132] For FRAX, authors stratified results by study risk of bias. Authors concluded with very low certainty that FRAX demonstrated poor performance among the high risk of bias studies (13 studies for hip fracture without BMD, 12 studies for MOF without BMD, 13 studies for hip fracture with BMD, and 16 studies for MOF with BMD). The authors concluded with low (hip fractures, 3 studies) and moderate (MOF, 3 studies) certainty that FRAX without BMD may be well-calibrated among the three studies with unclear risk of bias that were specifically evaluating external validations of the FRAX-Canada model.[131, 132] Further, authors concluded with low certainty that FRAX with BMD may perform poorly for hip fractures (3 studies) but had moderate certainty that it was probably well-calibrated for MOF (3 studies).[131, 132]

Authors of the Sun et al SR reported that calibration measures were only reported by primary study authors for 33 (24%) of the 138 models evaluated, and 31 (22%) showed "good fitness."[137] However, the SR authors reported only 22 (16%) used suitable methods for measuring calibration.

Of the primary research studies we identified, 25 cohorts were described in 40 articles[72, 73, 139, 141-158, 160, 161, 163-173, 176, 177, 181, 311-313] that reported on calibration. Most focused on FRAX calibration. In the WHI cohort, the overall observed vs. expected ratio for FRAX was 1.0 (range of 0.76 to 1.15 across risk categories), and the calibration slope was 1.04.[139] For hip fracture, the overall observed-to-expected ratio was 1.0 (range, 0.27 to 1.63 across risk categories), and the calibration slope was 1.59, with significant overprediction at the lowest risk categories and significant underprediction at the three highest categories.[139] Calibration appeared similar when stratified by race and ethnicity in the two analyses among women ages 50 to 64 years that reported data by race or ethnicity.[141, 143] Data from the other two U.S. cohorts of women were somewhat limited; FRAX appeared to underestimate risk in older age groups[165] and underestimate risk in women with obesity.[152, 153] Data from the male U.S. cohort (MrOs) were also somewhat limited; one analysis suggested the risk of MOF (with or without BMD) was overestimated and the risk of hip fracture (with BMD) was underestimated.[145] In the other analysis of the MrOs cohort, the Hosmer-Lemeshow goodness-of-fit values suggested poor calibration for both MOF and hip fracture.[146] Data related to the calibration of other instruments were limited and are reported in **Appendix D Table 11**.

Discrimination

Six SRs[131, 132, 134-138] and 16 cohorts published in 25 articles[73, 139-143, 146, 159, 160, 163, 165, 172-181, 311, 312] not included in one or more of the SRs (with some exceptions) reported discrimination outcomes (AUC, sensitivity, specificity). Detailed findings are in **Appendix D Table 10** (primary studies) and **Table 12** (SRs). **Figures 4a** (women), **4b** (men), and **4c** (mixed-sex) summarize predictive discrimination with respect to AUC outcomes organized by instrument and by whether results were obtained from SRs or primary research studies. The AUCs varied widely depending on instrument, inclusion of BMD input as a risk, fracture type predicted, age range of the population evaluated, and whether authors were reporting the overall AUC (maximum AUC possible over all potential thresholds) or an AUC associated with a specific threshold. Of the three instruments that can assess risk with or without a BMD input (FRAX, FRC, Garvan), models with BMD generally reported higher AUCs than the same model without the inclusion of BMD. Studies evaluated FRAX, FRC, Garvan, and QFracture in men, women, and mixed-sex populations, and findings appeared similar across these populations. OST and WHI were only evaluated in women. Of the four instruments predicting risk for both MOF and hip fracture (FRAX, FRC, Garvan, QFracture), predictive accuracy appeared generally higher for prediction of hip fracture than MOF. For studies reporting outcomes specifically for women ages 50 to 64 years, the AUCs ranged from 0.52 to 0.71 across instruments. For the other studies of women (not reporting by age), the AUCs ranged from 0.63 to 0.89. For studies reporting outcomes for men, the AUCs ranged from 0.63 to 0.93. For studies reporting outcomes for mixed-sex populations, the AUCs ranged from 0.61 to 0.88.

Compared with the number of studies reporting AUC outcomes, fewer studies reported on sensitivity and specificity, and across studies the thresholds evaluated varied. FRAX was the most commonly reported-on instrument. In the WHI cohort of women ages 50 to 64 years not taking osteoporosis medication (n=62,492), the sensitivity of FRAX without BMD input for MOF risk greater than 9.3 percent was 26 percent, and the specificity was 83 percent.[139] In a cohort of Spanish women (n=1,090), the sensitivity of a 5 percent threshold was 61 percent, and the specificity was 72 percent.[174] For MOF with BMD input, the sensitivity of a fracture risk of 20 percent or higher was 20 percent, and the specificity was 93 percent in the Manitoba BMD registry cohort (n=54,459).[160] From the same cohort, the sensitivity was 62 percent and specificity 79 percent for a fracture risk of 3 percent or higher for prediction of hip fracture with BMD input.

With respect to instruments other than FRAX, sensitivity and specificity varied. One study reported on the sensitivity (83%) and specificity (65%) of a 3 percent risk threshold for QFracture in predicting hip fracture in an Australian cohort of men and women ages 40 to 89 years with diabetes (n=1,251).[173] Two studies reported on the sensitivity and specificity of OST at a score threshold of less than 2. Among women ages 50 to 64 years from the WHI cohort (n=99,431), the sensitivity was 40 percent, and the specificity was 61 percent.[139] In an analysis of women ages 40 to 59 years from the Manitoba BMD registry, the sensitivity was 46 percent, and the specificity was 62 percent (n=8,254).[159]

The studies reporting on the predictive accuracy in the MrOs cohort of men (n=5,200) selected risk thresholds equivalent to a sensitivity of 90 percent for all instruments evaluated (FRAX, Garvan, QFracture), precluding a comparison with findings from other cohorts.[146] No studies reported on the sensitivity or specificity for the FRC.

Accuracy outcomes by race and ethnicity. Accuracy results stratified by race or ethnicity were only reported by one cohort published in three articles.[140, 142, 143] The WHI cohort, which was 89 percent White, reported findings for FRAX, Garvan Fracture Risk Calculator, and OST stratified by White, African American, Hispanic, and Asian race or ethnicity. However, results were only reported for women ages 50 to 64 years. CIs for AUC estimates were largely overlapping for the various race and ethnicity subgroups, precluding any conclusions about differences in predictive accuracy by race or ethnicity.

KQ 2b. Predictive Accuracy of BMD Measurement for Incident Fractures

Summary

Twenty-eight publications from 22 unique cohorts, a third of which we deemed poor quality, reported on the accuracy of BMD measurement (typically at the FN) for prediction of incident fractures.[15, 146, 150, 151, 154, 156, 157, 160-162, 166, 172, 175, 176, 179, 182-194] Most studies were conducted among women, and the mean age of participants varied from 49 years to 75 years with followup for incident fractures of 8 to 12 years, although some had shorter or longer followups. Four cohorts reported at least one type of calibration outcome, but few reported detailed information or the same information to allow for comparison across studies.[146, 150, 185, 192] Thirteen unique cohorts reported on the discrimination of BMD as a continuous variable for predicting MOF with AUCs ranging from 0.60 to 0.80.[146, 150, 151, 154, 156, 157, 166, 172, 175, 176, 179, 183, 189, 190, 193] Twelve cohorts reported AUC outcomes for predicting hip fracture ranging from 0.64 to 0.86.[146, 150, 151, 154, 157, 166, 172, 176, 179, 183, 185, 187, 193, 194] Substantial heterogeneity precluded quantitative synthesis, but the AUC estimates for hip fracture appeared higher than the estimates for MOF. Fewer studies reported sensitivity and specificity and thresholds varied. Discrimination outcomes for men appeared similar to estimates for the overall body of evidence, which was predominantly in women. Discrimination outcomes for women younger than age 65 years were limited to two studies.[160, 190]

Study Characteristics

Twenty-eight publications from 22 unique cohorts reported on the accuracy of BMD measurement for prediction of incident fractures.[15, 146, 150, 151, 154, 156, 157, 160-162, 166, 172, 175, 176, 179, 182-194] Individual study details are in **Appendix D Table 3**. We assessed 10 of these analyses (covering 7 unique cohorts) as poor quality.[15, 151, 156, 175, 176, 179, 185, 186, 189, 194] The rest were fair quality. Detailed study quality ratings are in **Appendix D Tables 38–42**. Three unique cohorts were from the United States;[146, 187, 192] the rest were from Canada (2 cohorts[154, 157, 160-162, 183, 184, 193]), countries in Europe (8 cohorts[15, 175, 176, 179, 185, 188-190]), countries in Asia (5 cohorts[150, 151, 156, 182, 194]), Australia (2 cohorts[186, 191]), New Zealand (1 cohort[166]), or Israel (1 cohort[172]).

Most analyses used data collected from prospective, population-based epidemiologic studies focused either on bone health, osteoporosis, or aging generally. The exceptions were an analysis based on the WHI (United States) that used data from both the clinical trial and observational study components,[187] an analysis based on electronic health data and administrative billing data collected through usual care in Israel,[172] and a provincial BMD registry with administrative billing data in Canada.[157, 160-162, 183, 193] The cohort sample sizes varied from 400 to 68,730 persons, and the mean age of included persons varied from 49 years to 75 years. Two cohorts were exclusively men,[146, 182] four cohorts included both men and women,[15, 154, 157, 161, 162, 176, 193] and the rest were exclusively women. Only persons with available BMD measurements were included in the reported analyses; other inclusion/exclusion criteria varied across cohorts. For example, some of the reported analyses excluded persons who were known to be taking antiosteoporosis medication[146, 150, 151, 161, 166, 189, 191, 193] or who were known to have secondary osteoporosis or metabolic bone disease,[156] while other cohorts allowed persons on treatment or did not specify treatment status. Similarly, the proportion of persons with a history of fracture at baseline enrollment also varied and was reported using different definitions. Among the studies conducted exclusively or mostly in women that reported mean T-score at the FN at baseline, T-scores ranged from -1.0 to -1.5. The two studies conducted exclusively in men did not report baseline T-scores.[146, 182] Among the studies reporting the prevalence of osteoporosis at baseline, the range was 4.9 to 31.7 percent.

Findings

Individual cohort findings are in **Appendix D Table 13**. Across these cohorts, incident fractures were reported over a followup ranging from 4 to 25 years; however, followup of 8 to 12 years was most common because many studies also evaluated the predictive accuracy of fracture risk assessments designed to predict fracture over a 10-year time period. In the cohorts reporting on men, the incidence of MOF was 3.7 to 10.7 percent, and the incidence of hip fracture ranged from 1.5 to 4.2 percent over a followup of 5.4 to 15.8 years. Among the cohorts reporting on women, the incidence of MOF ranged from 3.3 to 15.0 percent, and the incidence of hip fracture ranged from 0.5 to 15.9 percent over followup of 4.5 to 25 years. In addition to differences in length of followup, the anatomic site of BMD measurement varied across studies; FN was most commonly used, but some studies reported outcomes based on measurement at the TH or LS or based on the lowest measurement from the FN, TH, or LS.

Calibration

Four cohorts reported at least one type of calibration outcome, but few reported detailed information or the same information to allow for comparison across studies (**Appendix E.2 Table 1**).[146, 150, 185, 192] Only two cohorts reported goodness-of-fit outcomes (poor in 1 cohort[185] and good in the other[146]). Only one cohort reported calibration plots, which showed a dose-response effect across quartiles of predicted risk but no other information to interpret the calibration.[150] One cohort reported hip and nonvertebral fractures after 25 years; participants in the highest quartile of BMD had lower incidence of fractures compared with those in the lower quartile of BMD.[192] Lastly, one study reported the observed-to-expected ratio for hip fracture (0.83 [95% CI, 0.65 to 1.04]), suggesting poor calibration.[185]

Discrimination

Twenty-six articles reporting on 20 unique cohorts reported discrimination outcomes (**Figure 5**).[15, 146, 150, 151, 154, 156, 157, 160-162, 166, 172, 175, 176, 179, 183-191, 193, 194] Thirteen unique cohorts reported on the discrimination of BMD as a continuous variable for predicting MOF with AUCs ranging from 0.60 to 0.80.[146, 150, 151, 154, 156, 157, 166, 172, 175, 176, 179, 183, 189, 190, 193] Twelve unique cohorts reported AUC outcomes for predicting hip fracture ranging from 0.64 to 0.86.[146, 150, 151, 154, 157, 166, 172, 176, 179, 183, 185, 187, 193, 194] Substantial heterogeneity precluded quantitative synthesis, but the AUC estimates for hip fracture appeared higher than the estimates for MOF.

Fewer studies reported sensitivity and specificity outcomes. In studies that used a BMD T-score of less than -2.5 as the threshold for a positive test, the sensitivity ranged from 17.5 to 51.3 percent for MOF[150, 160-162, 175] and from 25.0 to 66.7 percent for hip fractures.[15, 150, 160, 162, 187] The specificity for MOF ranged from 70.9 to 95.4 percent[160, 161, 175] and from 88.6 to 94.0 percent for hip fractures.[15, 160, 162, 187]

Discrimination outcomes in younger women. Only two studies reported on the discrimination of BMD alone specifically in younger women.[160, 190] In one population-based prospective cohort study of women ages 45 to 54 years in the United Kingdom, the AUC for predictive accuracy of continuous BMD at the FN was 0.64 (95% CI, 0.63 to 0.66) over a followup of 3 to 12 years; sensitivity and specificity were not reported.[190] In a BMD registry from Manitoba, Canada, the prediction of MOF, based on a T-score less than -2.5, had a sensitivity of 6.7 percent for women ages 40 to 49 years, 9.7 percent for women ages 50 to 59 years, and 18.5 percent for women ages 60 to 69 years compared with 30.1 percent for women ages 70 to 79 years and 49 percent for women age 80 years or older.[160] Similarly, specificity decreased from 98 percent in women ages 40 to 49 years to 69 percent for women age 80 years or older.[160] For the prediction of hip fractures, a similar pattern was observed with the lowest sensitivity for women ages 40 to 49 years (19%) and the highest sensitivity for women age 80 years or older (54%).[160]

Discrimination outcomes in men. Only one study that exclusively enrolled men reported discrimination outcomes.[146] In this retrospective analysis of participants in a community-based population study of mostly White men age 65 years or older, the AUC for continuous BMD over a followup of 15.8 years was 0.76 (95% CI, 0.71 to 0.80) for the prediction of MOF and was very similar for the prediction of hip fracture (0.76 [95%, CI, 0.721 to 0.81]).[146] The T-score threshold cutoff associated with a sensitivity of 90 percent for MOF prediction was -0.21 and for hip fracture was -0.36; both used a young, White female reference range for T-scores.[146]

Three analyses reported outcomes separately for women and men from within the same study.[15, 162, 176] One analysis reported data from three population-based cohort studies in Portugal (N=1,897). Marques et al reported AUC estimates in men that were higher for prediction of both MOF (0.80 vs. 0.66) and hip fracture (0.82 vs. 0.68).[176] In Trajanoska et al, a population-based study from the Netherlands (N=11,052), AUCs were not reported, but the sensitivity was lower in men (20% vs. 38%) and specificity was higher (94% vs. 91%) for the prediction of hip fracture over 11 years of followup based on a threshold T-score of less than -2.5.[15] A similar pattern was observed for nonvertebral fractures. In data from the Manitoba BMD registry, sensitivity was also lower in men compared with women for the prediction of MOF (18% vs. 28%) and the prediction of hip fracture (31% vs. 43%), while the specificities were very similar for each fracture type between sexes (89% for women, 92% for men).[162]

Two studies reported on mixed-sex populations of men and women;[154, 157] these estimates appear similar to the estimates from studies that exclusively analyzed women or men with AUCs ranging from 0.66 to 0.68 for MOF prediction and 0.76 to 0.80 for hip fracture prediction.

Accuracy outcomes by race and ethnicity. Of the studies reporting the race or ethnicity of participants, studies enrolled exclusively or predominantly White participants (89% or more). No studies reported calibration or discrimination outcomes by race or ethnicity.

KQ 2c. Diagnostic Accuracy of Risk Assessment Instruments for Identifying Osteoporosis

Summary

Forty-three unique cohorts (published in 54 articles) reported on diagnostic test accuracy of 15 risk assessment instruments for identifying osteoporosis.[141, 143, 159, 195-245] More than half of the studies enrolled populations with a mean age between 60 and 69 years and studies included women, men, or both. Differences in reference standards, risk assessment score thresholds, and study populations precluded a quantitative synthesis. In women, AUCs ranged from 0.32 to 0.87 across 35 articles evaluating 11 instruments (**Figure 6a**). Five articles reported results from three independent cohorts that retrospectively evaluated the accuracy of a FRAX MOF risk threshold of 8.4 percent or 9.3 percent in women ages 50 to 64 years; AUCs ranged from 0.55 to 0.62. In men, AUCs ranged from 0.62 to 0.94 across 18 articles evaluating 12 instruments (**Figure 6b**). Three articles reported on accuracy among mixed populations of men and women for the FRAX, OST, or Garvan Fracture Risk Calculator (**Figure 6c**).[196, 239, 241] Findings in these studies were consistent with the findings reported for men and women separately. Several studies reported findings stratified by age, but few studies reported findings stratified by race or ethnicity.

Study Characteristics

We identified 54 articles reporting on diagnostic test accuracy of risk assessment instruments for identifying osteoporosis (**Appendix D Table 4**) from 43 unique cohorts.[141, 159, 195-245] Sixteen studies[197, 198, 201, 202, 212, 216, 218, 219, 221, 229, 232, 235, 237, 238, 240, 244] were conducted exclusively in men, three studies[196, 239, 241] were conducted among a mixed population of men and women (proportion of women ranged from 45% to 87%) but did not report results separately for men and women, and two studies[233, 243] included men and women but reported results separately by sex; the rest of the studies were conducted exclusively in women. The mean age across studies ranged from 50.5 to 80.4 years, with just over half of the studies enrolling populations with a mean age between 60 and 69 years. We rated one study[205] as good quality; the rest were fair quality. Detailed study quality ratings are in **Appendix D Tables 43–47**. Common risk-of-bias issues included lack of consecutive or random enrollment of patients, no information about blinding of index and reference tests, and lack of information about interval between risk assessment and DXA testing. Further, about a third of studies were conducted on data collected during usual care from persons referred for DXA; the rest of the studies recruited persons from healthcare settings or were population-based cohort studies. Twenty-two analyses were conducted in U.S. cohorts.[141, 143, 195, 197, 198, 202, 216, 219-222, 224, 226, 230-232, 234-238, 241]

Fifteen unique risk assessment instruments were evaluated as index tests for identifying osteoporosis. Most instruments were originally developed to identify persons at high risk for osteoporosis; however, three instruments (FRAX, Garvan, and the **V**eterans **A**ffairs **F**racture **A**bsolute **R**isk **A**ssessment [VA-FARA]) were originally designed as fracture risk prediction instruments.[141, 195, 196, 200, 220, 232-239, 241] Authors evaluated instruments against a reference standard of a T-score based on DXA BMD measurement most commonly at the FN, but many studies also used measurements at the TH or LS or against the lowest T-score from across the three sites. Methods used to determine discrimination varied; authors either computed AUC across the range of all possible threshold (i.e., "continuous" or "overall" AUC) or computed AUC with respect to a specific threshold, or both.

Findings

Studies reported discrimination outcomes including AUC, sensitivity, and specificity to describe the accuracy of these risk assessments for identifying osteoporosis. Some studies reported accuracy outcomes for more than one risk assessment instrument for the same study population, and some studies reported sensitivity and specificity outcomes using different risk assessment score thresholds, often prespecified but sometimes empirically derived to maximize sensitivity. In some cases, results for women and men were presented separately, and in other cases results for the "mixed" population of men and women were reported as one estimate. This heterogeneity precluded a quantitative synthesis of accuracy results.

The instrument most commonly evaluated was the **O**steoporosis **S**elf-assessment **T**ool (OST), which was reported in 26 unique cohorts from 31 articles.[143, 159, 195, 196, 198-201, 203, 205, 208-210, 214, 216-219, 221, 223, 225, 228, 230, 231, 234-237, 240, 242, 243] Other instruments reported in more than 10 articles included the **O**steoporosis **R**isk **A**ssessment **I**nstrument (ORAI) reported in 21 unique cohorts in 22 articles,[199, 203, 206-210, 214, 217, 223-228, 230, 231, 233, 234, 236, 242, 245] the **S**imple **C**alculated **O**steoporosis **R**isk **E**stimation (SCORE) reported in 18 cohorts in 20 articles,[195, 203, 206-211, 217, 222, 224, 226, 228, 230, 231, 233, 234, 236, 242, 245] and FRAX reported from 12 unique cohorts in 15 articles.[141, 195, 196, 200, 220, 232-239, 241] A summary of findings is depicted in **Figures 6a** (women), **6b** (men), and **6c** (mixed populations) with detailed findings in **Appendix D Table 14**. A detailed narrative description of findings organized by risk assessment instrument is in **Appendix E.3**.

Accuracy in Women

We identified 28 unique cohorts (reported in 35 articles) for 11 risk assessment instruments evaluated in populations that were exclusively women or that reported results separately for women (**Figure 6a**).[141, 143, 159, 195, 199, 200, 203-211, 213-215, 217, 220, 222-228, 230, 231, 233, 234, 236, 242, 243, 245] The instruments evaluated in women included **A**ge, **B**one, **N**o **E**strogen (ABONE), **A**ge, **ME**nopause, **M**enarche, **B**MI (AMMEB), FRAX, Garvan, **N**ational **O**steoporosis **F**oundation tool (NOF), ORAI, **OS**teoporosis **I**ndex of **RIS**k (OSIRIS), OST, OST for **A**sians (OSTA), SCORE, and **S**tudy of **O**steoporotic **F**ractures Research Group **S**tudy **U**tilizing **R**isk **F**actors (SOFSURF). Across instruments, the AUC ranged from 0.32 to 0.87, excluding one study that we deemed an outlier because of extreme values.[245] Sensitivity ranged from 5 to 100 percent, and specificity ranged from 0 to 100 percent; however, these ranges represent different score thresholds, some of which were prespecified and some of which were empirically derived to maximize sensitivity. A detailed description of findings for each risk assessment instrument is in **Appendix E.3**.

Accuracy in Women Younger Than Age 65 Years

Several articles reported on accuracy of risk assessment instrument specifically among women ages 50 to 64 years (selected parts of **Figure 6a**). Five articles[141, 195, 220, 234, 236] reported results from three independent cohorts that retrospectively evaluated the accuracy of the USPSTF's present (8.4%) or former (9.3%) suggested FRAX MOF risk threshold for DXA screening in women younger than age 65 years. The AUC in these studies ranged from 0.55 to 0.62, sensitivity ranged from 5 to 49 percent and specificity ranged from 63 to 96 percent. In one study from the WHI (N=8,134), the sensitivity was 5 percent among women ages 50 to 54 years, 17 percent among women ages 55 to 59 years, and 49 percent among women ages 60 to 64 years.[141] The sensitivities of FRAX for the USPSTF's suggested threshold reported by the other included articles ranged from 24 to 37 percent. The specificity across these five studies ranged from 63.4 to 95.8 percent. In addition to reporting on the accuracy of a specific FRAX risk threshold, two studies using data from the WHI also reported the AUC of FRAX when treated continuously without respect to a specific risk threshold; the AUC was 0.59 for MOF[140] and was 0.68 for hip.[143]

Several articles also reported on the accuracy of other risk assessment instruments among women younger than age 65 years. Six cohorts (in 8 articles[143, 159, 195, 200, 209, 228, 234, 236]) reported an AUC for OST of 0.63 to 0.83, six cohorts[195, 209, 226, 228, 234, 236] reported an AUC of 0.58 to 0.87 for SCORE, and five cohorts[209, 226, 228, 234, 236] reported an AUC for ORAI of 0.60 to 0.84.

Accuracy in Men

We identified six studies for four risk assessment instruments that were developed exclusively for men (**Figure 6b**).[197, 202, 216, 218, 232, 237] The **M**ale **O**steoporosis **R**isk **E**stimation **S**core (MORES, 2 cohorts in 3 articles[197, 202, 232]), the **M**ale **O**steoporosis **S**creening **T**ool (MOST, 1 cohort[218]), the **M**ale **S**imple **C**alculated **O**steoporosis **R**isk **E**stimation (MSCORE, 1 cohort[216]) and the VA-FARA (1 cohort[237]). In these studies, AUCs ranged from 0.64 to 0.88. These estimates were similar to those observed for other non-male specific risk assessment instruments (e.g., OST) evaluated within these cohorts.

We also identified studies evaluating other risk instruments that were not developed specifically for men (**Figure 6b**). These included the ABONE, FRAX, Garvan, ORAI, OSIRIS, OST, OSTA, and SCORE. Across these other instruments, the AUCs ranged from 0.62 to 0.94.[198, 201, 212, 216, 218, 219, 221, 229, 232, 233, 235, 237, 238, 240, 243, 244] A detailed description of findings for each risk assessment instrument is in **Appendix E.3**.

Accuracy in Mixed-Sex Populations

Three studies reported on accuracy among mixed populations of men and women for FRAX, OST, and Garvan (**Figure 6c**).[196, 239, 241] Findings in these studies were consistent with the findings reported for men and women separately.

Accuracy by Age

In addition to the studies related to FRAX for women younger than age 65 years discussed above, nine cohorts reported in 11 articles reported findings on other instruments stratified by age.[159, 195, 198, 200, 209, 221, 226, 228, 234, 236, 240] Three cohorts reported findings exclusively among men[198, 221, 240] for the OST instrument, while eight articles reported findings exclusively among women for the NOF, ORAI, OST, and SCORE instruments.[159, 195, 200, 209, 226, 228, 234, 236]

Among women, the AUCs in the studies reporting by age ranged from 0.58 to 0.87 across instruments. Sensitivity ranged from 44 to 100 percent, and specificity ranged from 10 to 81 percent, but score thresholds used to determine sensitivity and specificity varied by study, precluding direct comparisons. Meaningful differences in findings by age were observed for the same instrument evaluated by different studies. For example, in a population-based sample of postmenopausal women from the Rochester, MN, region (N=202), authors reported age-stratified results for the women ages 45 to 64 years and the women age 65 years or older for the NOF, ORAI, and SCORE. Differences in AUCs and sensitivity between age strata were small with overlapping CIs, suggesting no meaningful differences between age groups.[226] However, large differences in specificity were observed for ORAI and SCORE, with specificity across the three instruments ranging from 0 to 8 percent in the older women and from 19 to 69 percent in the younger women.[226] Yet, in a study published 10 years later from the same clinical setting among women ages 50 to 64 years (N=290) and using the same score thresholds that were used in the earlier study, the AUCs reported for ORAI and SCORE were more than 0.2 units lower than those reported in the earlier study for both instruments, and sensitivity was also meaningfully lower (sensitivity 99% and 100% for ORAI and SCORE, respectively, vs. 52% and 74% in the later study).[234] However, these differences may be partially explained by the use of a different reference standard in the later study (BMD at the FN or LS vs. BMD at the FN only used in the earlier study).

One study conducted among Caucasian women (N=4,025) referred for DXA in a single Belgian city reported discrimination stratified by age (45 to 64 years, 65 years or older).[209] In this study, the AUCs for ORAI, OST, and SCORE were similar in both age strata (range, 0.75 to 0.76); however, the authors chose different scoring thresholds to determine a positive test for the different age groups, precluding a direct comparison of sensitivity between age groups.[209]

In a study of men enrolled from specialty clinics in Veterans Affairs (VA) settings (N=181), the AUCs for the OST instrument ranged from 0.70 to 0.99 across four age categories from 50 to 59 years to 80 years or older; however, there was not a clear linear trend: the lowest AUC was in the age group 70 to 79 years and the highest was in the age group 80 years or older.[221] In a separate study of men enrolled from four VA sites (N=518), the sensitivity of the OST was higher and specificity was lower among men older than age 65 years compared with younger men at both of the score thresholds reported (OST≤6, OST≤0).[198] In another study among men referred for DXA at an academic health center in Taiwan (N=834), the AUCs for the OST instrument were similar among men younger than age 65 years (0.66) and men age 65 years or older (0.68).[240]

Accuracy by Race or Ethnicity

Six cohorts reported findings stratified by race or ethnicity.[140, 142, 143, 198, 202, 216, 221, 224] Four of the six cohorts were men, and three of those studies were among men recruited from VA clinical settings. Substantial heterogeneity with respect to instruments, score thresholds used, and racial categories evaluated precludes any definitive conclusion about differences in accuracy by race or ethnicity. In one VA study (N=518), the sensitivity of the OST was higher and the specificity lower for both score thresholds reported (OST≤6, OST≤0) among Caucasians compared with African Americans; for example, sensitivity for the less than 0 threshold was 25 percent in African Americans and 42 percent in Caucasian participants, and specificity was 87 percent and 85 percent, respectively.[198] In another VA study (N=197), authors reported on accuracy data for MSCORE, OST, and the reduced MSCORE.[216] The sensitivity of these instruments was higher and specificity was lower for African Americans compared with Caucasians, but only when using a Caucasian reference range for calculating T-scores from raw BMD measurements (which is the standard method for calculating T-scores for persons of all races).[216] Outcomes were similar when an African American reference range was used.[216] In the third VA study (N=181), the AUCs reported for White persons (0.85) were reasonably similar to the AUCs reported for Black persons (0.80).[221] In an analysis of U.S. NHANES data (N=2,944 men age 50 years or older), authors reported on the accuracy of MORES for White, African American, Mexican American, and "other" race and ethnicities.[202] Across the groups, sensitivity ranged from 60 percent (White) to 95 percent (other), and specificity ranged from 55 percent (other) to 69 percent (White).[202]

Two cohorts reported on differences in accuracy by race and ethnicity among women. In a cohort of postmenopausal women identified from a university-based family practice (N=226), AUCs were similar for Hispanic and African American persons compared with the full study population for the ORAI and SCORE instruments; sensitivities and specificities varied but were quite imprecise, precluding any definitive conclusions about differences by race or ethnicity.[224] Among women ages 50 to 64 years in the WHI cohort, no discernible pattern of differences in AUC were observed between AUC estimates among persons of different race or ethnicity for the FRAX, Garvan, and OST instruments.[140, 142, 143]

KQ 2d. What Is the Evidence to Determine Screening Intervals, and How Do These Intervals Vary by Baseline or Current Individual Fracture Risk?

Study Characteristics

We identified three new cohort studies[248-250] for this update for a total of five included studies for this KQ.[246-250] We rated two as poor quality[247, 248] and the rest as fair quality; detailed study quality ratings are in **Appendix D Tables 48–52**. Study characteristics are detailed in **Appendix D Table 5** and findings are detailed in **Appendix D Table 15**.[246-248] Four studies were conducted among U.S. cohorts (Framingham Osteoporosis Study,[246] WHI,[249] MrOs,[250] SOF[247]) and the fifth study used data from the Manitoba BMD registry in Canada.[248] The mean age of participants was 60 in the Manitoba cohort,[248] 66 in the WHI cohort,[249] and 72 to 74 years in the other three cohorts. The Framingham Cohort was 61 percent women,[246] MrOs was 100 percent men,[250] and the rest were exclusively women.

All studies used a similar design that evaluated the accuracy of a fracture risk prediction model based on an initial BMD measurement and a repeat BMD measurement at a subsequent interval, which ranged from 4 to 8 years across studies. Followup for fracture ascertainment occurred for 8 to 11 years after the second BMD measurement. Notably, because of this study design, authors excluded participants who experienced a fracture event during the interval between the initial and repeat BMD test.

Findings

In four of the five studies, authors reported similar accuracy when comparing models including only the initial BMD compared with models based on the change in BMD or in models that included both initial BMD and change in BMD. As an illustrative example, the AUC for baseline BMD for predicting MOF in the SOF cohort was 0.68 (95% CI, 0.66 to 0.71), the AUC for BMD change (as a % of initial) was 0.63 (95% CI, 0.61 to 0.66), and the AUC for a model combining initial BMD and change in BMD had an AUC of 0.69 (95% CI, 0.66 to 0.71).[250] In the fifth study, authors reported no association between change in spine, TH, or FN BMD and MOF (HRs, 0.93 to 1.02 per SD increase in BMD, all statistically nonsignificant).[248]

KQ 3. What Are the Harms of Screening for Fracture Risk or Osteoporosis?

Of the three RCTs included for KQ 1, only the SCOOP trial reported on harms of screening.[120, 121] Study and population characteristics for the SCOOP trial are detailed in the KQ 1 section.

Anxiety was assessed using the Strait-Trait Anxiety Inventory-Short Form at repeated intervals over the 5-year study period.[120, 121] Authors observed no difference in anxiety between screening participants (both those deemed low risk and those deemed high risk who were invited to DXA testing) and the control group participants (p=0.515) (**Appendix D Table 8**).

The included SR for KQ 1 conducted to inform the CTFPHC recommendation on screening for primary prevention of fractures reported on overdiagnosis.[131, 132] Based on the data reported in the SCOOP and SOS RCTs, the SR authors estimated the proportion of participants overdiagnosed ranged from 11.8 to 24.1 percent.[131, 132] The method for calculating overdiagnosis in context of being labeled as “high risk” was described in detail in a companion publication and was characterized as evolving by review authors.[314]

KQ 4. What Is the Effectiveness of Pharmacotherapy With Selected FDA-Approved Medications on Fracture Incidence and Fracture-Related Morbidity and Mortality?

We identified 21 RCTs (reported in 27 articles[251-277]) comparing bisphosphonates (alendronate, ibandronate, risedronate, or zoledronic acid) with placebo and six RCTs (reported in 9 articles[278-286]) comparing denosumab with placebo that reported fracture, mortality, or both. Two RCTs of alendronate,[276, 277] two RCTs of zoledronic acid,[269-274] one RCT of ibandronate,[275] and two RCTs of denosumab[278, 286] were new to this update. Three RCTs were good quality;[251, 254, 255, 269-271] the rest were fair quality. Detailed study quality ratings are in **Appendix D Tables 53–57**. A summary of study characteristics is in **Table 2** with additional narrative description in **Appendix E.4**. One RCT of zoledronic acid[251] and one study of denosumab[278] were conducted exclusively in men; three studies (all evaluating bisphosphonates) included men, but the proportions comprised between 1 and 8 percent of the enrolled population.[295, 298, 299] The rest were conducted exclusively among postmenopausal women. T-score criteria for enrollment across studies varied, but only six required T-scores in the osteoporotic range. The rest enrolled participants with T-scores spanning the range considered low bone mass and osteoporosis or low bone mass only.

Detailed study characteristics are in **Appendix D Table 6**, and detailed study findings are in **Appendix D Table 16**.

Bisphosphonates: Findings

The findings from included trials evaluating the benefits of bisphosphonates compared with placebo for the outcomes of vertebral fractures, nonvertebral fractures, hip fractures, and mortality are summarized in this section and depicted in **Figure 7**. Findings were consistent for each outcome when alternative pooling methods or alternative doses other than FDA-approved doses were used (**Appendix E.4 Table 1**). One study of alendronate,[276] one study of zoledronic acid,[262] and one study of ibandronate[275] reported fractures other than vertebral, nonvertebral, and hip; these findings are reported in **Appendix D Table 16**.[262]

Vertebral Fracture

The impact of bisphosphonates on vertebral fracture outcomes reported in 10 trials is summarized in **Appendix E.4 Figure 1**.[251-254, 256, 258, 260, 261, 269, 277] These studies reported a mix of clinical vertebral fractures, radiographic vertebral fractures, or both. Five of these trials compared alendronate with placebo,[252-254, 256, 277] two compared risedronate with placebo,[258, 261] and three compared zoledronic acid with placebo.[251, 260, 269] The pooled RR for FDA-approved dosages was 0.51 (95% CI, 0.39 to 0.66; 10 RCTs; 9,015 participants; I^2=0%). This corresponds to an ARD of 18 fewer vertebral fractures per 1,000 participants treated (95% CI, from 23 fewer to 13 fewer). One study comparing alendronate with placebo showed a statistically significant reduction in vertebral fractures (1.9% vs. 3.5%; RR, 0.55 [95% CI, 0.38 to 0.80]),[254] and two studies comparing zoledronic acid with placebo showed a statistically significant reduction in vertebral fractures (1.5% vs. 4.6%; RR, 0.33 [95% CI, 0.16 to 0.70];[251] 2.3% vs. 4.9%; RR, 0.47 [95% CI, 0.29 to 0.76]).[269] Seven trials were not powered to evaluate vertebral fractures and individually found no statistically significant differences in reported vertebral fracture outcomes.[252, 253, 256, 258, 260, 261, 277] Five studies reported zero vertebral fracture events in at least one study arm.[252, 253, 258, 260, 261]

We conducted a sensitivity analysis based on type of vertebral fracture (**Appendix E.4 Table 2**). Four studies reported clinical vertebral fractures,[252, 253, 261, 269] three of which reported zero events in both study arms. The pooled RR for clinical vertebral fractures comparing treatment with placebo was 0.44 (95% CI, 0.24 to 0.79; 4 RCTs; 2,373 participants; I^2=0%). Six studies reported radiographic vertebral fractures,[251, 254, 256, 258, 260, 269, 277] two of which reported zero events in at least one study arm. The pooled RR for radiographic vertebral fractures comparing treatment with placebo was 0.51 (95% CI, 0.39 to 0.66; 7 RCTs; 8,642 participants; I^2=0%).

Nonvertebral Fracture

The impact of bisphosphonates on nonvertebral fracture outcomes reported in 13 trials is summarized in **Appendix E.4 Figure 2**.[251, 252, 254, 256-261, 268, 269, 272, 277] Six of these studies compared alendronate with placebo,[252, 254, 256, 259, 268, 277] three compared risedronate with placebo,[257, 258, 261] and four compared zoledronic acid with placebo.[251, 260, 269, 272] The pooled RR was 0.81 (95% CI, 0.74 to 0.88; 13 RCTs; 20,929 participants; I^2=0%). This corresponds to an ARD of 28 fewer nonvertebral fractures per 1,000 participants treated (95% CI, from 38 fewer to

18 fewer). Two studies reported zero events in at least one study arm.[252, 258] Eleven trials were not powered to evaluate nonvertebral fractures. Three trials individually reported a statistically significant benefit of active medication compared with placebo.[257, 259, 269] These studies included one evaluating alendronate (2.0% vs. 3.9%; RR, 0.52 [95% CI, 0.30 to 0.89]),[259] one evaluating risedronate (9.4% vs. 11.2%; RR, 0.84 [95% CI, 0.74 to 0.95]),[257] and one evaluating zoledronic acid (10.1% vs. 14.8%; RR, 0.68 [95% CI, 0.54 to 0.87]).[269]

Hip Fractures

The impact of bisphosphonates on hip fracture outcomes in six trials is summarized in **Appendix E.4 Figure 3**.[254, 256-259, 269] Three of these studies compared alendronate with placebo,[254, 256, 259] two compared risedronate with placebo,[257, 258] and one compared zoledronic acid with placebo.[269] We identified no trials of ibandronate that reported hip fracture outcomes. The pooled RR was 0.67 (95% CI, 0.45 to 1.00; 6 RCTs; 12,055 participants; I^2=0%). This corresponds to an ARD of 3 fewer hip fractures per 1,000 participants treated (95% CI, from 5 fewer to 0 fewer). One study reported zero events in both study arms.[258] None of the trials were powered to look at hip fractures as benefits, and none found statistically significant differences in reported hip fracture outcomes.

Mortality

The impact of bisphosphonates on mortality outcomes reported in six trials is summarized in **Appendix E.4 Figure 4**.[251, 264-266, 269, 275] Four of these studies compared ibandronate with placebo[264-266, 275] and two compared zoledronic acid with placebo.[251, 269] The pooled RR was 0.71 (95% CI, 0.49 to 1.05; 6 RCTs; 3,714 participants, I^2=0%). This corresponds to an ARD of 10 fewer deaths per 1,000 participants (95% CI, from 17 fewer to 2 more). Three studies reported zero events in at least one study arm.[264, 266, 275] None of the trials were powered to look at mortality as benefits, and none found statistically significant differences in mortality outcomes.

Bisphosphonates: Findings for Specific Populations of Interest

Only one trial of a bisphosphonate agent was conducted among men.[251] This trial reported on the effectiveness of zoledronic acid in 1,199 men with mean FN T-scores of -2.2. Men were eligible to participate if they had a T-score of -1.5 or less (based on the device-specific reference values for men). The authors found a reduced risk of morphometric vertebral fractures in the treatment arm (1.5% vs. 4.6%; RR, 0.33 [95% CI, 0.16 to 0.70]) but no significant difference in nonvertebral fractures (0.9% vs. 1.3%; RR 0.65 [95% CI, 0.21 to 1.97]).[251]

One study new to this update reported similar effectiveness of zoledronic acid compared with placebo among persons stratified by baseline BMD as well as when stratified by baseline fracture risk as measured by FRAX (hip and MOF) and the Garvan Fracture Risk Calculator.[269]

The study population of one large multicenter trial investigated the impact of risedronate on hip fractures in a study population with 41 percent of participants having a prior vertebral fracture at baseline. When including all participants in the study population, the pooled RR for bisphosphonates was 0.72 (95% CI, 0.58 to 0.91; 18,740 participants; I^2=0%). When including only participants ages 70 to 79 years without prior vertebral fracture, the pooled RR for bisphosphonates was 0.67 (95% CI, 0.45 to 1.00; 12,057 participants).[257]

Denosumab: Findings

The findings from included trials studying the benefits of denosumab compared with placebo are summarized in this section and include outcomes of vertebral fractures, nonvertebral fractures, hip fractures, and mortality (**Figure 7**). One trial was conducted exclusively in men;[278] the rest were conducted exclusively in postmenopausal women. Findings were consistent for each outcome when alternative pooling methods or alternative doses other than FDA-approved doses were used (**Appendix E.4 Table 3**).

Fractures

The impact of denosumab on fracture outcomes reported in five trials[278-280, 284, 285] are summarized in **Appendix E.4 Figure 5**. Four studies[278, 279, 284, 285] were not powered to look at fractures as outcomes, and events were rare in both study arms of these trials (range, 0 to 7 fracture events) such that the pooled RRs were dominated by results of the large FREEDOM trial.[278, 279, 284, 285] Authors of the FREEDOM trial (N=7,808) reported a statistically significant difference in incident radiographic vertebral fractures (2.3% vs. 7.2%; RR, 0.32 [95% CI, 0.26 to 0.41]), incident clinical vertebral fractures (0.8% vs. 2.5%; RR 0.31 [95% CI, 0.20 to 0.47]), nonvertebral fractures (6.1% vs. 7.5%; RR, 0.80 [95% CI, 0.67 to 0.95]), and hip fractures (0.7% vs. 1.1%; RR, 0.60 [95% CI, 0.37 to 0.97]).[280, 303] These correspond to an ARD of 48 fewer per 1,000 participants (95% CI, from 52 fewer to 42 fewer) for radiographic vertebral fractures, 17 fewer per 1,000 participants (95% CI, from 20 fewer to 13 fewer), 15 fewer per 1,000 participants (95% CI, from 24 fewer to 4 fewer) for nonvertebral fractures, and 4 fewer per 1,000 participants (95% CI, from 7 fewer to 0 fewer) for hip fractures. The FREEDOM study also reported significant reductions in multiple new vertebral fractures compared with placebo (see **Appendix D Table 16**).[280, 303]

We conducted a sensitivity analysis based on type of vertebral fracture (**Appendix E.4 Table 4**). Two studies reported clinical vertebral fractures,[278, 280] one of which reported zero events in the intervention arm. The pooled RR for clinical vertebral fractures was 0.31 (95% CI, 0.21 to 0.47; 7,635 participants; I^2=0%). One study evaluated radiographic vertebral fractures[285] but only reported one event in the placebo arm.

Mortality

Five trials reported mortality outcomes, but none were powered for this outcome.[280, 284-286] In the largest of the trials (FREEDOM, N=7,762 for this outcome) mortality was 1.8 percent in the denosumab arm compared with 2.3 percent in the placebo arm (calculated RR, 0.78 [95% CI, 0.57 to 1.06]).[280] Deaths were rare in the other four trials; one trial[285] reported zero deaths in the denosumab and placebo arms, and three trials[278, 284, 286] reported one death each in the denosumab arms. The pooled RR was 0.79 (95% CI, 0.58 to 1.07; 5 RCTs; 8,828 participants; I^2=0%) (**Appendix E.4 Figure 6**). This corresponds to an ARD of 4 fewer deaths per 1,000 participants treated (95% CI, from 9 fewer to 1 more).

Denosumab: Findings for Specific Populations of Interest

Authors of the FREEDOM trial reported on a preplanned analysis evaluating the effectiveness of denosumab as a function of baseline fracture risk.[280, 282] A linear model demonstrated no significant interaction between treatment effect and baseline fracture risk (p=0.72). However, analyses using a cubic spline function suggested a relationship (p<0.001). Compared with placebo, there was increasing efficacy of denosumab as baseline fracture risk increased between 5 percent and 18 percent with a leveling off (to slight decrease) in efficacy at baseline risks higher than 18 percent.

KQ 5. What Are the Harms Associated With Selected FDA-Approved Medications?

We identified 40 RCTs (reported in 48 articles[251-254, 256-269, 272-280, 284-304]) comparing bisphosphonates (alendronate, ibandronate, risedronate, or zoledronic acid) or denosumab with placebo that assessed harm outcomes. In addition, we identified three controlled cohort studies evaluating bisphosphonates compared with placebo.[305-307] We evaluated five RCTs as good quality;[251, 254, 269, 289, 290, 298, 300, 301] the rest of the RCTs and the controlled cohort studies were fair quality.

Bisphosphonates: Overview of the Evidence From RCTs

Thirty-four RCTs (published in 40 articles[251-254, 256-269, 272-277, 287-302]) reported on harms from bisphosphonates; four were new to this update.[275-277, 302] A summary of RCT characteristics is in **Table 2** with additional narrative description in **Appendix E.4**. Detailed study characteristics are in **Appendix D Table 6**, and detailed findings are in **Appendix D Table 16**.

Bisphosphonates: Findings From RCTs

The findings from included trials reporting the harms of bisphosphonates compared with placebo are summarized in this section, including discontinuations due to adverse events, serious adverse events, upper GI adverse events, and other rare harm outcomes (**Figure 7**). Findings were consistent for each outcome when alternative pooling methods or data from the non-FDA-approved doses were used (**Appendix E.4 Table 5**).

Discontinuations Due to Adverse Events

Twenty-seven RCTs reported discontinuations due to adverse events; however, none were powered for this outcome.[252-254, 256-261, 264, 266-268, 272, 275-277, 288, 291-294, 296-299, 302] Three RCTs reported only data for the intervention arm and thus could not be included in the pooled estimate.[253, 292, 293 272] The pooled RR was 1.00 (95% CI, 0.92 to 1.08; 25 RCT comparisons; 18,617 participants; I^2=0%; **Appendix E.4 Figure 7**). This corresponds to an ARD of 0 fewer discontinuations for adverse event per 1,000 participants treated (95% CI, from 9 fewer to 9 more). The two largest RCTs contributing to this pooled estimate were the Fracture Intervention Trial (FIT) study (N=4,432) comparing alendronate with placebo[254] and an international multicenter study comparing risedronate with placebo (N=9,331).[257] In FIT, discontinuations due to adverse events were 10.0 percent in the active drug group compared with 10.2 percent in the placebo group (RR, 0.98 [95% CI, 0.82 to 1.16]).[254, 289] In the risedronate trial, discontinuations due to adverse events were 17.7 percent in the active drug group compared with 18.0 percent in the placebo group (RR, 0.98 [95% CI, 0.89 to 1.10]).[257] One trial reported zero discontinuations due to adverse events in at least one study arm.[272]

Serious Adverse Events

Twenty-two RCTs reported serious adverse events; however, none were powered for this outcome.[251, 257, 259-261, 264, 266-268, 272, 275, 276, 287, 288, 291-293, 295-297, 299, 302] Two RCTs could not be included in the pooled estimate because authors did not report data for the control arms.[292, 293] The pooled RR was 0.97 (95% CI, 0.91 to 1.04; 21 RCT comparisons; 13,878 participants; I^2=0%; **Appendix E.4 Figure 8**). This corresponds to an ARD of 6 fewer serious adverse events per 1,000 participants treated (95% CI, from 18 fewer to 8 more). In the largest study contributing to this pooled estimate, an international multicenter RCT (N=9,331) comparing risedronate with placebo, serious adverse events occurred in 30.4 percent of the risedronate group and 31.0 percent of the placebo group.[257] Three RCTs reported zero events in both the placebo and active drug study arms.[266, 272, 295] The absolute incidence reported across this drug class was 0.7 to 25.3 percent across study arms, suggesting large variation in rigor of ascertainment methods across included studies.

GI Adverse Events

Twenty-six RCTs (representing 27 comparisons) reported GI adverse events.[252, 256-259, 261, 264, 266-269, 272, 275, 276, 289, 291-300, 302] None of the RCTs were powered for this outcome, and only one trial reported statistically significant differences in GI adverse events between the placebo and treatment arms.[272] In this trial, the outcome was described as "gastrointestinal acute phase reactions" reported by patients at 1 week postinfusion, potentially measuring a different outcome from the other RCTs. The pooled RR was 1.02 (95% CI, 0.98 to 1.06; 27 RCT comparisons; 22,280 participants; I^2=0%; **Appendix E.4 Figure 9**). This corresponds to an ARD of 5 more GI adverse events per 1,000 participants treated (95% CI, from 5 fewer to 16 more). The two largest RCTs contributing to this pooled estimate were the FIT study (N=4,432)[254] and the international, multicenter study of risedronate compared with placebo (N=9,331).[257] In FIT, the incidence of upper GI adverse events was 47.6 percent in the alendronate group and 46.2 percent in the placebo group (calculated RR, 1.03 [95% CI, 0.98 to 1.08]).[254, 289] In the study of risedronate, the incidence of upper GI adverse events was 21.2 percent in the risedronate group and 21.8 percent in the placebo group (calculated RR, 0.91 [95% CI, 0.88 to 1.07]).[257]

Cardiovascular Outcomes

Eight RCTs reported on one or more cardiovascular outcomes.[251, 254, 262, 269, 272, 287, 295, 300] Six RCTs reported on the incidence of atrial fibrillation.[251, 254, 262, 269, 272, 287] RR estimates ranged from 0.98 to 1.51; however, none were statistically significant. Furthermore, three of these RCTs reported zero events in both study arms.[262, 272, 287] Three RCTs reported on incidence of myocardial infarction.[251, 269, 300] RR estimates ranged from 0.61 to 4.68 and were very imprecise because of small sample sizes and rare events. The estimate for this harm was statistically significant in the study comparing zoledronic acid with placebo (RR 4.68 [95% CI, 1.02 to 21.5]) in men but was not statistically significant in the other two RCTs. One trial reported multiple other cardiac outcomes (stroke, transient ischemic attack, cardiac deaths) all of which were nonsignificant and imprecise.[269]

Osteonecrosis of the Jaw

Five RCTs, including two new to this update,[269, 272] reported no cases of osteonecrosis of the jaw.[251, 262, 287] Additional information about this rare outcome from studies not eligible for inclusion is addressed as **Contextual Question 6** in **Appendix F**.

Atypical Fractures of the Femur

We did not identify any RCTs that reported on the rare outcome of atypical femur fracture. Additional information about this rare outcome from studies not eligible for inclusion is addressed as **Contextual Question 6** in **Appendix F.3**.

Bisphosphonates: Evidence From Controlled Cohort Studies

Three fair-quality cohort studies set in Denmark,[305] Sweden and Denmark,[306] and South Korea[307] addressed potential harms of bisphosphonate use. Two studies were limited to new users;[305, 307] the third study provided sensitivity analyses for a treatment-naïve cohort.[306] The studies predominantly (86%[306] and 91%[307]) or solely comprised women.[305] Two studies did not report the prevalence of fractures among participants at the start of the study;[306, 307] one study reported differences in prevalence of fractures at baseline (12% for alendronate vs. 4% for nonusers).[305] One study was limited to zoledronic acid,[306] a second to alendronate,[305] and the third included all bisphosphonates (which may have included non-FDA-approved bisphosphonates).[307] Detailed study characteristics are in **Appendix D Table 7** and detailed findings are in **Appendix D Table 17**. Study quality ratings are in **Appendix D Tables 58–65**.

GI Cancers

One fair-quality controlled cohort study set in Denmark[305] reported on the incidence of GI cancers, specifically colon cancer[305] among women newly exposed to alendronate when compared with matched nonuser controls over 5 years of followup. The study reported a lower risk of developing colon cancer in new alendronate users when compared with matched nonusers of alendronate (aHR, 0.69 [95% CI, 0.60 to 0.79]).[305]

Cardiovascular Outcomes

One fair-quality controlled cohort study set in Sweden and Denmark[306] reported on cardiovascular outcomes. A propensity-score matched cohort of treatment-naïve users of zoledronic acid compared with nonusers in Sweden and Denmark reported no statistically significant differences in atrial fibrillation (aHR, 1.18 [95% CI, 0.99 to 1.40]), myocardial infarction (aHR, 0.92 [95% CI, 0.64 to 1.31]), and cardiovascular mortality (aHR, 0.97 [95% CI, 0.81 to 1.15]) but did find a statistically significant increased risk for heart failure (aHR, 1.32 [95% CI, 1.08 to 1.61]). This study did not control for known confounders of heart failure such as BMI, smoking and alcohol exposure, hypertension, and metabolic syndrome. It is possible that the zoledronic acid users may have had a higher inherent risk of heart failure.[306]

Atypical Femur Fractures

Two fair-quality controlled cohort studies set in Sweden and Denmark[306] and South Korea[307] consistently reported increased risk of atypical femur fractures with bisphosphonate exposure. The propensity-score matched cohort of new users of zoledronic acid compared with nonusers in Sweden and Denmark reported an increased risk of atypical femur fractures (aHR, 2.46 [95% CI, 1.17 to 5.15]). However, this study could not control for known confounders of fracture such as baseline levels of calcium and vitamin D, bone density, BMI, smoking and alcohol exposure, hypertension, and metabolic syndrome and could not rule out that zoledronic acid users may have higher inherent risks of frailty. The South Korean study of new bisphosphonate users reported an increased risk of atypical femur fractures with bisphosphonate use (aHR, 1.53 [95% CI, 1.36 to 1.73]) over a mean of 1 year followup when compared with matched bisphosphonate nonuser control participants.[307] The study did not adjust for confounders other than age, sex, systemic use of glucocorticoids, and comorbidity and may have included drugs not approved by the FDA for osteoporosis.

Denosumab: Overview of the Evidence

We identified six fair-quality RCTs (published in 8 articles[278-280, 284-286, 303, 304]) that assessed the harms of denosumab compared with placebo (**Figure 7**); two were new to this update.[278, 286] A summary of RCT characteristics is in **Table 2** with additional narrative description in **Appendix E.4**. Detailed study characteristics are in **Appendix D Table 6**, and detailed findings are in **Appendix D Table 16**. Findings were consistent for each outcome when alternative pooling methods or data from the non-FDA-approved doses were used (**Appendix E.4 Table 6**).

Denosumab: Findings

Discontinuation Due to Adverse Events

Five RCTs[278, 280, 284-286] reported discontinuations due to adverse events. However, none of the studies were powered for this outcome.[286] The pooled RR was 1.16 (95% CI, 0.87 to 1.54; 5 RCTs; 8,826 participants; I^2=0%; **Appendix E.4 Figure 10**). This corresponds to an ARD of 3 more discontinuations per 1,000 participants treated (95% CI, from 3 fewer to 11 more). This pooled estimate was mostly influenced by the large FREEDOM study (N=7,762) where the incidence of discontinuations due to adverse events was 2.4 percent in the denosumab arm compared with 2.1 percent in the placebo arm (calculated RR, 1.15 [95% CI, 0.85 to 1.54]).[280, 303]

Serious Adverse Events

Six RCTs[278-280, 284-286] reported serious adverse events; however, none of the studies were powered for this outcome. The pooled RR was 1.04 (95% CI, 0.97 to 1.12; 5 RCTs; 8,934 participants; I^2=0%; **Appendix E.4 Figure 11**). This corresponds to an ARD of 9 more serious adverse events per 1,000 participants treated (95% CI, from 7 fewer to 28 more). This pooled estimate was mostly influenced by the large FREEDOM study (N=7,762) where the incidence of serious adverse events was 25.8 percent in the denosumab group and 25.1 percent in the placebo group (calculated RR, 1.03 [95% CI, 0.95 to 1.11]).[280, 303]

Upper GI Adverse Events

Four RCTs reported upper GI adverse events; however, none of these studies were powered for this outcome.[279, 284-286] Events were rare across all study groups, including two RCTs with zero events in the placebo arm.[284, 285] The pooled RR was 2.18 (95% CI, 0.74 to 6.46; 4 RCTs; 932 participants; I^2=0%; **Appendix E.4 Figure 12**). This corresponds to an ARD of 14 more GI adverse events per 1,000 participants treated (95% CI, from 3 fewer to 66 more).

Cardiovascular Outcomes

Two RCTs reported cardiovascular outcomes.[280, 284, 303] In the large FREEDOM study, authors reported no significant difference in cardiovascular events (calculated RR, 1.04 [95% CI, 0.85 to 1.27]).[280, 303] A second trial reported no difference in "cardiac disorders," but events were rare and estimates imprecise (calculated RR, 0.45 [95% CI, 0.02 to 10.83]).[284]

Osteonecrosis of the Jaw

Three RCTs reported on the rare outcome of osteonecrosis of the jaw.[278, 280, 286] Zero events were reported in all studies, one of which was the large FREEDOM study.[280, 303] Additional information about this rare outcome from studies not eligible for inclusion is addressed as **Contextual Question 6** in **Appendix F.3**.

Atypical Femur Fracture

Two RCTs, new to this update, reported on the rare outcome of atypical femur fracture.[278, 286] Zero events occurred in both studies. Additional information about this rare outcome from studies not eligible for inclusion, such as the FREEDOM long-term extension study, is addressed as **Contextual Question 6** in **Appendix F.3**.

Rebound Vertebral Fractures

No studies that were included for KQ 5 had study designs sufficient to evaluate the outcome of rebound vertebral fractures after denosumab discontinuation. We describe findings for this outcome from studies not eligible for inclusion in this update as **Contextual Question 7** in **Appendix F.4**.

Other Adverse Events

Three RCTs reported additional harm outcomes related to skin disease and infection.[280, 284, 285] In the FREEDOM RCT, a higher incidence of eczema was observed in the denosumab arm compared with placebo (RR, 1.81 [95% CI, 1.34 to 2.44]), and a higher risk for serious skin infection was also observed but was imprecise (RR, 15.0 [95% CI, 1.98 to 113.2]).[280] There was no difference in the risk of serious infections (RR, 1.19 [95% CI, 0.95 to 1.49]).[280] Another RCT also reported a higher incidence of rash (calculated RR, 2.82 [95% CI, 1.04 to 7.64]) and serious infection (calculated RR, 8.1 [95% CI, 1.02 to 63.6]).[285] A third study reported no difference in serious infection (calculated RR, 3.5 [95% CI, 0.07 to 190.8]).[284]

Chapter 4. Discussion

Summary of Evidence

Table 3 summarizes the evidence synthesized in this report by KQ and provides our EPC's assessment of the strength of evidence (SOE) and applicability. Compared with the prior review on this topic,[3] our certainty as reflected in our SOE ratings has increased as a result of new direct evidence for KQ 1. Whereas our previous SOE rating for KQ 1 was insufficient for mortality and fracture outcomes except hip (which was rated as low SOE for benefit in the prior review), in this update we rated MOF and hip fracture outcomes as moderate SOE for benefit, osteoporotic fractures as low SOE for benefit, and mortality as low for no effect. We continue to grade the direct evidence as insufficient for harms of screening (KQ 3) but have identified additional data on overdiagnosis for consideration compared with the prior review.

We identified some new evidence related to treatment benefits (KQ 4) and harms (KQ 5) in this update. Our SOE ratings for treatment benefits (KQ 4) remained largely the same as the prior review: low to moderate SOE for benefit across multiple fracture outcomes for both bisphosphonates and denosumab. For treatment harms (KQ 5), we graded the evidence for each outcome separately as compared to the prior report; with low (denosumab) to moderate (bisphosphonates) SOE for both discontinuations due to adverse events and serious adverse events and moderate SOE for no effect on upper GI adverse events for bisphosphonates and low SOE for increased upper GI adverse events for denosumab. As in the prior report, we note that the evidence included for KQ 5 is not sufficient for evaluating the effect of treatment on very rare harms such as osteonecrosis of the jaw, atypical femur fractures, rebound vertebral fractures, or harms that may emerge after prolonged duration of treatment.

The scope of the KQs on accuracy changed between the prior report and the current update so direct SOE comparisons are not possible. Further, in this update we rated SOE for specific instruments and among subpopulations wherever possible, further limiting a direct comparison with the prior review's SOE ratings.

Benefits and Harms of Screening (KQs 1 and 3)

For this update, we included three trials (ROSE, SOS, and SCOOP) providing direct evidence for screening. All studies were pragmatic in nature, relying on participants' PCPs to initiate further evaluation and treatment in response to positive screening tests. As with most trials of screening, the proportion of participants who received treatment was a relatively small proportion of those randomized. We found moderate SOE for a small absolute benefit of screening on hip fractures (5 fewer per 1,000 screened) and MOF (6 fewer per 1,000 screened) and low SOE for osteoporotic fractures (5 fewer per 1,000 screened). The absolute magnitude of benefit observed is similar to that observed for hip fracture prevention from treatment with bisphosphonates or denosumab in persons with known osteoporosis, but smaller than the benefit observed for vertebral or nonvertebral fracture prevention. These estimates are based on pooling with the per-protocol population from the ROSE trial, which is methodologically most similar to the study designs used in SCOOP and SOS. Our sensitivity analysis using the ROSE ITT population (**Appendix E.1 Figure 1**) predictably led to smaller estimates of absolute effect and that do not exclude a null effect because of the large proportion of participants (nearly 40%) who were included in the analysis without receiving the intervention.

The only individually statistically significant fracture reduction outcome was for hip fractures (a secondary endpoint) in the SCOOP trial. This finding was unexpected given that hip fractures are a subset of MOF and are much rarer events than other fractures. The study authors suggested that because they used the 10-year estimated hip fracture risk to determine recommendations for DXA, they were perhaps preferentially targeting persons more likely to suffer hip fractures than other fracture types. The hip fracture outcome may be spurious or biased because the relative magnitude of effect is inconsistent with findings for the other fracture outcomes, which occurred with much higher frequency. It is also a relatively large relative reduction, considering few participants were actually treated with medication. However, the authors reported a post hoc analysis in which only participants with the highest percentile of FRAX baseline hip probability benefited from screening,[122] and findings from the ROSE trial also suggested that most of the benefit with respect to MOF could be attributed to reductions in hip fracture.

All three trials enrolled individuals at high risk for fracture. The SOS trial enrolled a higher risk population (43% had prior fractures) than ROSE and SCOOP and conducted a more extensive battery of tests as its screening intervention. Further, the populations in all three studies were likely at higher risk of fracture than an average screening population in the United States. For example, a 65-year-old White woman in the United States of average height (159.7 cm) and weight (75.6 kg) based on 2015–2018 NHANES data[315] with no additional clinical risk factors has a 10-year risk of MOF of 8.2 percent and a hip fracture risk of 1.0 percent according to FRAX (without BMD input).[316] These risks are well below the mean FRAX-estimated risks in the SCOOP, ROSE, and SOS study populations (MOF risks ranged from 19% to 24.6%; hip fracture risks ranged from 6.7% to 11.6%). If one considers that individuals with the risk factors of glucocorticoid use, rheumatoid arthritis, secondary osteoporosis, or prior fracture are not in the target population for screening (i.e., DXA testing would be indicated for these individuals as part of disease management), then the highest estimated risk possible for a 65-year-old White woman of average height and weight and unknown BMD without such risks but with all other FRAX-specified risks (i.e., smoking, alcohol use, parental hip fracture) is 19 percent for MOF and 2.9 percent for hip fracture. Those risks increase slightly for women with lower BMIs and decrease slightly for those with higher BMIs. For a 65-year-old Black woman with the same height, weight, and risks, the highest possible risk is 8.9 percent (MOF) and 1.3 percent (hip), which is well below the risk of women in the included trials. One of the SRs included for KQ 1 also reviewed the acceptability of screening by patients and reported women who are low risk based on age have a high intention of getting screened; however, no studies report on the intentions of higher-risk women.[131] The ROSE trial authors analyzed subjects who declined DXA testing and reported a higher level of comorbidities and health behaviors that also portend a higher fracture risk, suggesting a selection bias toward healthy individuals.[128] Thus, achieving population-level benefits of screening likely requires implementation strategies to ensure it is reaching those at highest risk.

We judged the SOE as low for no effect on mortality because of imprecision and study limitations. Only one trial reported on a single harm outcome (anxiety);[120] no differences were observed between groups. We judged the strength of evidence for these anxiety harms as insufficient because of study limitations related to modest uptake and adherence, and because of a single study body of evidence. One of the included SRs reported estimates for overdiagnosis of between 118 and 241 per 1,000 women screened. We assessed the SOE for overdiagnosis as insufficient, primarily because of study limitations in the underlying RCTs included in the SR and evolving methods for estimating this harm, which involves extrapolation. Overdiagnosis for identifying a high-risk probability is conceptually different than overdiagnosis of overt conditions (e.g., cancer), and the exact methods to estimate overdiagnosis in this context are still evolving and will generally be limited by less than perfect calibration of risk prediction instruments.[314]

Another consideration is the applicability of the screening interventions used in these trials. The 2018 USPSTF strategy recommends universal BMD assessment in women age 65 years or older and a two-staged approach (formal risk assessment followed by BMD) for postmenopausal women younger than age 65 years. A two-staged approach was used with FRAX in SCOOP and ROSE for women of all ages; however, country-specific FRAX prediction models were used with thresholds unique to each study. SCOOP used an age-dependent hip fracture risk threshold to offer DXA, which varied from 5.2 to 8.4 percent, whereas ROSE offered DXA to participants above a 15 percent MOF risk threshold, regardless of age. If a two-stage approach were replicated in the United States, it is not entirely clear what thresholds should be used and whether thresholds should be fixed or vary based on age or other factors. It is also not clear how patient values and preferences about getting screened should be incorporated. The implications of using fixed risk thresholds vs. age-dependent thresholds are addressed further in **Contextual Question 5** in **Appendix F.2**. In brief, fixed thresholds may result in over or under screening or treatment while age-dependent thresholds may be difficult to use in practice. In contrast to SCOOP and ROSE, the SOS trial used an intensive intervention consisting of DXA, vertebral fracture assessment (imaging test), FRAX (country specific), fall risk assessment, and laboratory evaluation to evaluate for secondary causes of osteoporosis for women allocated to screening. Whether such an intensive intervention is feasible in usual primary care settings in the United States is not clear, nor is it clear whether the intensity of the intervention is warranted because this intervention had a similar magnitude of benefit compared with the less intensive interventions used in SCOOP and ROSE.

Accuracy (KQ 2)

Although this update includes more direct evidence for the benefits of screening compared with the prior review for older postmenopausal women, it may still be useful to consider the indirect evidence pathway for screening given the limitations and applicability of the direct evidence and because direct evidence for men and younger postmenopausal women is lacking entirely. Currently the USPSTF recommends universal DXA testing in women beginning at age 65 years, without regard to clinical risks. Because most fragility fractures occur in persons without osteoporosis, accurate risk assessment instruments could help identify the highest risk persons for subsequent risk reduction treatment, including but not limited to pharmacotherapy. However, because pharmacotherapy trials to date have not enrolled persons based on fracture risk, the role of such instruments with respect to decisions about DXA screening and treatment remains unclear.

Predictive Accuracy

The evidence for predictive accuracy of risk assessments and BMD measurement was very heterogeneous; further, it was poor methodologic quality. This poor quality partly reflects increased rigor of design and reporting standards for prognostic studies in recent years. The predictive accuracy of some risk assessment instruments (KQ 2a) appears to be similar to that of BMD alone (KQ 2b). Although many measures of accuracy exist, most studies reported discrimination measures only and specifically AUC.

We rated the SOE for the predictive accuracy of risk assessment instruments for discrimination as either low (FRAX, FRC, Garvan, QFracture) or insufficient (OST, WHI) and for calibration as low for FRAX and insufficient for all others evaluated. Accuracy appears higher for instruments that can incorporate a BMD input (FRAX, Garvan, FRC); however, this may not be particularly useful when considering such instruments as the first assessment step for determining who to refer for further DXA testing in a two-stage screening approach. Thus, for USPSTF consideration, our findings related to instruments without a BMD input are likely the most applicable to decision making.

A particular challenge to using risk assessment instruments in practice is determining the risk threshold to apply for clinical action. The evidence in this update suggests that multiple instruments can reasonably predict MOF or hip fractures at various thresholds, but the inconsistency of findings across the evidence base limits a strong conclusion about the use of a specific instrument at a specific threshold. Commonly applied thresholds (3% for hip fracture risk and 20% for MOF risk) were derived as thresholds for considering treatment (not screening) and were based on a cost-effectiveness analysis.[80, 85] We note that the predictive accuracy of more complex risk assessment instruments involving multiple clinical or demographic risks appears to be similar to the accuracy of simpler assessments with fewer risks.

We rated the SOE for the predictive accuracy of BMD alone as low for discrimination outcomes because results were inconsistent and study quality was poor, and we rated the SOE for calibration outcomes as insufficient because the evaluation of BMD alone as a predictor was not the primary study aim for any of these studies, so authors reported limited calibration information. Discriminative accuracy varied widely among the different cohorts when considering BMD as a continuous measure; it appears better for hip prediction than for MOF, which could be explained by the fact that FN was the site most often used for measuring BMD, but also because MOF is a heterogenous outcome compared with hip fracture. Predictive accuracy in men appears to be similar or better than in women, though we note that the men enrolled in studies of accuracy may not be generalizable to the general primary care population as they may have been identified from referrals for BMD testing, specialty care, or primary care clinics caring for medically complex patients, such as VA settings. BMD alone is already used in practice for clinical decision making related to treatment; however, the evidence in this update confirms that the T-score threshold defining osteoporosis (<-2.5) is not very sensitive and is only modestly specific for predicting future fragility fracture. This appears particularly true among younger women, but the evidence for this group is limited.

Diagnostic Accuracy

Given that the evidence base for pharmacotherapy is based on treating persons with osteoporosis or low bone mass, the accuracy of risk assessment instruments to identify which persons are likely to be candidates based on BMD is critical. Like the evidence base for predictive accuracy, the evidence base for the diagnostic accuracy of various risk assessment instruments (KQ 2c) was very heterogeneous in terms of populations evaluated, reference standards used, and score thresholds evaluated. Accuracy as evaluated by AUC was modest to good, but sensitivity and specificity ranges within and across studies were wide. We rated the SOE for discrimination outcomes for FRAX, OST, and OSTA as low in both women and men. For women, we rated the SOE as low for ABONE, NOF, ORAI, OSIRIS, and SCORE and insufficient for AMMEB, Garvan, and SOFSURF. For men, we rated the SOE as low for MORES and MOST and insufficient for ABONE, Garvan, MSCORE, ORAI, OSIRIS, SCORE, and VA-FARA. To be used in clinical practice, thresholds need to be established to determine a positive screening test. Many studies evaluated alternative thresholds than the ones established in development cohorts to optimize sensitivity in their population. In some cases, the alternative thresholds may have involved a slight tweak to the score threshold, but in other studies it may have involved a much larger adjustment to the threshold. A test whose threshold is not robust across a spectrum of populations may not be suitable for widespread use.

Repeat Screening

We did not identify any direct evidence (KQ 1) evaluating a strategy of repeat screening. As part of our assessment of the indirect evidence (KQ 2d), we did identify studies comparing the predictive accuracy of repeat screening with DXA BMD after 4 to 8 years with a baseline DXA BMD and we rated the SOE as moderate for similar predictive accuracy (**Table 3**). Further evidence from contextual question 1 (**Appendix F.1**) provides evidence from studies evaluating the time taken for women to transition across various BMD categories. In an analysis of the SOF cohort of postmenopausal women age 65 years or older, it took on average 17 years for 10 percent of women with a normal BMD at baseline to transition to an osteoporotic range, and a similar figure was observed for women with T-scores between -1.0 and -1.49.[317] The transition interval decreased for women with T-scores between -1.50 and -1.99 (4.7 years) at baseline and women with T-scores between -2.0 and -2.49 (1.1 years) at baseline.[317] Several other studies have attempted to identify optimal screening intervals by assessing the time to transition to osteoporosis or a 10 percent fracture risk and the time for 1 percent of women to transition to an actual fracture event. These authors estimate various intervals, but a pattern of shorter intervals with advancing age is consistent across studies.

Considerations Regarding Race-Based Prediction Models

Of special note are the findings related to using FRAX in women younger than age 65 years with the MOF risk thresholds suggested by the USPSTF's current (8.4%) or past (9.3%) recommendation. Analyses of this threshold in three unique cohorts suggested poor sensitivity and only modest specificity. A further concern with these thresholds is that they represent the risks for a 65-year-old White woman of average height and weight. This benchmark was selected by the USPSTF based on its existing recommendation that women age 65 years or older should be screened with DXA. However, if the goal is to use the risk of a 65-year-old woman with no other clinical risks as a benchmark, then the risk for any given individual may be more fairly evaluated against the risk of a 65-year-old of the same height, weight, and race. For example, the estimated risk for a younger Black woman could be evaluated against the risk for a 65-year-old Black woman (equivalent to 4.2% for BMI of 25.0 kg/m^2). Such an approach would solve issues inherent in using a fixed threshold based on risk of a White woman; however, across the population, it would result in persons of varying fracture risks being referred for DXA. Further, this approach is likely not feasible to implement in time-constrained primary care settings. The issue related to referral of persons with varying fracture risk is already inherent in the recommendation for universal DXA screening in women at age 65 years. MOF risks for women with a BMI of 25 kg/m^2 vary between a low of 4.2 percent (Black woman, no other clinical risk) and 22 percent (White woman with parental history, smoking, and alcohol use). See **Appendix F.2 Contextual Question 5** for additional information concerning use of fixed thresholds with risk assessment tools.

For further consideration is whether a race-based prediction model should be used for clinical decision making at all. The United States is one of only a few countries that use FRAX calibrated for specific racial groups; other countries with multiethnic populations (e.g., Canada, United Kingdom) use one race-neutral FRAX calculator. Race-informed prediction models do not accommodate multiracial individuals or the underlying heterogeneity in risk that exists within a single racial group. Other race-neutral prediction models with fewer inputs appear to be as accurate or more accurate than FRAX or similarly complex assessments. However, challenges remain as to what the appropriate threshold for decision making would be for these instruments and whether it should vary for different populations or clinical contexts.

Benefits and Harms of Treatment

Treatment of osteoporosis is well established in clinical practice. We found moderate SOE for treatment with bisphosphonates for the primary prevention of vertebral and nonvertebral fractures and low SOE for benefit on hip fractures and mortality. We analyzed studies of bisphosphonates as a class; however, it is important to note that not all drugs in the class have demonstrated efficacy with respect to hip fracture outcomes. We found moderate SOE for denosumab with respect to the primary prevention of vertebral fractures and low SOE for nonvertebral fractures, hip fractures, and mortality. An SR and network meta-analysis[318] in support of the January 2023 clinical practice guideline on pharmacotherapy issued by the American College of Physicians[319] found similar conclusions regarding bisphosphonate and denosumab treatment with respect to fracture outcomes; however, this review included a broader scope that did not limit to primary prevention populations and did not exclude poor-quality studies. We rated serious adverse events and discontinuations due to adverse event outcomes as either low or moderate SOE for no effect comparing active drug (bisphosphonate or denosumab) and placebo. For upper GI adverse events, we rated the evidence as moderate SOE for no effect for bisphosphonates and low for harm for denosumab.

We identified several applicability concerns with this body of evidence. Of the studies that reported fracture or mortality outcomes, a minority specifically required T-scores less than -2.5; the rest enrolled participants with T-scores spanning the range considered low bone mass and osteoporosis or low bone mass only. All but one study of denosumab and one study of bisphosphonates were conducted in postmenopausal women. Although studies of abaloparatide, teriparatide, and romosozumab were eligible for inclusion for men, we did not identify any such studies because studies of those agents were conducted among populations with prior fracture or with secondary osteoporosis; these populations were not eligible for inclusion in this update review. For this update, we identified only one study published since the prior review (the other five studies newly included in this update were published between 1996 and 2012 and we identified them through handsearches of systematic reviews identified in the current update). Because treatment is now standard of care for osteoporosis, we think future placebo-controlled trials are unlikely. Future updates for this topic may want to consider the treatment benefit and harm evidence as foundational.[318]

Limitations of the Evidence

We note several limitations with the three trials included for KQ 1. First, they were all pragmatic trials conducted among older European women (median ages, 71 to 76 years) using interventions that included non-U.S. fracture prediction models. The proportion of persons eligible who participated was low (about one-third) in one trial[120] with evidence of healthy selection bias, and the receipt of the screening intervention was suboptimal in the other two trials (55% in ROSE[126] and 76% in SOS[124]). The three trials were underpowered because the observed proportion of women with treatment indications and who adhered to treatment were lower than expected and because of contamination in control arms from secular trends in screening and treatment. For all these reasons, the estimate of benefits from these trials probably represents the lower bounds of screening efficacy for the eligible population. Yet, these findings may reflect the real-world effectiveness of a systematic screening program. It is not clear whether similar findings would be observed if screening were offered entirely through the participant's PCP office, which is a model more applicable to USPSTF considerations. Although these estimates represent the lower bounds of efficacy, it is not entirely clear that the findings are applicable to populations with lower fracture risk or U.S. settings given the use of country-specific FRAX prediction models and the thresholds for action (i.e., further screening with DXA or treatment) used in these trials. As described earlier, the women in these trials represented a population with a higher risk than we might expect to encounter in a primary prevention population of women with a screening indication based on age alone. Whether it is possible to conduct a large-scale trial of screening among women age 65 years or older in the United States given that universal screening with DXA is a common practice is not clear.

Although we identified many studies for the KQ on accuracy, heterogeneity in populations, thresholds used, and incomplete reporting precluded robust conclusions. For both predictive and diagnostic accuracy, a number of studies were conducted using retrospectively assembled datasets of persons referred for BMD, some of whom may already have a diagnosis of osteoporosis or take medication or may have had a prior fracture. Many predictive accuracy studies focused only on discrimination outcomes and did not report sufficient information about calibration. Some used proxy data for selected risk factors or omitted those factors if data were not available, or participants were observed for fewer years than the duration used in the risk model development studies. Further, it is unclear whether data on FRAX from other countries is applicable to the U.S. setting given that FRAX is calibrated to each country's fracture incidence. We tried to mitigate this issue by limiting the KQs on predictive accuracy to countries with similar hip fracture incidence as the United States (moderate incidence). The diagnostic accuracy studies varied in how the DXA reference standard was measured (e.g., different anatomic sites for T-score, different reference range used to calculate T-scores from raw BMD measure).

The major limitation in the treatment literature for primary prevention is that few studies include men, and studies enrolled persons based on BMD T-scores and not fracture risk. Although data suggest treatments are probably safe compared with placebo, few studies eligible for this update review were sufficiently designed to report on rare or duration-dependent harms such as osteonecrosis of the jaw, atypical femur fractures, or rebound vertebral fractures.

A concern across the evidence for all KQs relates to the lack of diverse populations enrolled in studies. Many studies did not report the race or ethnicity of enrolled populations, and those that did mostly enrolled exclusively or vast majority White populations. Given the differences in fracture incidence among persons of different races and ethnicities in the United States, studies enrolling sufficient numbers from diverse populations are needed to determine the applicability of findings in different populations.

Future Research Needs

Because the predictive accuracy of most risk assessment instruments is similar to that of BMD alone, trials that randomize participants to fracture risk assessment vs. DXA for screening and then treat based on fracture risk or T-score would provide direct evidence for comparing such screening strategies and would address a gap about whether pharmacotherapy based on fracture risk is effective for reducing fractures. It is not clear what screening strategy should be evaluated in such trials, including whether the focus should be on identifying osteoporosis to treat with medication or whether a more comprehensive screening strategy to address a broader set of fracture determinants should be evaluated. Further, if a country-specific risk assessment tool, such as FRAX, is used, then trials conducted in the United States using the U.S. version of FRAX would ensure applicability of findings to U.S. settings. In addition, future trials of screening should specify harm outcomes a priori and use adequate ascertainment methods. In the absence of future trials offering direct evidence, decision analyses could help fill in gaps regarding optimal starting and stopping stages for women or identifying optimal screening approaches; however, it is not clear whether enough screening trial evidence exists for robust inputs into such analyses.

Rigorously designed research on instruments for fracture risk prediction or osteoporosis identification that are applicable to general, unselected primary care populations and that are feasible for use in such settings is needed. Thoughtful consideration for whether and how race is used in such instruments is critical as is research associated with selecting thresholds for action resulting from the use of such instruments in practice. Whether the focus of future research should be on improving existing instruments, such as with the addition of fall history or propensity, on developing new instruments, or on improving provider and patient understanding and decision making from the use of current instrument is unclear.

Given that treatment of osteoporosis in older, screen-detected women without contraindications is considered standard of care, it is unlikely that future placebo-controlled trials of treatment in such populations will be conducted for ethical reasons. However, research that evaluates treatment of osteoporosis among screen-detected men and younger women without known clinical risks would likely have equipoise. To date, most studies that have enrolled men or younger women focus on persons with a history of prior fracture or who have underlying medical conditions or take medications associated with secondary osteoporosis.

Our search of trial registries identified three ongoing studies (**Appendix H**) but none that appear to address the specific research needs described in this section.

Limitations of the Review

This review focused on only one aspect of fracture prevention, which was to identify and pharmacologically treat osteoporosis. We did not evaluate comprehensive approaches to fracture prevention that might include screening, counseling, medication, physical therapy, and other interventions to prevent falls or improve physical function in older adults. Preventing falls is addressed by a separate USPSTF recommendation.[320]

This review did not address the use of DXA testing as part of disease management in persons with a history of fragility fracture or medical conditions or medications associated with secondary osteoporosis. DXA testing in such persons is clinically indicated along with other medical tests or interventions for risk mitigation. Thus, we do not consider DXA testing in such individuals as screening, so results from this review cannot be applied to such populations.

We did not evaluate the comparative effectiveness and harms of alternative pharmacotherapies, and we did not evaluate evidence concerning duration of treatment or temporary drug holidays. For treatment benefits and harms, we focused on studies for primary prevention and did not include trials conducted predominantly amond persons with secondary osteoporosis or history of fragility fracture. Our review scope was not comprehensive for evaluting rare harms of treatment; several authors have reported on these harms using study designs broader than what we used for the KQs in this update (**Appendix F, Contextual Questions 6 and 7**).

Our review was limited to English-language publications published in peer-reviewed journals and conducted in very highly developed countries. We did not include conference abstracts or data from completed but unpublished studies posted in trial registries.

Conclusions

Screening in older, higher-risk women was associated with a small absolute risk reduction in hip fractures and MOF compared with usual care. Screening strategies varied and no direct evidence evaluated screening using DXA alone or screening in women younger than age 65 years or in men. Risk assessment instruments and BMD at the hip or spine has poor to modest discrimination in men and older women for predicting fracture and studies of calibration were limited. For identifying osteoporosis, risk assessment instruments had modest to good accuracy in men and modest accuracy in older women. In women younger than age 65 years, risk assessment instruments had poor predictive (fracture) and diagnostic (osteoporosis) discrimination. Treatment of osteoporosis with FDA-approved bisphosphonates or denosumab was associated with reductions in vertebral, nonvertebral, and hip fractures with no increase in discontinuations due to adverse events or serious adverse events compared with placebo in studies conducted over one to several years' duration; however, data about rare and longer-term harms were limited from the evidence included in this update.

References

1. Curry SJ, Krist AH, Owens DK, et al. Screening for osteoporosis to prevent fractures: US Preventive Services Task Force recommendation statement. *Jama*. 2018 Jun 26;319(24):2521-31. doi: 10.1001/jama.2018.7498. PMID: 29946735.
2. U.S. Preventive Services Task Force. Screening for osteoporosis: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2011 Mar 1;154(5):356-64. doi: 10.7326/0003-4819-154-5-201103010-00307. PMID: 21242341.
3. Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: an evidence review for the U.S. Preventive Services Task Force. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Report No.: 15-05226-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; June. 2018.
4. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med*. 1993 Jun;94(6):646-50. doi: 10.1016/0002-9343(93)90218-e. PMID: 8506892.
5. Kanis JA, Melton LJ, 3rd, Christiansen C, et al. The diagnosis of osteoporosis. *Journal of Bone & Mineral Research*. 1994 Aug;9(8):1137-41. doi: 10.1002/jbmr.5650090802. PMID: 7976495.
6. Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. *Osteoporosis international*. 1994;4(6):368-81.
7. Looker AC, Wahner HW, Dunn WL, et al. Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int*. 1998;8(5):468-89. doi: 10.1007/s001980050093. PMID: 9850356.
8. Binkley N, Adler R, Bilezikian JP. Osteoporosis diagnosis in men: the T-score controversy revisited. *Curr Osteoporos Rep*. 2014 Dec;12(4):403-9. doi: 10.1007/s11914-014-0242-z. PMID: 25255867.
9. De Laet CE, van Hout BA, Burger H, et al. Bone density and risk of hip fracture in men and women: cross sectional analysis. *BMJ*. 1997 Jul 26;315(7102):221-5. doi: 10.1136/bmj.315.7102.221. PMID: 9253270.
10. Alswat KA. Gender disparities in osteoporosis. *J Clin Med Res*. 2017 May;9(5):382-7. doi: 10.14740/jocmr2970w. PMID: 28392857.
11. International Society for Clinical Densitometry. Adult positions. Indications for bone mineral density (BMD) testing. Middletown, CT: International Society for Clinical Densitometry, Inc.; 2019. http://www.iscd.org/official-positions/2015-iscd-official-positions-adult/ Accessed 20 May 2021.
12. Kanis JA, World Health Organization Scientific Group. Assessment of osteoporosis at the primary health care level. Technical report. Sheffield, United Kingdom: World Health Organization Collaborating Centre for Metabolic Bone Diseases; 2008.
13. Schuit SC, van der Klift M, Weel AE, et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone*. 2004 Jan;34(1):195-202. PMID: 14751578.
14. Siris ES, Chen YT, Abbott TA, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med*. 2004 May 24;164(10):1108-12. doi: 10.1001/archinte.164.10.1108. PMID: 15159268.
15. Trajanoska K, Schoufour JD, de Jonge EAL, et al. Fracture incidence and secular trends between 1989 and 2013 in a population based cohort: the Rotterdam Study. *Bone*. 2018 Sep;114:116-24. doi: 10.1016/j.bone.2018.06.004. PMID: 29885926.
16. Sornay-Rendu E, Munoz F, Garnero P, et al. Identification of osteopenic women at high risk of fracture: the OFELY study. *Journal of Bone & Mineral Research*. 2005 Oct;20(10):1813-9. doi: 10.1359/jbmr.050609. PMID: 16160738.
17. Lyritis G, Rizou S. A revision in the definition of osteoporosis. *J Frailty Sarcopenia Falls*. 2016;1(1):01-3. doi: 10.22540/jfsf-01-001.
18. National Institute for Health and Care Excellence. Osteoporosis overview. National Institute for Health and Care Excellence. London, England. 2020. https://www.ncbi.nlm.nih.gov/books/NBK554920/bin/nicecg146guid_pathway1.pdf
19. Siris ES, Adler R, Bilezikian J, et al. The clinical diagnosis of osteoporosis: a position statement from the National Bone Health Alliance Working Group. *Osteoporos Int*. 2014 May;25(5):1439-43. doi: 10.1007/s00198-014-2655-z. PMID: 24577348.
20. Kanis JA, McCloskey EV, Harvey NC, et al. The need to distinguish intervention thresholds and diagnostic thresholds in the management of osteoporosis. *Osteoporos Int*. 2022 Oct 25. doi: 10.1007/s00198-022-06567-9. PMID: 36282342.

21. Sarafrazi N, Wambogo EA, Shepherd JA. Osteoporosis or low bone mass in older adults: United States, 2017–2018. NCHS Data Brief, no. 45. National Center for Health Statistics. Hyattsville, MD. 2021.
22. QuickStats: percentage* of adults aged ≥50 years with osteoporosis,(†) by race and hispanic origin(§) - United States, 2017-2018. *MMWR: Morbidity & Mortality Weekly Report*. 2021 May 14;70(19):731. doi: 10.15585/mmwr.mm7019a5. PMID: 33983912.
23. Nazrun AS, Tzar MN, Mokhtar SA, Mohamed IN. A systematic review of the outcomes of osteoporotic fracture patients after hospital discharge: morbidity, subsequent fractures, and mortality. *Therapeutics & Clinical Risk Management*. 2014;10:937-48. doi: 10.2147/TCRM.S72456. PMID: 25429224.
24. Andrich S, Haastert B, Neuhaus E, et al. Excess mortality after pelvic fractures among older people. *Journal of Bone & Mineral Research*. 2017 Sep;32(9):1789-801. doi: 10.1002/jbmr.3116. PMID: 28272751.
25. Hodsman AB, Leslie WD, Tsang JF, Gamble GD. 10-year probability of recurrent fractures following wrist and other osteoporotic fractures in a large clinical cohort: an analysis from the Manitoba Bone Density Program. *Arch Intern Med*. 2008 Nov 10;168(20):2261-7. doi: 10.1001/archinte.168.20.2261. PMID: 19001204.
26. Hsiao PC, Chen TJ, Li CY, et al. Risk factors and incidence of repeat osteoporotic fractures among the elderly in Taiwan: a population-based cohort study. *Medicine (Baltimore)*. 2015 Feb;94(7):e532. doi: 10.1097/MD.0000000000000532. PMID: 25700317.
27. Kanis JA, Johansson H, Oden A, et al. Characteristics of recurrent fractures. *Osteoporos Int*. 2018 Aug;29(8):1747-57. doi: 10.1007/s00198-018-4502-0. PMID: 29947869.
28. Dyer SM, Crotty M, Fairhall N, et al. A critical review of the long-term disability outcomes following hip fracture. *BMC Geriatr*. 2016 Sep 2;16:158. doi: 10.1186/s12877-016-0332-0. PMID: 27590604.
29. Old JL, Calvert M. Vertebral compression fractures in the elderly. *Am Fam Physician*. 2004 Jan 1;69(1):111-6. PMID: 14727827.
30. Schousboe JT. Mortality after osteoporotic fractures: what proportion is caused by fracture and is preventable? *Journal of Bone & Mineral Research*. 2017 Sep;32(9):1783-8. doi: 10.1002/jbmr.3216. PMID: 28691759.
31. Kanis JA, Oden A, Johnell O, et al. The components of excess mortality after hip fracture. *Bone*. 2003 May;32(5):468-73. doi: 10.1016/s8756-3282(03)00061-9. PMID: 12753862.
32. Bliuc D, Nguyen ND, Milch VE, et al. Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. *JAMA*. 2009 Feb 4;301(5):513-21. doi: 10.1001/jama.2009.50. PMID: 19190316.
33. Center JR, Nguyen TV, Schneider D, et al. Mortality after all major types of osteoporotic fracture in men and women: an observational study. *Lancet*. 1999 Mar 13;353(9156):878-82. doi: 10.1016/s0140-6736(98)09075-8. PMID: 10093980.
34. Kannegaard PN, van der Mark S, Eiken P, Abrahamsen B. Excess mortality in men compared with women following a hip fracture. National analysis of comedications, comorbidity and survival. *Age & Ageing*. 2010 Mar;39(2):203-9. doi: 10.1093/ageing/afp221. PMID: 20075035.
35. Lewiecki EM, Chastek B, Sundquist K, et al. Osteoporotic fracture trends in a population of US managed care enrollees from 2007 to 2017. *Osteoporos Int*. 2020 Jul;31(7):1299-304. doi: 10.1007/s00198-020-05334-y. PMID: 32062687.
36. Tran T, Bliuc D, van Geel T, et al. Population-wide impact of non-hip non-vertebral fractures on mortality. *Journal of Bone & Mineral Research*. 2017 Sep;32(9):1802-10. doi: 10.1002/jbmr.3118. PMID: 28256011.
37. Veronese N, Kolk H, Maggi S. Epidemiology of fragility fractures and social impact. In: Falaschi P, Marsh D, eds. Orthogeriatrics: the management of older patients with fragility fractures. Cham (CH): Springer; 2021:19-34.
38. Wu F, Laslett LL, Wills K, et al. Effects of individualized bone density feedback and educational interventions on osteoporosis knowledge and self-efficacy: a 12-yr prospective study. *J Clin Densitom*. 2014 Oct-Dec;17(4):466-72. doi: 10.1016/j.jocd.2014.07.008. PMID: 25220886.
39. Brauer CA, Coca-Perraillon M, Cutler DM, Rosen AB. Incidence and mortality of hip fractures in the United States. *JAMA*. 2009 Oct 14;302(14):1573-9. doi: 10.1001/jama.2009.1462. PMID: 19826027.
40. Lewiecki EM, Wright NC, Curtis JR, et al. Hip fracture trends in the United States, 2002 to 2015. *Osteoporos Int*. 2018 Mar;29(3):717-22. doi: 10.1007/s00198-017-4345-0. PMID: 29282482.
41. Cranney A, Jamal SA, Tsang JF, et al. Low bone mineral density and fracture burden in postmenopausal women. *CMAJ*. 2007 Sep 11;177(6):575-80. doi: 10.1503/cmaj.070234. PMID: 17846439.

42. Pasco JA, Seeman E, Henry MJ, et al. The population burden of fractures originates in women with osteopenia, not osteoporosis. *Osteoporos Int*. 2006;17(9):1404-9. doi: 10.1007/s00198-006-0135-9. PMID: 16699736.
43. Stone KL, Seeley DG, Lui LY, et al. BMD at multiple sites and risk of fracture of multiple types: long-term results from the Study of Osteoporotic Fractures. *Journal of Bone & Mineral Research*. 2003 Nov;18(11):1947-54. doi: 10.1359/jbmr.2003.18.11.1947. PMID: 14606506.
44. Wainwright SA, Marshall LM, Ensrud KE, et al. Hip fracture in women without osteoporosis. *Journal of Clinical Endocrinology & Metabolism*. 2005 May;90(5):2787-93. doi: 10.1210/jc.2004-1568. PMID: 15728213.
45. Sanders KM, Nicholson GC, Watts JJ, et al. Half the burden of fragility fractures in the community occur in women without osteoporosis. When is fracture prevention cost-effective? *Bone*. 2006 May;38(5):694-700. doi: 10.1016/j.bone.2005.06.004. PMID: 16507356.
46. Sobh MM, Abdalbary M, Elnagar S, et al. Secondary osteoporosis and metabolic bone diseases. *J Clin Med*. 2022 Apr 24;11(9):2382. doi: 10.3390/jcm11092382. PMID: 35566509.
47. Fitzpatrick LA. Secondary causes of osteoporosis. *Mayo Clin Proc*. 2002 May;77(5):453-68. doi: 10.4065/77.5.453. PMID: 12004995.
48. Sowers MR, Zheng H, Greendale GA, et al. Changes in bone resorption across the menopause transition: effects of reproductive hormones, body size, and ethnicity. *Journal of Clinical Endocrinology & Metabolism*. 2013 Jul;98(7):2854-63. doi: 10.1210/jc.2012-4113. PMID: 23666961.
49. Kanis JA, Johnell O, Oden A, et al. Ten year probabilities of osteoporotic fractures according to BMD and diagnostic thresholds. *Osteoporos Int*. 2001 Dec;12(12):989-95. doi: 10.1007/s001980170006 [doi]. PMID: 11846333.
50. Heaney RP. Bone mass, bone loss, and osteoporosis prophylaxis. *Ann Intern Med*. 1998 Feb 15;128(4):313-4. PMID: 9471936.
51. Richelson LS, Wahner HW, Melton LJ, 3rd, Riggs BL. Relative contributions of aging and estrogen deficiency to postmenopausal bone loss. *N Engl J Med*. 1984 Nov 15;311(20):1273-5. doi: 10.1056/NEJM198411153112002. PMID: 6493283.
52. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med*. 1995 Mar 23;332(12):767-73. doi: 10.1056/nejm199503233321202. PMID: 7862179.
53. Poiana C, Capatina C. Osteoporosis and fracture risk in patients with type 2 diabetes mellitus. *Acta Endocrinol (Buchar)*. 2019 Apr-Jun;15(2):231-6. doi: 10.4183/aeb.2019.231. PMID: 31508182.
54. Davie GS, Pal K, Orton E, et al. Incident Type 2 Diabetes and Risk of Fracture: A Comparative Cohort Analysis Using U.K. Primary Care Records. *Diabetes Care*. 2021 Jan;44(1):58-66. doi: 10.2337/dc20-1220. PMID: 33148635.
55. Axelsson KF, Litsne H, Kousoula K, et al. Risk of fracture in adults with type 2 diabetes in Sweden: A national cohort study. *PLoS Med*. 2023 Jan;20(1):e1004172. doi: 10.1371/journal.pmed.1004172. PMID: 36701363.
56. Drake MT, Murad MH, Mauck KF, et al. Clinical review. Risk factors for low bone mass-related fractures in men: a systematic review and meta-analysis. *Journal of Clinical Endocrinology & Metabolism*. 2012 Jun;97(6):1861-70. doi: 10.1210/jc.2011-3058. PMID: 22466344.
57. Liu LH, Chandra M, Gonzalez JR, Lo JC. Racial and ethnic differences in hip fracture outcomes in men. *Am J Manag Care*. 2017 2017/09;23(9):560-4.
58. Lo JC, Zheng P, Grimsrud CD, et al. Racial/ethnic differences in hip and diaphyseal femur fractures. *Osteoporos Int*. 2014 2014/09;25(9):2313-8. doi: 10.1007/s00198-014-2750-1.
59. Hansen D, Pelizzari PM, Pyenson BS. Medicare cost of osteoporotic fractures: 2021 updated report.
60. Lo JC, Chandra M, Lee C, et al. Bone mineral density in older U.S. Filipino, Chinese, Japanese, and White women. *J Am Geriatr Soc*. 2020 2020;68(11):2656-61. doi: 10.1111/jgs.16785.
61. Donovan Walker M, Babbar R, Opotowsky AR, et al. A referent bone mineral density database for Chinese American women. *Osteoporos Int*. 2006 2006/05/11;17(6):878-87. doi: 10.1007/s00198-005-0059-9.
62. Lo JC, Kim S, Chandra M, Ettinger B. Applying ethnic-specific bone mineral density T-scores to Chinese women in the USA. *Osteoporos Int*. 2016 Dec;27(12):3477-84. doi: 10.1007/s00198-016-3673-9. PMID: 27468900.
63. Ward RJ, Roberts CC, Bencardino JT, et al. ACR Appropriateness Criteria(®) osteoporosis and bone mineral density. *J Am Coll Radiol*. 2017 May;14(5s):S189-s202. doi: 10.1016/j.jacr.2017.02.018. PMID: 28473075.

64. Nelson HD, Haney EM, Chou R, et al. Screening for osteoporosis. systematic review to update the 2002 U.S. Preventive Services Task Force Recommendation. Evidence Synthesis No. 77. Report No.: 10-05145-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; July. 2010.
65. Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. *Lancet*. 2002 Jun 1;359(9321):1929-36. doi: 10.1016/s0140-6736(02)08761-5. PMID: 12057569.
66. Cummings SR, Black D. Bone mass measurements and risk of fracture in Caucasian women: a review of findings from prospective studies. *Am J Med*. 1995 Feb 27;98(2A):24S-8S. doi: 10.1016/s0002-9343(05)80041-5. PMID: 7709929.
67. Eastell R, Wahner HW, O'Fallon WM, et al. Unequal decrease in bone density of lumbar spine and ultradistal radius in Colles' and vertebral fracture syndromes. *J Clin Invest*. 1989 Jan;83(1):168-74. doi: 10.1172/JCI113854. PMID: 2910906.
68. Melton LJ, 3rd, Atkinson EJ, O'Fallon WM, et al. Long-term fracture prediction by bone mineral assessed at different skeletal sites. *Journal of Bone & Mineral Research*. 1993 Oct;8(10):1227-33. doi: 10.1002/jbmr.5650081010. PMID: 8256660.
69. Fuggle NR, Curtis EM, Ward KA, et al. Fracture prediction, imaging and screening in osteoporosis. *Nat Rev Endocrinol*. 2019 Sep;15(9):535-47. doi: 10.1038/s41574-019-0220-8. PMID: 31189982.
70. Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the US Preventive Services Task Force. *Jama*. 2018 Jun 26;319(24):2532-51. doi: 10.1001/jama.2018.6537. PMID: 29946734.
71. Kanis JA, Oden A, Johnell O, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int*. 2007 Aug;18(8):1033-46. doi: 10.1007/s00198-007-0343-y. PMID: 17323110.
72. Ettinger B, Liu H, Blackwell T, et al. Validation of FRC, a fracture risk assessment tool, in a cohort of older men: the Osteoporotic Fractures in Men (MrOS) Study. *J Clin Densitom*. 2012 Jul-Sep;15(3):334-42. doi: 10.1016/j.jocd.2012.01.011. PMID: 22445858.
73. Lo JC, Pressman AR, Chandra M, Ettinger B. Fracture risk tool validation in an integrated healthcare delivery system. *Am J Manag Care*. 2011 Mar;17(3):188-94. doi: 48148 [pii]. PMID: 21504255.
74. Nguyen ND, Frost SA, Center JR, et al. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. *Osteoporos Int*. 2008 Oct;19(10):1431-44. doi: 10.1007/s00198-008-0588-0. PMID: 18324342.
75. Nguyen ND, Frost SA, Center JR, et al. Development of a nomogram for individualizing hip fracture risk in men and women. *Osteoporos Int*. 2007 Aug;18(8):1109-17. doi: 10.1007/s00198-007-0362-8. PMID: 17370100.
76. Lydick E, Cook K, Turpin J, et al. Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. *Am J Manag Care*. 1998 Jan;4(1):37-48. PMID: 10179905.
77. Kanis JA, Harvey NC, Johansson H, et al. Overview of fracture prediction tools. *J Clin Densitom*. 2017 Jul-Sep;20(3):444-50. doi: 10.1016/j.jocd.2017.06.013. PMID: 28716500.
78. FRAX. FRAX® fracture risk assessment tool. Charts to download. UK: University of Sheffield; 2008. https://www.sheffield.ac.uk/FRAX/charts.aspx. Accessed 19 May 2021.
79. Lewiecki EM, Erb SF. Racial disparities and inequalities in the management of patients with osteoporosis. *Orthop Nurs*. 2022 Mar-Apr 01;41(2):125-34. doi: 10.1097/NOR.0000000000000832. PMID: 35358132.
80. Dawson-Hughes B, Tosteson AN, Melton LJ, 3rd, et al. Implications of absolute fracture risk assessment for osteoporosis practice guidelines in the USA. *Osteoporos Int*. 2008 Apr;19(4):449-58. doi: 10.1007/s00198-008-0559-5. PMID: 18292975.
81. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020 Aug 27;383(9):874-82. doi: 10.1056/NEJMms2004740. PMID: 32853499.
82. Simpson EL, Martyn-St James M, Hamilton J, et al. Clinical effectiveness of denosumab, raloxifene, romosozumab, and teriparatide for the prevention of osteoporotic fragility fractures: a systematic review and network meta-analysis. *Bone*. 2020 Jan;130:115081. doi: 10.1016/j.bone.2019.115081. PMID: 31626995.
83. Barrionuevo P, Kapoor E, Asi N, et al. Efficacy of pharmacological therapies for the prevention of fractures in postmenopausal women: a network meta-analysis. *Journal of Clinical Endocrinology & Metabolism*. 2019 May 1;104(5):1623-30. doi: 10.1210/jc.2019-00192. PMID: 30907957.

84. LeBoff MS, Greenspan SL, Insogna KL, et al. The clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int*. 2022 Oct;33(10):2049-102. doi: 10.1007/s00198-021-05900-y. PMID: 35478046.
85. Tosteson AN, Melton LJ, 3rd, Dawson-Hughes B, et al. Cost-effective osteoporosis treatment thresholds: the United States perspective. *Osteoporos Int*. 2008 Apr;19(4):437-47. doi: 10.1007/s00198-007-0550-6. PMID: 18292976.
86. Dawson-Hughes B, Looker AC, Tosteson AN, et al. The potential impact of the National Osteoporosis Foundation guidance on treatment eligibility in the USA: an update in NHANES 2005-2008. *Osteoporos Int*. 2012 Mar;23(3):811-20. doi: 10.1007/s00198-011-1694-y. PMID: 21717247.
87. Dawson-Hughes B, Looker AC, Tosteson AN, et al. The potential impact of new National Osteoporosis Foundation guidance on treatment patterns. *Osteoporos Int*. 2010 Jan;21(1):41-52. doi: 10.1007/s00198-009-1034-7. PMID: 19705046.
88. Kanis JA, Harvey NC, Cooper C, et al. A systematic review of intervention thresholds based on FRAX: a report prepared for the National Osteoporosis Guideline Group and the International Osteoporosis Foundation. *Arch Osteoporos*. 2016 Dec;11(1):25. doi: 10.1007/s11657-016-0278-z. PMID: 27465509.
89. Kanis JA, Harvey NC, Johansson H, et al. FRAX update. *J Clin Densitom*. 2017 Jul - Sep;20(3):360-7. doi: 10.1016/j.jocd.2017.06.022. PMID: 28732576.
90. Thériault G, Limburg H, Klarenbach S, et al. Recommendations on screening for primary prevention of fragility fractures. *CMAJ*. 2023 May 8;195(18):e639-e49. doi: 10.1503/cmaj.221219. PMID: 37156553.
91. Camacho PM, Petak SM, Binkley N, et al. American Association of Clinical Endocrinologists/American College of Endocrinology clinical practice guidelines for the diagnosis and treatment of postmenopausal osteoporosis—2020 update. *Endocr Pract*. 2020 May;26(Suppl 1):1-46. doi: 10.4158/gl-2020-0524suppl. PMID: 32427503.
92. French S, Choden S, Schmajuk G. Quality measures and quality improvement initiatives in osteoporosis—an update. *Curr Osteoporos Rep*. 2019 Dec;17(6):491-509. doi: 10.1007/s11914-019-00547-5. PMID: 31782030.
93. National Committee for Quality Assurance. 2015 state of health care quality table of contents. Washington, DC: National Committee for Quality Assurance; 2015. http://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality/2015-table-of-contents/osteoporosis. Accessed 10 Aug 2016.
94. Centers for Medicare & Medicaid Services. Measures inventory tool. Screening for osteoporosis for women aged 65-85 years of age. Baltimore, MD: Centers for Medicare & Medicaid Services; n.d. https://mdinteractive.com/mips_quality_measure/2022-mips-quality-measure-03920 May 2021.
95. Curtis JR, Sharma P, Arora T, et al. Physicians' explanations for apparent gaps in the quality of rheumatology care: results from the US Medicare Physician Quality Reporting System. *Arthritis Care Res (Hoboken)*. 2013 Feb;65(2):235-43. doi: 10.1002/acr.21713. PMID: 22556118.
96. Amarnath AL, Franks P, Robbins JA, et al. Underuse and overuse of osteoporosis screening in a regional health system: a retrospective cohort study. *J Gen Intern Med*. 2015 Dec;30(12):1733-40. doi: 10.1007/s11606-015-3349-8. PMID: 25986135.
97. Curtis JR, McClure LA, Delzell E, et al. Population-based fracture risk assessment and osteoporosis treatment disparities by race and gender. *J Gen Intern Med*. 2009 Aug;24(8):956-62. doi: 10.1007/s11606-009-1031-8. PMID: 19551449.
98. Choosing Wisely. Bone-density tests. Philadelphia, PA ABIM Foundation; 2016. https://www.choosingwisely.org/patient-resources/bone-density-tests/. Accessed 9 Dec 2022.
99. Good Stewardship Working Group. The "top 5" lists in primary care: meeting the responsibility of professionalism. *Arch Intern Med.* 2011 Aug 08;171(15):1385-90. doi: 10.1001/archinternmed.2011.231. PMID: 21606090.
100. Centers for Medicare & Medicaid Services. Measures inventory tool. Appropriate use of DXA scans in women under 65 years who do not meet the risk factor profile for osteoporotic fracture (eCQM). Baltimore, MD: Centers for Medicare & Medicaid Services; n.d. https://mdinteractive.com/mips_quality_measure/2022-mips-quality-measure-039. Accessed 20 May 2021.
101. Kale MS, Bishop TF, Federman AD, Keyhani S. "Top 5" lists top $5 billion. *Arch Intern Med.* 2011 Nov 14;171(20):1856-8. doi: 10.1001/archinternmed.2011.501. PMID: 21965814.
102. Reynolds K, Muntner P, Cheetham TC, et al. Primary non-adherence to bisphosphonates in an integrated healthcare setting. *Osteoporos Int*. 2013 Sep;24(9):2509-17. doi: 10.1007/s00198-013-2326-5. PMID: 23595561.

103. Cromer SJ, D'Silva KM, Yu EW, et al. Secular trends in the pharmacologic treatment of osteoporosis and malignancy-related bone disease from 2009 to 2020. *J Gen Intern Med*. 2022 Jun;37(8):1917-24. doi: 10.1007/s11606-021-06938-8. PMID: 34100235.
104. United Nations Development Programme. Human Development Index (HDI) ranking. In: Human Development Report 2020: The Next Frontier: Human Development and the Anthropocene. 2020. https://hdr.undp.org/en/content/latest-human-development-index-ranking. Accessed 14 Nov 2022.
105. Kanis JA, Oden A, McCloskey EV, et al. A systematic review of hip fracture incidence and probability of fracture worldwide. *Osteoporos Int*. 2012 Sep;23(9):2239-56. doi: 10.1007/s00198-012-1964-3. PMID: 22419370.
106. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj*. 2019 Aug 28;366:l4898. doi: 10.1136/bmj.l4898. PMID: 31462531.
107. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919. PMID: 27733354.
108. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.
109. Whiting P, Savović J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016 Jan;69:225-34. doi: 10.1016/j.jclinepi.2015.06.005. PMID: 26092286.
110. Venema E, Wessler BS, Paulus JK, et al. Large-scale validation of the prediction model risk of bias assessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. *J Clin Epidemiol*. 2021 Oct;138:32-9. doi: 10.1016/j.jclinepi.2021.06.017. PMID: 34175377.
111. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019 Jan 1;170(1):51-8. doi: 10.7326/M18-1376. PMID: 30596875.
112. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019 Jan 1;170(1):W1-w33. doi: 10.7326/m18-1377. PMID: 30596876.
113. U.S. Preventive Services Task Force. U.S. Preventive Services Task Force procedure manual. U.S. Preventive Services Task Force. Rockville, MD. 2021. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual
114. West SL, Gartlehner G, Mansfield AJ, et al. Comparative effectiveness review methods: clinical heterogeneity. Methods Research Paper. AHRQ Publication No. 10-EHC070-EF. Rockville MD: Agency for Healthcare Research and Quality; Sept. 2010. https://effectivehealthcare.ahrq.gov/products/clinical-heterogeneity-methods/research
115. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986 Sep;7(3):177-88. PMID: 3802833.
116. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons. Chichester, UK; 12 Nov. 2022. https://training.cochrane.org/handbook/current
117. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002 Jun 15;21(11):1539-58. doi: 10.1002/sim.1186. PMID: 12111919.
118. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003 Sep 6;327(7414):557-60. doi: 10.1136/bmj.327.7414.557. PMID: 12958120.
119. Berkman ND, Lohr KN, Ansari MT, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. *J Clin Epidemiol*. 2015 Nov;68(11):1312-24. doi: 10.1016/j.jclinepi.2014.11.023. PMID: 25721570.
120. Shepstone L, Lenaghan E, Cooper C, et al. Screening in the community to reduce fractures in older women (SCOOP): a randomised controlled trial. *Lancet*. 2018 Feb 24;391(10122):741-7. doi: 10.1016/s0140-6736(17)32640-5. PMID: 29254858.
121. Shepstone L, Fordham R, Lenaghan E, et al. A pragmatic randomised controlled trial of the effectiveness and cost-effectiveness of screening older women for the prevention of fractures: rationale, design and methods for the SCOOP study. *Osteoporos Int*. 2012 Oct;23(10):2507-15. doi: 10.1007/s00198-011-1876-7. PMID: 22314936.
122. McCloskey E, Johansson H, Harvey NC, et al. Management of patients with high baseline hip fracture risk by FRAX reduces hip fractures-a post hoc analysis of the SCOOP study. *Journal of Bone & Mineral Research*. 2018 Jun;33(6):1020-6. doi: 10.1002/jbmr.3411. PMID: 29480960.

123. Parsons CM, Harvey N, Shepstone L, et al. Systematic screening using FRAX(®) leads to increased use of, and adherence to, anti-osteoporosis medications: an analysis of the UK SCOOP trial. *Osteoporos Int*. 2020 Jan;31(1):67-75. doi: 10.1007/s00198-019-05142-z. PMID: 31606826.
124. Merlijn T, Swart KM, van Schoor NM, et al. The effect of a screening and treatment program for the prevention of fractures in older women: a randomized pragmatic trial. *Journal of Bone & Mineral Research*. 2019 Nov;34(11):1993-2000. doi: 10.1002/jbmr.3815. PMID: 31220365.
125. Elders PJM, Merlijn T, Swart KMA, et al. Design of the SALT osteoporosis study: a randomised pragmatic trial, to study a primary care screening and treatment program for the prevention of fractures in women aged 65 years or older. *BMC Musculoskelet Disord*. 2017 Oct 27;18(1):424. doi: 10.1186/s12891-017-1783-y. PMID: 29078762.
126. Rubin KH, Rothmann MJ, Holmberg T, et al. Effectiveness of a two-step population-based osteoporosis screening program using FRAX: the randomized Risk-stratified Osteoporosis Strategy Evaluation (ROSE) study. *Osteoporos Int*. 2018 Mar;29(3):567-78. doi: 10.1007/s00198-017-4326-3. PMID: 29218381.
127. Rubin KH, Holmberg T, Rothmann MJ, et al. The risk-stratified osteoporosis strategy evaluation study (ROSE): a randomized prospective population-based study. Design and baseline characteristics. *Calcif Tissue Int*. 2015 Feb;96(2):167-79. doi: 10.1007/s00223-014-9950-8. PMID: 25578146.
128. Rothmann MJ, Möller S, Holmberg T, et al. Non-participation in systematic screening for osteoporosis-the ROSE trial. *Osteoporos Int*. 2017 Dec;28(12):3389-99. doi: 10.1007/s00198-017-4205-y. PMID: 28875257.
129. Høiberg MP, Rubin KH, Holmberg T, et al. Use of antiosteoporotic medication in the Danish ROSE population-based screening study. *Osteoporos Int*. 2019 Jun;30(6):1223-33. doi: 10.1007/s00198-019-04934-7. PMID: 30911782.
130. Merlijn T, Swart KMA, van der Horst HE, et al. Fracture prevention by screening for high fracture risk: a systematic review and meta-analysis. *Osteoporos Int*. 2020 Feb;31(2):251-7. doi: 10.1007/s00198-019-05226-w. PMID: 31838551.
131. Gates M, Pillay J, Nuspl M, et al. Screening for the primary prevention of fragility fractures among adults aged 40 years and older in primary care: systematic reviews of the effects and acceptability of screening and treatment, and the accuracy of risk prediction tools. *Syst Rev*. 2023 Mar 21;12(1):51. doi: 10.1186/s13643-023-02181-w. PMID: 36945065.
132. Gates M, Pillay J, Thériault G, et al. Screening to prevent fragility fractures among adults 40 years and older in primary care: protocol for a systematic review. *Syst Rev*. 2019 Aug 23;8(1):216. doi: 10.1186/s13643-019-1094-5. PMID: 31443711.
133. Auais M, Angermann H, Grubb M, et al. The effectiveness and cost-effectiveness of clinical fracture-risk assessment tools in reducing future osteoporotic fractures among older adults: a structured scoping review. *Osteoporos Int*. 2023 May;34(5):823-40. doi: 10.1007/s00198-022-06659-6. PMID: 36598523.
134. Marques A, Ferreira RJ, Santos E, et al. The accuracy of osteoporotic fracture risk prediction tools: a systematic review and meta-analysis. *Ann Rheum Dis*. 2015 Nov;74(11):1958-67. doi: 10.1136/annrheumdis-2015-207907. PMID: 26248637.
135. Beaudoin C, Moore L, Gagné M, et al. Performance of predictive tools to identify individuals at risk of non-traumatic fracture: a systematic review, meta-analysis, and meta-regression. *Osteoporos Int*. 2019 Apr;30(4):721-40. doi: 10.1007/s00198-019-04919-6. PMID: 30877348.
136. Jiang X, Gruner M, Tremollieres F, et al. Diagnostic accuracy of FRAX in predicting the 10-year risk of osteoporotic fractures using the USA treatment thresholds: a systematic review and meta-analysis. *Bone*. 2017 Jun;99:20-5. doi: 10.1016/j.bone.2017.02.008. PMID: 28274799.
137. Sun X, Chen Y, Gao Y, et al. Prediction models for osteoporotic fractures risk: a systematic review and critical appraisal. *Aging & Disease*. 2022 Jul 11;13(4):1215-38. doi: 10.14336/ad.2021.1206. PMID: 35855348.
138. Adami G, Biffi A, Porcu G, et al. A systematic review on the performance of fracture risk assessment tools: FRAX, DeFRA, FRA-HS. *J Endocrinol Invest*. 2023 Nov;46(11):2287-97. doi: 10.1007/s40618-023-02082-8. PMID: 37031450.
139. Crandall CJ, Larson JC, Watts NB, et al. Comparison of fracture risk prediction by the US Preventive Services Task Force strategy and two alternative strategies in women 50-64 years old in the Women's Health Initiative. *Journal of Clinical Endocrinology & Metabolism*. 2014;99(12):4514-22. doi: 10.1210/jc.2014-2332. PMID: 25322268.

140. Crandall CJ, Larson J, LaCroix A, et al. Predicting fracture risk in younger postmenopausal women: Comparison of the Garvan and FRAX risk calculators in the Women's Health Initiative Study. *J Gen Intern Med*. 2019 Feb;34(2):235-42. doi: 10.1007/s11606-018-4696-z. PMID: 30334182.
141. Crandall CJ, Larson J, Manson JE, et al. A comparison of US and Canadian osteoporosis screening and treatment strategies in postmenopausal women. *Journal of Bone & Mineral Research*. 2019 Apr;34(4):607-15. doi: 10.1002/jbmr.3636. PMID: 30536628.
142. Crandall CJ, Larson J, Cauley JA, et al. Do additional clinical risk factors improve the performance of Fracture Risk Assessment Tool (FRAX) among postmenopausal women? Findings from the Women's Health Initiative observational study and clinical trials. *JBMR Plus*. 2019 Dec;3(12):e10239. doi: 10.1002/jbm4.10239. PMID: 31844827.
143. Crandall CJ, Larson JC, Schousboe JT, et al. Race and ethnicity and fracture prediction among younger postmenopausal women in the Women's Health Initiative Study. *JAMA Intern. Med*. 2023. doi: 10.1001/jamainternmed.2023.1253.
144. Tebe Cordomi C, Del Rio LM, Di Gregorio S, et al. Validation of the FRAX predictive model for major osteoporotic fracture in a historical cohort of Spanish women. *J Clin Densitom*. 2013 Apr-Jun;16(2):231-7. doi: 10.1016/j.jocd.2012.05.007. PMID: 22748778.
145. Ettinger B, Ensrud KE, Blackwell T, et al. Performance of FRAX in a cohort of community-dwelling, ambulatory older men: the Osteoporotic Fractures in Men (MrOS) study. *Osteoporos Int*. 2013 Apr;24(4):1185-93. doi: 10.1007/s00198-012-2215-3. PMID: 23179575.
146. Gourlay ML, Ritter VS, Fine JP, et al. Comparison of fracture risk assessment tools in older men without prior hip or spine fracture: the MrOS study. *Arch Osteoporos*. 2017 Oct 20;12(1):91. doi: 10.1007/s11657-017-0389-1. PMID: 29052793.
147. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *BMJ*. 2012 May 22;344:e3427. doi: 10.1136/bmj.e3427. PMID: 22619194.
148. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ*. 2009 Nov 19;339:b4229. doi: 10.1136/bmj.b4229. PMID: 19926696.
149. Gonzalez-Macias J, Marin F, Vila J, Diez-Perez A. Probability of fractures predicted by FRAX(R) and observed incidence in the Spanish ECOSAP Study cohort. *Bone*. 2012 Jan;50(1):373-7. doi: 10.1016/j.bone.2011.11.006. PMID: 22129640.
150. Cheung EY, Bow CH, Cheung CL, et al. Discriminative value of FRAX for fracture prediction in a cohort of Chinese postmenopausal women. *Osteoporos Int*. 2012 Mar;23(3):871-8. doi: 10.1007/s00198-011-1647-5. PMID: 21562875.
151. Tamaki J, Iki M, Kadowaki E, et al. Fracture risk prediction using FRAX(R): a 10-year follow-up survey of the Japanese Population-Based Osteoporosis (JPOS) Cohort Study. *Osteoporos Int*. 2011 Dec;22(12):3037-45. doi: 10.1007/s00198-011-1537-x. PMID: 21279504.
152. Ensrud KE, Lui LY, Taylor BC, et al. A comparison of prediction models for fractures in older women: is more better? *Arch Intern Med*. 2009 Dec 14;169(22):2087-94. doi: 10.1001/archinternmed.2009.404. PMID: 20008691.
153. Premaor M, Parker RA, Cummings S, et al. Predictive value of FRAX for fracture in obese older women. *Journal of Bone & Mineral Research*. 2013 Jan;28(1):188-95. doi: 10.1002/jbmr.1729. PMID: 22890977.
154. Fraser LA, Langsetmo L, Berger C, et al. Fracture prediction and calibration of a Canadian FRAX(R) tool: a population-based report from CaMos. *Osteoporos Int*. 2011 Mar;22(3):829-37. doi: 10.1007/s00198-010-1465-1. PMID: 21161508.
155. Langsetmo L, Nguyen TV, Nguyen ND, et al. Independent external validation of nomograms for predicting risk of low-trauma fracture and hip fracture. *CMAJ*. 2011 Feb 8;183(2):E107-14. doi: 10.1503/cmaj.100458. PMID: 21173069.
156. Tanaka S, Yoshimura N, Kuroda T, et al. The Fracture and Immobilization Score (FRISC) for risk assessment of osteoporotic fracture and immobilization in postmenopausal women--a joint analysis of the Nagano, Miyama, and Taiji Cohorts. *Bone*. 2010 Dec;47(6):1064-70. doi: 10.1016/j.bone.2010.08.019. PMID: 20832514.
157. Leslie WD, Lix LM, Johansson H, et al. Independent clinical validation of a Canadian FRAX tool: fracture prediction and model calibration. *Journal of Bone & Mineral Research*. 2010 Nov;25(11):2350-8. doi: 10.1002/jbmr.123. PMID: 20499367.

158. Brennan SL, Leslie WD, Lix LM, et al. FRAX provides robust fracture prediction regardless of socioeconomic status. *Osteoporos Int*. 2014 Jan;25(1):61-9. doi: 10.1007/s00198-013-2525-0. PMID: 24190425.
159. Morin S, Tsang JF, Leslie WD. Weight and body mass index predict bone mineral density and fractures in women aged 40 to 59 years. *Osteoporos Int*. 2009 Mar;20(3):363-70. doi: 10.1007/s00198-008-0688-x. PMID: 18633665.
160. Crandall CJ, Schousboe JT, Morin SN, et al. Performance of FRAX and FRAX-based treatment thresholds in women aged 40 years and older: the Manitoba BMD registry. *Journal of Bone & Mineral Research*. 2019 Aug;34(8):1419-27. doi: 10.1002/jbmr.3717. PMID: 30920022.
161. Leslie WD, Majumdar SR, Lix LM, et al. Direct comparison of FRAX(R) and a simplified fracture risk assessment tool in routine clinical practice: a registry-based cohort study. *Osteoporos Int*. 2016 Sep;27(9):2689-95. doi: 10.1007/s00198-016-3613-8. PMID: 27108119.
162. Leslie WD, Majumdar SR, Morin SN, et al. Performance of FRAX in clinical practice according to sex and osteoporosis definitions: the Manitoba BMD registry. *Osteoporos Int*. 2018 Mar;29(3):759-67. doi: 10.1007/s00198-018-4415-y. PMID: 29404625.
163. Möller S, Skjødt MK, Yan L, et al. Prediction of imminent fracture risk in Canadian women and men aged 45 years or older: external validation of the Fracture Risk Evaluation Model (FREM). *Osteoporos Int*. 2022 Jan;33(1):57-66. doi: 10.1007/s00198-021-06165-1. PMID: 34596704.
164. Leslie WD, Möller S, Skjødt MK, et al. FREM predicts 10-year incident fracture risk independent of FRAX® probability: a registry-based cohort study. *Osteoporos Int*. 2022 Jul;33(7):1457-63. doi: 10.1007/s00198-022-06349-3. PMID: 35175395.
165. Pressman AR, Lo JC, Chandra M, Ettinger B. Methods for assessing fracture risk prediction models: experience with FRAX in a large integrated health care delivery system. *J Clin Densitom*. 2011 Oct-Dec;14(4):407-15. doi: 10.1016/j.jocd.2011.06.006. PMID: 21958955.
166. Bolland MJ, Siu AT, Mason BH, et al. Evaluation of the FRAX and Garvan fracture risk calculators in older women. *Journal of Bone & Mineral Research*. 2011 Feb;26(2):420-7. doi: 10.1002/jbmr.215. PMID: 20721930.
167. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. *BMJ*. 2011 Jun 22;342:d3651. doi: 10.1136/bmj.d3651. PMID: 21697214.
168. Azagra R, Roca G, Martin-Sanchez JC, et al. [FRAX(R) thresholds to identify people with high or low risk of osteoporotic fracture in Spanish female population]. *Med Clin (Barc)*. 2015 Jan 6;144(1):1-8. doi: 10.1016/j.medcli.2013.11.014. PMID: 24461732.
169. Dagan N, Cohen-Stavi C, Leventer-Roberts M, Balicer RD. External validation and comparison of three prediction tools for risk of osteoporotic fractures using data from population based electronic health records: retrospective cohort study. *BMJ*. 2017 Jan 19;356:i6755. doi: 10.1136/bmj.i6755. PMID: 28104610.
170. Klop C, de Vries F, Bijlsma JW, et al. Predicting the 10-year risk of hip and major osteoporotic fracture in rheumatoid arthritis and in the general population: an independent validation and update of UK FRAX without bone mineral density. *Ann Rheum Dis*. 2016 Dec;75(12):2095-100. doi: 10.1136/annrheumdis-2015-208958. PMID: 26984006.
171. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. *BMJ Open*. 2014 Aug 28;4(8):e005809. doi: 10.1136/bmjopen-2014-005809. PMID: 25168040.
172. Goldshtein I, Gerber Y, Ish-Shalom S, Leshno M. Fracture risk assessment with FRAX using real-world data in a population-based cohort from Israel. *Am J Epidemiol*. 2018 Jan 1;187(1):94-102. doi: 10.1093/aje/kwx128. PMID: 28520844.
173. Davis WA, Hamilton EJ, Bruce DG, Davis TME. Development and validation of a simple hip fracture risk prediction tool for type 2 diabetes: the Fremantle Diabetes Study phase I. *Diabetes Care*. 2019 Jan;42(1):102-9. doi: 10.2337/dc18-1486. PMID: 30455327.
174. Azagra R, Zwart M, Aguyé A, et al. Fracture experience among participants from the FROCAT study: what thresholding is appropriate using the FRAX tool? *Maturitas*. 2016 Jan;83:65-71. doi: 10.1016/j.maturitas.2015.10.002. PMID: 26546077.
175. Chapurlat R, Bui M, Sornay-Rendu E, et al. Deterioration of cortical and trabecular microstructure identifies women with osteopenia or normal bone mineral density at imminent and long-term risk for

fragility fracture: a prospective study. *Journal of Bone & Mineral Research*. 2020 May;35(5):833-44. doi: 10.1002/jbmr.3924. PMID: 31821619.

176. Marques A, Lucas R, Simões E, et al. Do we need bone mineral density to estimate osteoporotic fracture risk? A 10-year prospective multicentre validation study. *RMD Open*. 2017;3(2):e000509. doi: 10.1136/rmdopen-2017-000509. PMID: 29018567.

177. Desbiens LC, Sidibé A, Beaudoin C, et al. Comparison of fracture prediction tools in individuals without and with early chronic kidney disease: a population-based analysis of CARTaGENE. *Journal of Bone & Mineral Research*. 2020 Jun;35(6):1048-57. doi: 10.1002/jbmr.3977. PMID: 32022942.

178. Lu T, Forgetta V, Keller-Baruch J, et al. Improved prediction of fracture risk leveraging a genome-wide polygenic risk score. *Genome Med*. 2021 Feb 3;13(1):16. doi: 10.1186/s13073-021-00838-6. PMID: 33536041.

179. Baleanu F, Iconaru L, Charles A, et al. Independent external validation of FRAX and Garvan fracture risk calculators: a sub-study of the FRISBEE cohort. *JBMR Plus*. 2021 Sep;5(9):e10532. doi: 10.1002/jbm4.10532. PMID: 34532617.

180. Tebé C, Pallarès N, Reyes C, et al. Development and external validation of a 1- and 5-year fracture prediction tool based on electronic medical records data: The EPIC risk algorithm. *Bone*. 2022 Sep;162:116469. doi: 10.1016/j.bone.2022.116469. PMID: 35691583.

181. García-Sempere A, Hurtado I, Peiró S, et al. Predictive Performance of the FRAX Tool Calibrated for Spain vs. an Age and Sex Model: Prospective Cohort Study with 9082 Women and Men Followed for up to 8 Years. *Journal of Clinical Medicine*. 2022;11(9). doi: 10.3390/jcm11092409.

182. Kwok T, Khoo CC, Leung J, et al. Predictive values of calcaneal quantitative ultrasound and dual energy X ray absorptiometry for non-vertebral fracture in older men: results from the MrOS study (Hong Kong). *Osteoporos Int*. 2012 Mar;23(3):1001-6. doi: 10.1007/s00198-011-1634-x. PMID: 21528361.

183. Hans D, Goertzen AL, Krieg MA, Leslie WD. Bone microarchitecture assessed by TBS predicts osteoporotic fractures independent of bone density: the Manitoba study. *Journal of Bone & Mineral Research*. 2011 Nov;26(11):2762-9. doi: 10.1002/jbmr.499. PMID: 21887701.

184. Leslie WD, Aubry-Rozier B, Lamy O, Hans D. TBS (trabecular bone score) and diabetes-related fracture risk. *Journal of Clinical Endocrinology & Metabolism*. 2013 Feb;98(2):602-9. doi: 10.1210/jc.2012-3118. PMID: 23341489.

185. Sund R, Honkanen R, Johansson H, et al. Evaluation of the FRAX model for hip fracture predictions in the population-based Kuopio Osteoporosis Risk Factor and Prevention Study (OSTPRE). *Calcif Tissue Int*. 2014 Jul;95(1):39-45. doi: 10.1007/s00223-014-9860-9. PMID: 24792689.

186. Nguyen TV, Center JR, Eisman JA. Bone mineral density-independent association of quantitative ultrasound measurements and fracture risk in women. *Osteoporos Int*. 2004 Dec;15(12):942-7. doi: 10.1007/s00198-004-1717-z. PMID: 15309384.

187. Robbins J, Aragaki AK, Kooperberg C, et al. Factors associated with 5-year risk of hip fracture in postmenopausal women. *JAMA*. 2007 Nov 28;298(20):2389-98. doi: 10.1001/jama.298.20.2389. PMID: 18042916.

188. Sornay-Rendu E, Munoz F, Delmas PD, Chapurlat RD. The FRAX tool in French women: How well does it describe the real incidence of fracture in the OFELY cohort? *Journal of Bone & Mineral Research*. 2010 Oct;25(10):2101-7. doi: 10.1002/jbmr.106. PMID: 20499352.

189. Tremollieres FA, Pouilles JM, Drewniak N, et al. Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: sensitivity of the WHO FRAX tool. *Journal of Bone & Mineral Research*. 2010 May;25(5):1002-9. doi: 10.1002/jbmr.12. PMID: 20200927.

190. Stewart A, Kumar V, Reid DM. Long-term fracture prediction by DXA and QUS: a 10-year prospective study. *Journal of Bone & Mineral Research*. 2006 Mar;21(3):413-8. doi: 10.1359/jbmr.051205. PMID: 16491289.

191. Prince RL, Lewis JR, Lim WH, et al. Adding lateral spine imaging for vertebral fractures to densitometric screening: improving ascertainment of patients at high risk of incident osteoporotic fractures. *Journal of Bone & Mineral Research*. 2019 Feb;34(2):282-9. doi: 10.1002/jbmr.3595. PMID: 30395687.

192. Black DM, Cauley JA, Wagman R, et al. The ability of a single BMD and fracture history assessment to predict fracture over 25 years in postmenopausal women: the study of osteoporotic fractures. *Journal of Bone & Mineral Research*. 2018 Mar;33(3):389-95. doi: 10.1002/jbmr.3194. PMID: 28719727.

193. Agarwal A, Leslie WD, Nguyen TV, et al. Predictive performance of the Garvan Fracture Risk Calculator: a registry-based cohort study. *Osteoporos Int*. 2022 Mar;33(3):541-8. doi: 10.1007/s00198-021-06252-3. PMID: 34839377.

194. Iki M, Winzenrieth R, Tamaki J, et al. Predictive ability of novel volumetric and geometric indices derived from dual-energy X-ray absorptiometric images of the proximal femur for hip fracture compared with conventional areal bone mineral density: the Japanese Population-based Osteoporosis (JPOS) Cohort Study. *Osteoporos Int*. 2021 May 26;32(11):2289-99. doi: 10.1007/s00198-021-06013-2. PMID: 34041560.
195. Crandall CJ, Larson J, Gourlay ML, et al. Osteoporosis screening in postmenopausal women 50 to 64 years old: comparison of US Preventive Services Task Force strategy and two traditional strategies in the Women's Health Initiative. *Journal of Bone & Mineral Research*. 2014 Jul;29(7):1661-6. doi: 10.1002/jbmr.2174. PMID: 24431262.
196. Pang WY, Inderjeeth CA. FRAX without bone mineral density versus osteoporosis self-assessment screening tool as predictors of osteoporosis in primary screening of individuals aged 70 and older. *J Am Geriatr Soc*. 2014 Mar;62(3):442-6. doi: 10.1111/jgs.12696. PMID: 24617899.
197. Cass AR, Shepherd AJ. Validation of the Male Osteoporosis Risk Estimation Score (MORES) in a primary care setting. *J Am Board Fam Med*. 2013 Jul-Aug;26(4):436-44. doi: 10.3122/jabfm.2013.04.120182. PMID: 23833159.
198. Richards JS, Lazzari AA, Teves Qualler DA, et al. Validation of the osteoporosis self-assessment tool in US male veterans. *Journal of clinical densitometry*. 2014;17(1):32-7.
199. D'Amelio P, Spertino E, Martino F, Isaia GC. Prevalence of postmenopausal osteoporosis in Italy and validation of decision rules for referring women for bone densitometry. *Calcif Tissue Int*. 2013 May;92(5):437-43. doi: 10.1007/s00223-013-9699-5. PMID: 23334353.
200. Leslie WD, Lix LM, Johansson H, et al. Selection of women aged 50-64 yr for bone density measurement. *J Clin Densitom*. 2013 Oct-Dec;16(4):570-8. doi: 10.1016/j.jocd.2013.01.004. PMID: 23452870.
201. Machado P, Coutinho M, da Silva JA. Selecting men for bone densitometry: performance of osteoporosis risk assessment tools in Portuguese men. *Osteoporos Int*. 2010 Jun;21(6):977-83. doi: 10.1007/s00198-009-1036-5. PMID: 19727909.
202. Shepherd AJ, Cass AR, Ray L. Determining risk of vertebral osteoporosis in men: validation of the male osteoporosis risk estimation score. *J Am Board Fam Med*. 2010 Mar-Apr;23(2):186-94. doi: 10.3122/jabfm.2010.02.090027. PMID: 20207929.
203. Jimenez-Nunez FG, Manrique-Arija S, Urena-Garnica I, et al. Reducing the need for central dual-energy X-ray absorptiometry in postmenopausal women: efficacy of a clinical algorithm including peripheral densitometry. *Calcif Tissue Int*. 2013 Jul;93(1):62-8. doi: 10.1007/s00223-013-9728-4. PMID: 23608922.
204. Oh SM, Nam BH, Rhee Y, et al. Development and validation of osteoporosis risk-assessment model for Korean postmenopausal women. *Journal of Bone & Mineral Metabolism*. 2013 Jul;31(4):423-32. doi: 10.1007/s00774-013-0426-0. PMID: 23420298.
205. McLeod KM, Johnson S, Rasali D, Verma A. Discriminatory performance of the calcaneal quantitative ultrasound and osteoporosis self-assessment tool to select older women for dual-energy x-ray absorptiometry. *J Clin Densitom*. 2015 Apr-Jun;18(2):157-64. doi: 10.1016/j.jocd.2015.02.006. PMID: 25937306.
206. Cadarette SM, Jaglal SB, Murray TM, et al. Evaluation of decision rules for referring women for bone densitometry by dual-energy x-ray absorptiometry. *JAMA*. 2001 Jul 4;286(1):57-63. doi: 10.1001/jama.286.1.57. PMID: 11434827.
207. Chan SP, Teo CC, Ng SA, et al. Validation of various osteoporosis risk indices in elderly Chinese females in Singapore. *Osteoporos Int*. 2006;17(8):1182-8. doi: 10.1007/s00198-005-0051-4. PMID: 16699739.
208. Cook RB, Collins D, Tucker J, Zioupos P. Comparison of questionnaire and quantitative ultrasound techniques as screening tools for DXA. *Osteoporos Int*. 2005 Dec;16(12):1565-75. doi: 10.1007/s00198-005-1864-x. PMID: 15883661.
209. Gourlay ML, Miller WC, Richy F, et al. Performance of osteoporosis risk assessment tools in postmenopausal women aged 45-64 years. *Osteoporos Int*. 2005 Aug;16(8):921-7. doi: 10.1007/s00198-004-1775-2. PMID: 16028108.
210. Richy F, Gourlay M, Ross PD, et al. Validation and comparative evaluation of the osteoporosis self-assessment tool (OST) in a Caucasian population from Belgium. *Qjm*. 2004 Jan;97(1):39-46. doi: 10.1093/qjmed/hch002. PMID: 14702510.
211. Ben Sedrine W, Devogelaer JP, Kaufman JM, et al. Evaluation of the simple calculated osteoporosis risk estimation (SCORE) in a sample of white women from Belgium. *Bone*. 2001 Oct;29(4):374-80. doi: 10.1016/s8756-3282(01)00583-x. PMID: 11595621.

212. Kung AW, Ho AY, Ross PD, Reginster JY. Development of a clinical assessment tool in identifying Asian men with low bone mineral density and comparison of its usefulness to quantitative bone ultrasound. *Osteoporos Int*. 2005 Jul;16(7):849-55. doi: 10.1007/s00198-004-1778-z. PMID: 15611839.
213. Kung AW, Ho AY, Sedrine WB, et al. Comparison of a simple clinical risk index and quantitative bone ultrasound for identifying women at increased risk of osteoporosis. *Osteoporos Int*. 2003 Sep;14(9):716-21. doi: 10.1007/s00198-003-1428-x. PMID: 12897978.
214. Martinez-Aguila D, Gomez-Vaquero C, Rozadilla A, et al. Decision rules for selecting women for bone mineral density testing: application in postmenopausal women referred to a bone densitometry unit. *J Rheumatol*. 2007 Jun;34(6):1307-12. PMID: 17552058.
215. Park HM, Sedrine WB, Reginster JY, Ross PD. Korean experience with the OSTA risk index for osteoporosis: a validation study. *J Clin Densitom*. 2003 Fall;6(3):247-50. doi: 10.1385/jcd:6:3:247. PMID: 14514994.
216. Zimering MB, Shin JJ, Shah J, et al. Validation of a novel risk estimation tool for predicting low bone density in Caucasian and African American men veterans. *J Clin Densitom*. 2007 Jul-Sep;10(3):289-97. doi: 10.1016/j.jocd.2007.03.001. PMID: 17459748.
217. Harrison EJ, Adams JE. Application of a triage approach to peripheral bone densitometry reduces the requirement for central DXA but is not cost effective. *Calcif Tissue Int*. 2006 Oct;79(4):199-206. doi: 10.1007/s00223-005-0302-6. PMID: 16969598.
218. Lynn HS, Woo J, Leung PC, et al. An evaluation of osteoporosis screening tools for the osteoporotic fractures in men (MrOS) study. *Osteoporos Int*. 2008 Jul;19(7):1087-92. doi: 10.1007/s00198-007-0553-3. PMID: 18239959.
219. Sinnott B, Kukreja S, Barengolts E. Utility of screening tools for the prediction of low bone mass in African American men. *Osteoporos Int*. 2006;17(5):684-92. doi: 10.1007/s00198-005-0034-5. PMID: 16523248.
220. Bansal S, Pecina JL, Merry SP, et al. US Preventative Services Task Force FRAX threshold has a low sensitivity to detect osteoporosis in women ages 50-64 years. *Osteoporos Int*. 2015 Apr;26(4):1429-33. doi: 10.1007/s00198-015-3026-0. PMID: 25614141.
221. Adler RA, Tran MT, Petkov VI. Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men. *Mayo Clin Proc*. 2003 Jun;78(6):723-7. doi: 10.4065/78.6.723. PMID: 12934782.
222. Brenneman SK, Lacroix AZ, Buist DS, et al. Evaluation of decision rules to identify postmenopausal women for intervention related to osteoporosis. *Dis Manag*. 2003 Fall;6(3):159-68. doi: 10.1089/109350703322425509. PMID: 14570384.
223. Cadarette SM, McIsaac WJ, Hawker GA, et al. The validity of decision rules for selecting women with primary osteoporosis for bone mineral density testing. *Osteoporos Int*. 2004 May;15(5):361-6. doi: 10.1007/s00198-003-1552-7. PMID: 14730421.
224. Cass AR, Shepherd AJ, Carlson CA. Osteoporosis risk assessment and ethnicity: validation and comparison of 2 clinical risk stratification instruments. *J Gen Intern Med*. 2006 Jun;21(6):630-5. doi: 10.1111/j.1525-1497.2006.00459.x. PMID: 16808748.
225. D'Amelio P, Tamone C, Pluviano F, et al. Effects of lifestyle and risk factors on bone mineral density in a cohort of Italian women: suggestion for a new decision rule. *Calcif Tissue Int*. 2005 Aug;77(2):72-8. doi: 10.1007/s00223-004-0253-3. PMID: 16059776.
226. Mauck KF, Cuddihy MT, Atkinson EJ, Melton LJ, 3rd. Use of clinical prediction rules in detecting osteoporosis in a population-based sample of postmenopausal women. *Arch Intern Med*. 2005 Mar 14;165(5):530-6. doi: 10.1001/archinte.165.5.530. PMID: 15767529.
227. Nguyen TV, Center JR, Pocock NA, Eisman JA. Limited utility of clinical indices for the prediction of symptomatic fracture risk in postmenopausal women. *Osteoporos Int*. 2004 Jan;15(1):49-55. doi: 10.1007/s00198-003-1511-3. PMID: 14593453.
228. Rud B, Jensen JE, Mosekilde L, et al. Performance of four clinical screening tools to select peri- and early postmenopausal women for dual X-ray absorptiometry. *Osteoporos Int*. 2005 Jul;16(7):764-72. doi: 10.1007/s00198-004-1748-5. PMID: 15986263.
229. Oh SM, Song BM, Nam BH, et al. Development and validation of osteoporosis risk-assessment model for Korean men. *Yonsei Med J*. 2016 Jan;57(1):187-96. doi: 10.3349/ymj.2016.57.1.187. PMID: 26632400.
230. Gourlay ML, Powers JM, Lui LY, Ensrud KE. Clinical performance of osteoporosis risk assessment tools in women aged 67 years and older. *Osteoporos Int*. 2008 Aug;19(8):1175-83. doi: 10.1007/s00198-007-0555-1. PMID: 18219434.

231. Geusens P, Hochberg MC, van der Voort DJ, et al. Performance of risk indices for identifying low bone density in postmenopausal women. *Mayo Clin Proc*. 2002 Jul;77(7):629-37. doi: 10.4065/77.7.629. PMID: 12108600.
232. Cass AR, Shepherd AJ, Asirot R, et al. Comparison of the Male Osteoporosis Risk Estimation Score (MORES) with FRAX in identifying men at risk for osteoporosis. *Ann Fam Med*. 2016 Jul;14(4):365-9. doi: 10.1370/afm.1945. PMID: 27401426.
233. Chen SJ, Chen YJ, Cheng CH, et al. Comparisons of different screening tools for identifying fracture/osteoporosis risk among community-dwelling older people. *Medicine (Baltimore)*. 2016 May;95(20):e3415. doi: 10.1097/md.0000000000003415. PMID: 27196447.
234. Pecina JL, Romanovsky L, Merry SP, et al. Comparison of clinical risk tools for predicting osteoporosis in women ages 50-64. *J Am Board Fam Med*. 2016 Mar-Apr;29(2):233-9. doi: 10.3122/jabfm.2016.02.150237. PMID: 26957380.
235. Diem SJ, Peters KW, Gourlay ML, et al. Screening for osteoporosis in older men: operating characteristics of proposed strategies for selecting men for BMD testing. *J Gen Intern Med*. 2017 Nov;32(11):1235-41. doi: 10.1007/s11606-017-4153-4. PMID: 28815485.
236. Jiang X, Good LE, Spinka R, Schnatz PF. Osteoporosis screening in postmenopausal women aged 50-64 years: BMI alone compared with current screening tools. *Maturitas*. 2016 Jan;83:59-64. doi: 10.1016/j.maturitas.2015.09.009. PMID: 26471931.
237. Williams ST, Lawrence PT, Miller KL, et al. A comparison of electronic and manual fracture risk assessment tools in screening elderly male US veterans at risk for osteoporosis. *Osteoporos Int*. 2017 Nov;28(11):3107-11. doi: 10.1007/s00198-017-4172-3. PMID: 28756457.
238. Hamdy RC, Seier E, Whalen K, et al. FRAX calculated without BMD does not correctly identify Caucasian men with densitometric evidence of osteoporosis. *Osteoporos Int*. 2018 Apr;29(4):947-52. doi: 10.1007/s00198-017-4368-6. PMID: 29397409.
239. Inderjeeth CA, Raymond WD. Case finding with GARVAN fracture risk calculator in primary prevention of fragility fractures in older people. *Archives of Gerontology & Geriatrics*. 2020 Jan-Feb;86:103940. doi: 10.1016/j.archger.2019.103940. PMID: 31525558.
240. Chang SF, Yang RS. Optimal analysis to discriminate males' osteoporosis with simple physiological indicators: a cutoff point study. *Am J Mens Health*. 2016 Nov;10(6):487-94. doi: 10.1177/1557988315574857. PMID: 25766506.
241. Shuler FD, Scott K, Wilson-Byrne T, et al. Improving rural bone health and minimizing fracture risk in West Virginia: validation of the World Health Organization FRAX assessment tool as a phone survey for osteoporosis detection. *W V Med J*. 2016 May-Jun;112(3):84-8. PMID: 27301160.
242. Christodoulou S, Drosos GI, Ververidis A, et al. Risk assessment instruments for screening bone mineral density in a Mediterranean population. *World J Orthop*. 2016 Sep 18;7(9):577-83. doi: 10.5312/wjo.v7.i9.577. PMID: 27672571.
243. Erjiang E, Wang T, Yang L, et al. Utility of osteoporosis self-assessment tool as a screening tool for osteoporosis in Irish men and women: results of the DXA-HIP project. *J Clin Densitom*. 2021 Oct-Dec;24(4):516-26. doi: 10.1016/j.jocd.2021.03.003. PMID: 33789806.
244. Moon JH, Kim LO, Kim HJ, Kong MH. Evaluation of the predictive index for osteoporosis as a clinical tool to identify the risk of osteoporosis in Korean men by using the Korea National Health and Nutrition Examination Survey data. *Korean J Fam Med*. 2016 Nov;37(6):346-50. doi: 10.4082/kjfm.2016.37.6.346. PMID: 27900073.
245. Toh LS, Lai PSM, Wu DBC, et al. A comparison of 6 osteoporosis risk assessment tools among postmenopausal women in Kuala Lumpur, Malaysia. *Osteoporos Sarcopenia*. 2019 Sep;5(3):87-93. doi: 10.1016/j.afos.2019.09.001. PMID: 31728426.
246. Berry SD, Samelson EJ, Pencina MJ, et al. Repeat bone mineral density screening and prediction of hip and major osteoporotic fracture. *JAMA*. 2013 Sep 25;310(12):1256-62. doi: 10.1001/jama.2013.277817. PMID: 24065012.
247. Hillier TA, Stone KL, Bauer DC, et al. Evaluating the value of repeat bone mineral density measurement and prediction of fractures in older women: the study of osteoporotic fractures. *Arch Intern Med*. 2007 Jan 22;167(2):155-60. doi: 10.1001/archinte.167.2.155. PMID: 17242316.
248. Leslie WD, Majumdar SR, Morin SN, et al. Change in Trabecular Bone Score (TBS) with antiresorptive therapy does not predict fracture in women: the Manitoba BMD Cohort. *Journal of Bone & Mineral Research*. 2017 Mar;32(3):618-23. doi: 10.1002/jbmr.3054. PMID: 27933656.

249. Crandall CJ, Larson J, Wright NC, et al. Serial Bone Density Measurement and Incident Fracture Risk Discrimination in Postmenopausal Women. *JAMA Intern Med*. 2020 Sep 1;180(9):1232-40. doi: 10.1001/jamainternmed.2020.2986. PMID: 32730575.
250. Ensrud KE, Lui LY, Crandall CJ, et al. Repeat Bone Mineral Density Screening Measurement and Fracture Prediction in Older Men: A Prospective Cohort Study. *Journal of Clinical Endocrinology & Metabolism*. 2022 Aug 18;107(9):e3877-e86. doi: 10.1210/clinem/dgac324. PMID: 35587517.
251. Boonen S, Reginster JY, Kaufman JM, et al. Fracture risk and zoledronic acid therapy in men with osteoporosis. *N Engl J Med*. 2012 Nov 1;367(18):1714-23. doi: 10.1056/NEJMoa1204061. PMID: 23113482.
252. Ascott-Evans BH, Guanabens N, Kivinen S, et al. Alendronate prevents loss of bone density associated with discontinuation of hormone replacement therapy: a randomized controlled trial. *Arch Intern Med*. 2003 Apr 14;163(7):789-94. doi: 10.1001/archinte.163.7.789. PMID: 12695269.
253. Chesnut CH, 3rd, McClung MR, Ensrud KE, et al. Alendronate treatment of the postmenopausal osteoporotic woman: effect of multiple dosages on bone mass and bone remodeling. *Am J Med*. 1995 Aug;99(2):144-52. doi: 10.1016/s0002-9343(99)80134-x. PMID: 7625419.
254. Cummings SR, Black DM, Thompson DE, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA*. 1998 Dec 23-30;280(24):2077-82. doi: 10.1001/jama.280.24.2077. PMID: 9875874.
255. Quandt SA, Thompson DE, Schneider DL, et al. Effect of alendronate on vertebral fracture risk in women with bone mineral density T scores of -1.6 to -2.5 at the femoral neck: the Fracture Intervention Trial. *Mayo Clin Proc*. 2005 Mar;80(3):343-9. doi: 10.4065/80.3.343. PMID: 15757015.
256. Liberman UA, Weiss SR, Broll J, et al. Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. the Alendronate Phase III Osteoporosis Treatment Study Group. *N Engl J Med*. 1995 Nov 30;333(22):1437-43. doi: 10.1056/nejm199511303332201. PMID: 7477143.
257. McClung MR, Geusens P, Miller PD, et al. Effect of risedronate on the risk of hip fracture in elderly women. Hip Intervention Program Study Group. *N Engl J Med*. 2001 Feb 1;344(5):333-40. doi: 10.1056/nejm200102013440503. PMID: 11172164.
258. Mortensen L, Charles P, Bekker PJ, et al. Risedronate increases bone mass in an early postmenopausal population: two years of treatment plus one year of follow-up. *Journal of Clinical Endocrinology & Metabolism*. 1998 Feb;83(2):396-402. doi: 10.1210/jcem.83.2.4586. PMID: 9467547.
259. Pols HA, Felsenberg D, Hanley DA, et al. Multinational, placebo-controlled, randomized trial of the effects of alendronate on bone density and fracture risk in postmenopausal women with low bone mass: results of the FOSIT study. Fosamax International Trial Study Group. *Osteoporos Int*. 1999;9(5):461-8. doi: 10.1007/pl00004171. PMID: 10550467.
260. Reid IR, Brown JP, Burckhardt P, et al. Intravenous zoledronic acid in postmenopausal women with low bone mineral density. *N Engl J Med*. 2002 Feb 28;346(9):653-61. doi: 10.1056/NEJMoa011807. PMID: 11870242.
261. Valimaki MJ, Farrerons-Minguella J, Halse J, et al. Effects of risedronate 5 mg/d on bone mineral density and bone turnover markers in late-postmenopausal women with osteopenia: a multinational, 24-month, randomized, double-blind, placebo-controlled, parallel-group, phase III trial. *Clin Ther*. 2007 Sep;29(9):1937-49. doi: 10.1016/j.clinthera.2007.09.017. PMID: 18035193.
262. Grey A, Bolland M, Wattie D, et al. Prolonged antiresorptive activity of zoledronate: a randomized, controlled trial. *Journal of Bone & Mineral Research*. 2010 Oct;25(10):2251-5. doi: 10.1002/jbmr.103. PMID: 20499349.
263. Grey A, Bolland MJ, Wattie D, et al. The antiresorptive effects of a single dose of zoledronate persist for two years: a randomized, placebo-controlled trial in osteopenic postmenopausal women. *Journal of Clinical Endocrinology & Metabolism*. 2009 Feb;94(2):538-44. doi: 10.1210/jc.2008-2241. PMID: 19050050.
264. Ravn P, Clemmesen B, Riis BJ, Christiansen C. The effect on bone mass and bone markers of different doses of ibandronate: a new bisphosphonate for prevention and treatment of postmenopausal osteoporosis: a 1-year, randomized, double-blind, placebo-controlled dose-finding study. *Bone*. 1996 Nov;19(5):527-33. doi: 10.1016/s8756-3282(96)00229-3. PMID: 8922653.
265. Riis BJ, Ise J, von Stein T, et al. Ibandronate: a comparison of oral daily dosing versus intermittent dosing in postmenopausal osteoporosis. *Journal of Bone & Mineral Research*. 2001 Oct;16(10):1871-8. doi: 10.1359/jbmr.2001.16.10.1871. PMID: 11585352.

266. Reginster JY, Wilson KM, Dumont E, et al. Monthly oral ibandronate is well tolerated and efficacious in postmenopausal women: results from the monthly oral pilot study. *Journal of Clinical Endocrinology & Metabolism*. 2005 Sep;90(9):5018-24. doi: 10.1210/jc.2004-1750. PMID: 15972582.

267. Hosking D, Adami S, Felsenberg D, et al. Comparison of change in bone resorption and bone mineral density with once-weekly alendronate and daily risedronate: a randomised, placebo-controlled study. *Current Medical Research & Opinion*. 2003;19(5):383-94. doi: 10.1185/030079903125002009. PMID: 13678475.

268. Tucci JR, Tonino RP, Emkey RD, et al. Effect of three years of oral alendronate treatment in postmenopausal women with osteoporosis. *Am J Med*. 1996 Nov;101(5):488-501. doi: 10.1016/s0002-9343(96)00282-3. PMID: 8948272.

269. Reid IR, Horne AM, Mihov B, et al. Fracture prevention with zoledronate in older women with osteopenia. *N Engl J Med*. 2018 Dec 20;379(25):2407-16. doi: 10.1056/NEJMoa1808082. PMID: 30575489.

270. Reid IR, Horne AM, Mihov B, et al. Anti-fracture efficacy of zoledronate in subgroups of osteopenic postmenopausal women: secondary analysis of a randomized controlled trial. *J Intern Med*. 2019 Aug;286(2):221-9. doi: 10.1111/joim.12901. PMID: 30887607.

271. Reid IR, Horne AM, Mihov B, et al. Predictors of fracture in older women with osteopenic hip bone mineral density treated with zoledronate. *Journal of Bone & Mineral Research*. 2021 Jan;36(1):61-6. doi: 10.1002/jbmr.4167. PMID: 32835417.

272. Grey A, Bolland M, Wong S, et al. Low-dose zoledronate in osteopenic postmenopausal women: a randomized controlled trial. *Journal of Clinical Endocrinology & Metabolism*. 2012 Jan;97(1):286-92. doi: 10.1210/jc.2011-2081. PMID: 22072741.

273. Grey A, Bolland M, Mihov B, et al. Duration of antiresorptive effects of low-dose zoledronate in osteopenic postmenopausal women: a randomized, placebo-controlled trial. *Journal of Bone & Mineral Research*. 2014 Jan;29(1):166-72. doi: 10.1002/jbmr.2009. PMID: 23761303.

274. Grey A, Bolland MJ, Horne A, et al. Duration of antiresorptive activity of zoledronate in postmenopausal women with osteopenia: a randomized, controlled multidose trial. *CMAJ*. 2017 Sep 11;189(36):E1130-E6. doi: 10.1503/cmaj.161207. PMID: 28893875.

275. McClung MR, Bolognese MA, Sedarati F, et al. Efficacy and safety of monthly oral ibandronate in the prevention of postmenopausal bone loss. *Bone*. 2009 Mar;44(3):418-22. doi: 10.1016/j.bone.2008.09.011. PMID: 18950736.

276. Bell NH, Bilezikian JP, Bone HG, 3rd, et al. Alendronate increases bone mass and reduces bone markers in postmenopausal African-American women. *Journal of Clinical Endocrinology & Metabolism*. 2002 Jun;87(6):2792-7. doi: 10.1210/jcem.87.6.8575. PMID: 12050252.

277. Bone HG, Downs RW, Jr., Tucci JR, et al. Dose-response relationships for alendronate treatment in osteoporotic elderly women. Alendronate Elderly Osteoporosis Study Centers. *Journal of Clinical Endocrinology & Metabolism*. 1997 Jan;82(1):265-74. doi: 10.1210/jcem.82.1.3682. PMID: 8989272.

278. Orwoll E, Teglbjaerg CS, Langdahl BL, et al. A randomized, placebo-controlled study of the effects of denosumab for the treatment of men with low bone mineral density. *Journal of Clinical Endocrinology & Metabolism*. 2012 Sep;97(9):3161-9. doi: 10.1210/jc.2012-1569. PMID: 22723310.

279. Nakamura T, Matsumoto T, Sugimoto T, Shiraki M. Dose-response study of denosumab on bone mineral density and bone turnover markers in Japanese postmenopausal women with osteoporosis. *Osteoporos Int*. 2012(3):1131-40. doi: 10.1007/s00198-011-1786-8. PMID: CN-00843289.

280. Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*. 2009 Aug 20;361(8):756-65. doi: 10.1056/NEJMoa0809493. PMID: 19671655.

281. Simon JA, Recknor C, Moffett AH, Jr., et al. Impact of denosumab on the peripheral skeleton of postmenopausal women with osteoporosis: bone density, mass, and strength of the radius, and wrist fracture. *Menopause*. 2013 Feb;20(2):130-7. doi: 10.1097/gme.0b013e318267f909. PMID: 23010883.

282. McCloskey EV, Johansson H, Oden A, et al. Denosumab reduces the risk of osteoporotic fractures in postmenopausal women, particularly in those with moderate to high fracture risk as assessed with FRAX. *Journal of Bone & Mineral Research*. 2012 Jul;27(7):1480-6. doi: 10.1002/jbmr.1606. PMID: 22431426.

283. Palacios S, Kalouche-Khalil L, Rizzoli R, et al. Treatment with denosumab reduces secondary fracture risk in women with postmenopausal osteoporosis. *Climacteric*. 2015;18(6):805-12. doi: 10.3109/13697137.2015.1045484. PMID: 26029985.

284. Lewiecki EM, Miller PD, McClung MR, et al. Two-year treatment with denosumab (AMG 162) in a randomized phase 2 study of postmenopausal women with low BMD. *Journal of Bone & Mineral Research*. 2007 Dec;22(12):1832-41. doi: 10.1359/jbmr.070809. PMID: 17708711.
285. Bone HG, Bolognese MA, Yuen CK, et al. Effects of denosumab on bone mineral density and bone turnover in postmenopausal women. *Journal of Clinical Endocrinology & Metabolism*. 2008 Jun;93(6):2149-57. doi: 10.1210/jc.2007-2814. PMID: 18381571.
286. Koh JM, Chung DJ, Chung YS, et al. Assessment of denosumab in Korean postmenopausal women with osteoporosis: randomized, double-blind, placebo-controlled trial with open-label extension. *Yonsei Med J*. 2016 Jul;57(4):905-14. doi: 10.3349/ymj.2016.57.4.905. PMID: 27189284.
287. McClung M, Miller P, Recknor C, et al. Zoledronic acid for the prevention of bone loss in postmenopausal women with low bone mass: a randomized controlled trial. *Obstetrics & Gynecology*. 2009 Nov;114(5):999-1007. doi: 10.1097/AOG.0b013e3181bdce0a. PMID: 20168099.
288. Chapurlat RD, Laroche M, Thomas T, et al. Effect of oral monthly ibandronate on bone microarchitecture in women with osteopenia-a randomized placebo-controlled trial. *Osteoporos Int*. 2013 Jan;24(1):311-20. doi: 10.1007/s00198-012-1947-4. PMID: 22402673.
289. Bauer DC, Black D, Ensrud K, et al. Upper gastrointestinal tract safety profile of alendronate: the fracture intervention trial. *Arch Intern Med*. 2000 Feb 28;160(4):517-25. doi: 10.1001/archinte.160.4.517. PMID: 10695692.
290. Cummings SR, Schwartz AV, Black DM. Alendronate and atrial fibrillation. *N Engl J Med*. 2007 May 3;356(18):1895-6. doi: 10.1056/NEJMc076132. PMID: 17476024.
291. McClung MR, Wasnich RD, Recker R, et al. Oral daily ibandronate prevents bone loss in early postmenopausal women without osteoporosis. *Journal of Bone & Mineral Research*. 2004 Jan;19(1):11-8. doi: 10.1359/jbmr.0301202. PMID: 14753731.
292. Tanko LB, Felsenberg D, Czerwinski E, et al. Oral weekly ibandronate prevents bone loss in postmenopausal women. *J Intern Med*. 2003 Aug;254(2):159-67. doi: 10.1046/j.1365-2796.2003.01174.x. PMID: 12859697.
293. Thiebaud D, Burckhardt P, Kriegbaum H, et al. Three monthly intravenous injections of ibandronate in the treatment of postmenopausal osteoporosis. *Am J Med*. 1997 Oct;103(4):298-307. doi: 10.1016/s0002-9343(97)00249-0. PMID: 9382122.
294. Johnell O, Scheele WH, Lu Y, et al. Additive effects of raloxifene and alendronate on bone density and biochemical markers of bone remodeling in postmenopausal women with osteoporosis. *Journal of Clinical Endocrinology & Metabolism*. 2002 Mar;87(3):985-92. doi: 10.1210/jcem.87.3.8325. PMID: 11889149.
295. Shiraki M, Fukunaga M, Kushida K, et al. A double-blind dose-ranging study of risedronate in Japanese patients with osteoporosis (a study by the Risedronate Late Phase II Research Group). *Osteoporos Int*. 2003 May;14(3):225-34. doi: 10.1007/s00198-002-1369-9. PMID: 12730746.
296. Adachi JD, Faraawi RY, O'Mahony MF, et al. Upper gastrointestinal tolerability of alendronate sodium monohydrate 10 mg once daily in postmenopausal women: a 12-week, randomized, double-blind, placebo-controlled, exploratory study. *Clin Ther*. 2009 Aug;31(8):1747-53. doi: 10.1016/j.clinthera.2009.08.016. PMID: 19808133.
297. Cryer B, Binkley N, Simonelli C, et al. A randomized, placebo-controlled, 6-month study of once-weekly alendronate oral solution for postmenopausal osteoporosis. *Am J Geriatr Pharmacother*. 2005 Sep;3(3):127-36. doi: 10.1016/s1543-5946(05)80019-4. PMID: 16257815.
298. Eisman JA, Rizzoli R, Roman-Ivorra J, et al. Upper gastrointestinal and overall tolerability of alendronate once weekly in patients with osteoporosis: results of a randomized, double-blind, placebo-controlled study. *Current Medical Research & Opinion*. 2004 May;20(5):699-705. doi: 10.1185/030079904125003548. PMID: 15140336.
299. Greenspan S, Field-Munves E, Tonino R, et al. Tolerability of once-weekly alendronate in patients with osteoporosis: a randomized, double-blind, placebo-controlled study. *Mayo Clin Proc*. 2002 Oct;77(10):1044-52. doi: 10.4065/77.10.1044. PMID: 12374248.
300. Greenspan SL, Resnick NM, Parker RA. Combination therapy with hormone replacement and alendronate for prevention of bone loss in elderly women: a randomized controlled trial. *JAMA*. 2003 May 21;289(19):2525-33. doi: 10.1001/jama.289.19.2525. PMID: 12759324.
301. Reid IR, Horne AM, Mihov B, et al. Effects of zoledronate on cancer, cardiac events, and mortality in osteopenic older women. *Journal of Bone & Mineral Research*. 2020 Jan;35(1):20-7. doi: 10.1002/jbmr.3860. PMID: 31603996.

302. Devogelaer JP, Broll H, Correa-Rotter R, et al. Oral alendronate induces progressive increases in bone mass of the spine, hip, and total body over 3 years in postmenopausal women with osteoporosis. *Bone*. 1996 Feb;18(2):141-50. doi: 10.1016/8756-3282(95)00436-x. PMID: 8833208.
303. Watts NB, Roux C, Modlin JF, et al. Infections in postmenopausal women with osteoporosis treated with denosumab or placebo: coincidence or causal association? *Osteoporos Int*. 2012 Jan;23(1):327-37. doi: 10.1007/s00198-011-1755-2. PMID: 21892677.
304. McClung MR, Lewiecki EM, Cohen SB, et al. Denosumab in postmenopausal women with low bone mineral density. *N Engl J Med*. 2006 Feb 23;354(8):821-31. doi: 10.1056/NEJMoa044459. PMID: 16495394.
305. Pazianas M, Abrahamsen B, Eiken PA, et al. Reduced colon cancer incidence and mortality in postmenopausal women treated with an oral bisphosphonate—Danish National Register Based Cohort Study. *Osteoporos Int*. 2012 Nov;23(11):2693-701. doi: 10.1007/s00198-012-1902-4. PMID: 22392160.
306. Rubin KH, Möller S, Choudhury A, et al. Cardiovascular and skeletal safety of zoledronic acid in osteoporosis observational, matched cohort study using Danish and Swedish health registries. *Bone*. 2020 May;134:115296. doi: 10.1016/j.bone.2020.115296. PMID: 32097760.
307. Lee YK, Byun DW, Jung SM, et al. Bisphosphonates use and risk of subtrochanteric and diaphyseal femur fractures in Korea: results from the National Claim Registry. *Calcif Tissue Int*. 2019 Mar;104(3):313-9. doi: 10.1007/s00223-018-0493-2. PMID: 30446771.
308. Kern LM, Powe NR, Levine MA, et al. Association between screening for osteoporosis and the incidence of hip fracture. *Ann Intern Med*. 2005 Feb 1;142(3):173-81. PMID: 15684205.
309. Barr RJ, Stewart A, Torgerson DJ, Reid DM. Population screening for osteoporosis risk: a randomised control trial of medication use and fracture risk. *Osteoporos Int*. 2010 Apr;21(4):561-8. doi: 10.1007/s00198-009-1007-x [doi]. PMID: 19565176.
310. Azagra R, Zwart M, Encabo G, et al. Rationale of the Spanish FRAX model in decision-making for predicting osteoporotic fractures: an update of FRIDEX cohort of Spanish women. *BMC Musculoskelet Disord*. 2016 Jun 17;17:262. doi: 10.1186/s12891-016-1096-6. PMID: 27317560.
311. Jain RK, Weiner M, Polley E, et al. Electronic Health Records (EHRs) Can Identify Patients at High Risk of Fracture but Require Substantial Race Adjustments to Currently Available Fracture Risk Calculators. *J Gen Intern Med*. 2023 Dec;38(16):3451-9. doi: 10.1007/s11606-023-08347-5. PMID: 37715097.
312. Pluskiewicz W, Werner A, Bach M, et al. Optimal fracture prediction thresholds for therapy onset, established from FRAX and Garvan algorithms: a longitudinal observation of the population representative female cohort from the RAC-OST-POL Study. *Arch Osteoporos*. 2023 Nov 16;18(1):136. doi: 10.1007/s11657-023-01346-3. PMID: 37973685.
313. Zwart M, Azagra-Ledesma R, Saez M, et al. Predictive capacity of FRAX in a spanish region with a hip fracture rate close to the national mean. *BMC Musculoskelet Disord*. 2023 Jul 15;24(1):577. doi: 10.1186/s12891-023-06670-w. PMID: 37454058.
314. Theriault G, Reynolds D, Pillay JJ, et al. Expanding the measurement of overdiagnosis in the context of disease precursors and risk factors. *BMJ Evid Based Med*. 2023 Jan 10. doi: 10.1136/bmjebm-2022-112117. PMID: 36627178.
315. Fryar CD, Gu Q, Ogden CL. Anthropometric reference data for children and adults: United States 2007–2010. Vital and Health Statistics. Series 11, Number 252. Washington, DC: U.S. Department of Health and Human Services; October. 2012. http://www.cdc.gov/nchs/data/series/sr_11/sr11_252.pdf
316. Centre for Metabolic Bone Diseases. FRAX calculation tool. UK: University of Sheffield. https://www.sheffield.ac.uk/FRAX/tool.aspx?country=9. Accessed 14 Nov 2022.
317. Gourlay ML, Fine JP, Preisser JS, et al. Bone-density testing interval and transition to osteoporosis in older women. *N Engl J Med*. 2012 Jan 19;366(3):225-33. doi: 10.1056/NEJMoa1107142 [doi]. PMID: 22256806.
318. Ayers C, Kansagara D, Lazur B, et al. Effectiveness and safety of treatments to prevent fractures in people with low bone mass or primary osteoporosis: a living systematic review and network meta-analysis for the American College of Physicians. *Ann Intern Med*. 2023 Jan 3. doi: 10.7326/m22-0684. PMID: 36592455.
319. Qaseem A, Hicks LA, Etxeandia-Ikobaltzeta I, et al. Pharmacologic treatment of primary osteoporosis or low bone mass to prevent fractures in adults: a living clinical guideline from the American College of Physicians. *Ann Intern Med*. 2023 Jan 3. doi: 10.7326/m22-1034. PMID: 36592456.
320. U. S. Preventive Services Task Force. Interventions to prevent falls in community-dwelling older adults: US Preventive Services Task Force recommendation statement. *JAMA*. 2024 Jul 2;332(1):51-7. doi: 10.1001/jama.2024.8481. PMID: 38833246.

321. Bala Y, Chapurlat R, Cheung AM, et al. Risedronate slows or partly reverses cortical and trabecular microarchitectural deterioration in postmenopausal women. *Journal of Bone & Mineral Research*. 2014 Feb;29(2):380-8. doi: 10.1002/jbmr.2101 [doi]. PMID: 24115129.
322. Kanis JA. Welcome to FRAX®. UK: World Health Organization Collaborating Centre for Metabolic Bone Diseases, University of Sheffield, UK n.d. https://www.shef.ac.uk/FRAX/. Accessed 14 Nov 2022.
323. Kanis JA, McCloskey E, Johansson H, et al. FRAX((R)) with and without bone mineral density. *Calcif Tissue Int*. 2012 Jan;90(1):1-13. doi: 10.1007/s00223-011-9544-7. PMID: 22057815.
324. Reid HW, Selvan B, Batch BC, Lee RH. The break in FRAX: Equity concerns in estimating fracture risk in racial and ethnic minorities. *J Am Geriatr Soc*. 2021 Sep;69(9):2692-5. doi: 10.1111/jgs.17316. PMID: 34124772.
325. National Institute for Health and Care Excellence. Osteoporosis: assessing the risk of fragility fracture. Clinical guideline [CG146]. Chapter 1. Guidance London: National Institute for Health and Care Excellence; 2012. https://www.nice.org.uk/guidance/cg146/chapter/1-Guidance#targeting-risk-assessment. Accessed 20 May 2021.
326. Eastell R, Rosen CJ, Black DM, et al. Pharmacological management of osteoporosis in postmenopausal women: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology & Metabolism*. 2019 May 1;104(5):1595-622. doi: 10.1210/jc.2019-00221. PMID: 30907953.
327. Morin SN, Feldman S, Funnell L, et al. Clinical practice guideline for management of osteoporosis and fracture prevention in Canada: 2023 update. *Cmaj*. 2023 Oct 10;195(39):E1333-e48. doi: 10.1503/cmaj.221647. PMID: 37816527.
328. American College of Obstetricians and Gynecologists. Osteoporosis Prevention, Screening, and Diagnosis: ACOG Clinical Practice Guideline No. 1. *Obstetrics & Gynecology*. 2021 Sep 1;138(3):494-506. doi: 10.1097/aog.0000000000004514. PMID: 34412075.
329. American Academy of Family Physicians. Osteoporosis screening to prevent fractures. Leawood, KS: American Academy of Family Physicians; 2021. https://www.aafp.org/family-physician/patient-care/clinical-recommendations/all-clinical-recommendations/osteoporosis.html. Accessed 20 May 2021.
330. Lim LS, Hoeksema LJ, Sherin K, ACPM Prevention Practice Committee. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. *Am J Prev Med*. 2009 Apr;36(4):366-75. doi: 10.1016/j.amepre.2009.01.013. PMID: 19285200.
331. Watts NB, Adler RA, Bilezikian JP, et al. Osteoporosis in men: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology & Metabolism*. 2012 Jun;97(6):1802-22. doi: 10.1210/jc.2011-3045. PMID: 22675062.
332. Shoback D, Rosen CJ, Black DM, et al. Pharmacological management of osteoporosis in postmenopausal women: an Endocrine Society guideline update. *Journal of Clinical Endocrinology & Metabolism*. 2020 Mar 1;105(3). doi: 10.1210/clinem/dgaa048. PMID: 32068863.
333. Gregson CL, Armstrong DJ, Bowden J, et al. UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos*. 2022 Apr 5;17(1):58. doi: 10.1007/s11657-022-01061-5. PMID: 35378630.
334. North American Menopause Society. Management of osteoporosis in postmenopausal women: the 2021 position statement of The North American Menopause Society. *Menopause*. 2021 Sep 1;28(9):973-97. doi: 10.1097/gme.0000000000001831. PMID: 34448749.
335. UK National Screening Committee. Adult screening programme: osteoporosis. 2019. https://view-health-screening-recommendations.service.gov.uk/osteoporosis/. Accessed February 8, 2023.
336. Colón-Emeric CS, Colón-Emeric CS, Pieper CF, et al. Can historical and functional risk factors be used to predict fractures in community-dwelling older adults? Development and validation of a clinical tool. *Osteoporosis International*. 2002 2002/12/01;13(12):955-61. doi: 10.1007/s001980200133.
337. American Bone Health. 10-year fracture risk calculator. Raleigh, NC: American Bone Health; n.d. https://americanbonehealth.org/about-the-abh-fracture-risk-calculator/. Accessed 30 Nov 2022.
338. Kanis JA, Oden A, Johansson H, et al. FRAX® and its applications to clinical practice. *Bone*. 2009 2009/05/01;44(5):734-43. doi: 10.1016/j.bone.2009.01.373.
339. Kanis JA, Johansson H, Harvey NC, McCloskey EV. A brief history of FRAX. *Arch Osteoporos*. 2018 Oct 31;13(1):118. doi: 10.1007/s11657-018-0510-0. PMID: 30382424.
340. Kanis JA, Johansson H, Oden A, et al. The effects of a FRAX revision for the USA. *Osteoporos Int*. 2010 Jan;21(1):35-40. doi: 10.1007/s00198-009-1033-8 [doi]. PMID: 19705047.
341. Kanis JA, McCloskey EV, Johansson H, et al. Development and use of FRAX in osteoporosis. *Osteoporos Int*. 2010 Jun;21 Suppl 2:S407-13. doi: 10.1007/s00198-010-1253-y [doi]. PMID: 20464374.

342. Melton LJ, 3rd, Crowson CS, O'Fallon WM. Fracture incidence in Olmsted County, Minnesota: comparison of urban with rural rates and changes in urban rates over time. *Osteoporos Int*. 1999;9(1):29-37. doi: 10.1007/s001980050113. PMID: 10367027.
343. Kanis JA, Cooper C, Dawson-Hughes B, et al. FRAX and ethnicity. *Osteoporos Int*. 2020 Nov;31(11):2063-7. doi: 10.1007/s00198-020-05631-6. PMID: 32888046.
344. Lewiecki EM, Wright NC, Singer AJ. Racial disparities, FRAX, and the care of patients with osteoporosis. *Osteoporos Int*. 2020 Nov;31(11):2069-71. doi: 10.1007/s00198-020-05655-y. PMID: 32980922.
345. Doubeni CA, Simon M, Krist AH. Addressing systemic racism through clinical preventive service recommendations from the US Preventive Services Task Force. *Jama*. 2021 Feb 16;325(7):627-8. doi: 10.1001/jama.2020.26188. PMID: 33492333.
346. Wright NC, Chen L, Saag KG, et al. Racial disparities exist in outcomes after major fragility fractures. *J Am Geriatr Soc*. 2020 Aug;68(8):1803-10. doi: 10.1111/jgs.16455. PMID: 32337717.
347. Borak J, Fiellin M, Chemerynski S. Who is Hispanic? Implications for epidemiologic research in the United States. *Epidemiology*. 2004 2004;15(2):240-4.
348. ClinRisk. QFracture-2016 risk calculator. 2016.
349. Scotland's Census. Scotland's census at a glance: Ethnic groups. 2021. https://www.scotlandscensus.gov.uk/census-results/at-a-glance/ethnicity/. Accessed 13 Jan, 2022.
350. Northern Ireland Statistics and Research Agency. 2011 Census—key statistics for northern ireland. 2017. https://www.nisra.gov.uk/publications/2011-census-key-statistics-northern-ireland. Accessed 13 Jan 2023.
351. Aspinall PJ. Who is 'Black African' in Britain? Challenges to official categorisation of the Sub-Saharan African origin population. *African Identities*. 2011 2011/02;9(1):33-48. doi: 10.1080/14725843.2011.530443.
352. Landman J, Cruickshank J. A review of ethnicity, health and nutrition-related diseases in relation to migration in the United Kingdom. *Public Health Nutri*. 2001 2001/04;4(2b):647-57. doi: 10.1079/PHN2001148.
353. Araujo AB, Travison TG, Harris SS, et al. Race/ethnic differences in bone mineral density in men. *Osteoporos Int*. 2007 2007/07;18(7):943-53. doi: 10.1007/s00198-006-0321-9.
354. Barrett-Connor E, Siris ES, Wehren LE, et al. Osteoporosis and fracture risk in women of different ethnic groups. *J Bone Mine Res*. 2005 2005/02;20(2):185-94. doi: 10.1359/JBMR.041007.
355. Looker AC, Sarafrazi Isfahani N, Fan B, Shepherd JA. Trabecular bone scores and lumbar spine bone mineral density of US adults: comparison of relationships with demographic and body size variables. *Osteoporos Int*. 2016 2016/08/01;27(8):2467-75. doi: 10.1007/s00198-016-3550-6.
356. Nam HS, Shin MH, Zmuda JM, et al. Race/ethnic differences in bone mineral densities in older men. *Osteoporos Int*. 2010 2010/12;21(12):2115-23. doi: 10.1007/s00198-010-1188-3.
357. Nam H-S, Kweon S-S, Choi J-S, et al. Racial/ethnic differences in bone mineral density among older women. *J Bone Mine Res*. 2013 2013/03/01;31(2):190-8. doi: 10.1007/s00774-012-0402-0.
358. Nelson Da, Jacobsen G, Barondess Da, Parfitt Am. Ethnic differences in regional bone density, hip axis length, and lifestyle variables among healthy black and white men. *J Bone Mine Res*. 1995 1995;10(5):782-7. doi: 10.1002/jbmr.5650100515.
359. Noel SE, Mangano KM, Griffith JL, et al. Prevalence of osteoporosis and low bone mass among Puerto Rican older adults. *Journal of Bone & Mineral Research*. 2018 Mar;33(3):396-403. doi: 10.1002/jbmr.3315. PMID: 29044768.
360. Tracy JK, Meyer WA, Flores RH, et al. Racial differences in rate of decline in bone mass in older men: the Baltimore men's osteoporosis study. *J Bone Mine Res*. 2005 2005/07;20(7):1228-34. doi: 10.1359/JBMR.050310.
361. George A, Tracy JK, Meyer WA, et al. Racial differences in bone mineral density in older men. *J Bone Mine Res*. 2003 2003/12;18(12):2238-44. doi: 10.1359/jbmr.2003.18.12.2238.
362. Popp KL, Hughes JM, Martinez-Betancourt A, et al. Bone mass, microarchitecture and strength are influenced by race/ethnicity in young adult men and women. *Bone*. 2017 2017/10;103:200-8. doi: 10.1016/j.bone.2017.07.014.
363. Nelson DA, Barondess DA, Hendrix SL, Beck TJ. Cross-sectional geometry, bone strength, and bone mass in the proximal femur in black and white postmenopausal women. *J Bone Mine Res*. 2000 2000;15(10):1992-7. doi: 10.1359/jbmr.2000.15.10.1992.
364. Peacock M, Buckwalter KA, Persohn S, et al. Race and sex differences in bone mineral density and geometry at the femur. *Bone*. 2009 2009/08;45(2):218-25. doi: 10.1016/j.bone.2009.04.236.

365. Misra M, Ackerman KE, Bredella MA, et al. Racial differences in bone microarchitecture and estimated strength at the distal radius and distal tibia in older adolescent girls: a cross-sectional study. *J Racial Ethn Health Disparities*. 2017 2017/08;4(4):587-98. doi: 10.1007/s40615-016-0262-8.
366. Cauley JA, Lui L-Y, Stone KL, et al. Longitudinal study of changes in hip bone mineral density in caucasian and African-American women: bone loss by ethnicity and age. *J Am Geriatr Soc*. 2005 2005/02;53(2):183-9. doi: 10.1111/j.1532-5415.2005.53101.x.
367. Zheng X-y, Zhou Z, Gao Y, et al. Racial differences and factors associated with low femoral neck bone mineral density: an analysis of NHANES 2005–2014 data. *Arch Osteoporos*. 2021 2021/01/06;16(1):9. doi: 10.1007/s11657-020-00850-0.
368. Morton DJ, Barrett-Connor E, Kritz-Silverstein D, et al. Bone mineral density in postmenopausal Caucasian, Filipina, and Hispanic women. *Int J Epidemiol*. 2003 2003/02;32(1):150-6. doi: 10.1093/ije/dyg024.
369. Gong G, Haynatzki G, Haynatzka V, et al. Bone mineral density of recent African immigrants in the United States. *Journal of the National Medical Association*. 2006 2006/05;98(5):746-52.
370. Melton Iii LJ, Marquez MA, Achenbach SJ, et al. Variations in bone density among persons of African heritage. *Osteoporos Int*. 2002 2002/07;13(7):551-9. doi: 10.1007/s001980200072.
371. Demeke T, El-Gawad GA, Osmancevic A, et al. Lower bone mineral density in Somali women living in Sweden compared with African-Americans. *Arch Osteoporos*. 2015 2015;10:208. doi: 10.1007/s11657-015-0208-5.
372. Aspray TJ, Prentice A, Cole TJ, et al. Low bone mineral content is common but osteoporotic fractures are rare in elderly rural Gambian women. *J Bone Mine Res*. 1996 1996;11(7):1019-25. doi: 10.1002/jbmr.5650110720.
373. Dibba B, Prentice A, Laskey MA, et al. An investigation of ethnic differences in bone mineral, hip axis length, calcium metabolism and bone turnover between West African and Caucasian adults living in the United Kingdom. *Ann Hum Biol*. 1999 1999/06;26(3):229-42. doi: 10.1080/030144699282732.
374. Daniels ED, Pettifor JM, Schnitzler CM, et al. Differences in mineral homeostasis, volumetric bone mass and femoral neck axis length in black and white South African women. *Osteoporos Int*. 1997 1997;7(2):105-12. doi: 10.1007/BF01623684.
375. Wright NC, Looker AC, Saag KG, et al. The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine. *Journal of Bone & Mineral Research*. 2014 Nov;29(11):2520-6. doi: 10.1002/jbmr.2269. PMID: 24771492.
376. Finkelstein JS, Lee M-LT, Sowers M, et al. Ethnic variation in bone density in premenopausal and early perimenopausal women: effects of anthropometric and lifestyle factors. *Journal of Clinical Endocrinology & Metabolism*. 2002;87(7):3051-56. doi: 10.1210/jcem.87.7.8480.
377. Harvey NC, Glüer CC, Binkley N, et al. Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice. *Bone*. 2015 Sep;78:216-24. doi: 10.1016/j.bone.2015.05.016. PMID: 25988660.
378. Putman MS, Yu EW, Lee H, et al. Differences in skeletal microarchitecture and strength in African-American and white women. *Journal of Bone & Mineral Research*. 2013 Oct;28(10):2177-85. doi: 10.1002/jbmr.1953 [doi]. PMID: 23572415.
379. Theobald TM, Cauley JA, Gluer CC, et al. Black-white differences in hip geometry. Study of Osteoporotic Fractures Research Group. *Osteoporos Int*. 1998 1998;8(1):61-7. doi: 10.1007/s001980050049.
380. Jain RK, Narang DK, Hans D, Vokes TJ. Ethnic differences in trabecular bone score. *J Clin Densitom*. 2017 Apr-Jun;20(2):172-9. doi: 10.1016/j.jocd.2016.04.003. PMID: 27160503.
381. Marshall LM, Zmuda JM, Chan BK, et al. Race and ethnic variation in proximal femur structure and BMD among older men. *J Bone Mine Res*. 2008 2008/01;23(1):121-30. doi: 10.1359/jbmr.070908.
382. Walker MD, Liu XS, Stein E, et al. Differences in bone microarchitecture between postmenopausal Chinese-American and white women. *J Bone Mine Res*. 2011 2011/07;26(7):1392-8. doi: 10.1002/jbmr.352.
383. Leslie WD, Lix LM, Morin SN, et al. Hip axis length is a FRAX - and bone density-independent risk factor for hip fracture in women. *Journal of Clinical Endocrinology & Metabolism*. 2015 2015/05;100(5):2063-70. doi: 10.1210/jc.2014-4390.
384. Cummings SR, Cauley JA, Palermo L, et al. Racial differences in hip axis lengths might explain racial differences in rates of hip fracture. Study of Osteoporotic Fractures Research Group. *Osteoporos Int*. 1994 1994/07;4(4):226-9. doi: 10.1007/BF01623243.

385. Mikhail MB, Vaswani AN, Aloia JF. Racial differences in femoral dimensions and their relation to hip fracture. *Osteoporos Int*. 1996 1996;6(1):22-4. doi: 10.1007/BF01626533.
386. Clark P, Tesoriero LJ, Morton DJ, et al. Hip axis length variation: its correlation with anthropometric measurements in women from three ethnic groups. *Osteoporos Int*. 2008 2008/09;19(9):1301-6. doi: 10.1007/s00198-008-0572-8.
387. Greendale GA, Young JT, Huang M-H, et al. Hip axis length in mid-life Japanese and Caucasian U.S. residents: no evidence for an ethnic difference. *Osteoporos Int*. 2003 2003/06/01;14(4):320-5. doi: 10.1007/s00198-002-1367-y.
388. Shumway-Cook A, Ciol MA, Hoffman J, et al. Falls in the Medicare population: incidence, associated factors, and impact on health care. *Phys Ther*. 2009 2009/04;89(4):324-32. doi: 10.2522/ptj.20070107.
389. Kiely DK, Kim DH, Gross AL, et al. Fall risk is not black and white. *J Health Dispar Res Pract.* 2015 2015;8(3):72-84.
390. Faulkner KA, Cauley JA, Zmuda JM, et al. Ethnic differences in the frequency and circumstances of falling in older community-dwelling women. *J Am Geriatr Soc*. 2005 2005;53(10):1774-9. doi: 10.1111/j.1532-5415.2005.53514.x.
391. Amarnath ALD, Franks P, Robbins JA, et al. Underuse and overuse of osteoporosis screening in a regional health system: A retrospective cohort study. *Journal of General Internal Medicine*. 2015 2015/12;30(12):1733-40. doi: 10.1007/s11606-015-3349-8.
392. Gourlay ML, Callahan LF, Preisser JS, Sloane PD. Osteoporosis preventive care in white and black women in community family medicine settings. *Southern Medical Journal*. 2007 2007/07;100(7):677-82. doi: 10.1097/SMJ.0b013e318038d45e.
393. Hamrick I, Cao Q, Agbafe-Mosley D, Cummings DM. Osteoporosis healthcare disparities in postmenopausal women. *J Womens Health (Larchmt)*. 2012 Dec;21(12):1232-6. doi: 10.1089/jwh.2012.3812 [doi]. PMID: 23140203.
394. Lou Y, Edmonds SW, Jones MP, et al. Predictors of bone mineral density testing among older women on Medicare. *Osteoporos Int*. 2016 Dec;27(12):3577-86. doi: 10.1007/s00198-016-3688-2. PMID: 27358177.
395. Miller RG, Ashar BH, Cohen J, et al. Disparities in osteoporosis screening between at-risk african-american and white women. *J Gen Intern Med*. 2005 2005/09;20(9):847-51. doi: 10.1111/j.1525-1497.2005.0157.x.
396. Mudano AS, Casebeer L, Patino F, et al. Racial disparities in osteoporosis prevention in a managed care population. *Southern Medical Journal*. 2003 2003/05;96(5):445-51. doi: 10.1097/01.SMJ.0000053918.93363.B0.
397. Neuner JM, Zhang X, Sparapani R, et al. Racial and socioeconomic disparities in bone density testing before and after hip fracture. *J Gen Intern Med*. 2007 2007/09;22(9):1239-45. doi: 10.1007/s11606-007-0217-1.
398. Cunningham TD, Di Pace BS, Ullal J. Osteoporosis treatment disparities: a 6-year aggregate analysis from national survey data. *Osteoporos Int*. 2014 2014/09/01;25(9):2199-208. doi: 10.1007/s00198-014-2747-9.
399. Hamrick I, Whetstone LM, Cummings DM. Racial disparity in treatment of osteoporosis after diagnosis. *Osteoporos Int*. 2006 2006/11/01;17(11):1653-8. doi: 10.1007/s00198-006-0178-y.
400. Lin C-W, Karaca-Mandic P, McCullough JS, Weaver L. Access to oral osteoporosis drugs among female Medicare part D beneficiaries. *Womens Health Issues*. 2014 2014;24(4):e435-e45. doi: 10.1016/j.whi.2014.04.002.
401. Block AE, Solomon DH, Cadarette SM, et al. Patient and physician predictors of post-fracture osteoporosis management. *J Gen Intern Med*. 2008 2008/09;23(9):1447-51. doi: 10.1007/s11606-008-0697-7.
402. Boytsov NN, Crawford AG, Hazel-Fernandez LA, et al. Patient and provider characteristics associated with optimal post-fracture osteoporosis management. *Am J Med Qual*. 2017 Nov/Dec;32(6):644-54. doi: 10.1177/1062860617691123. PMID: 28693331.
403. Keshishian A, Boytsov N, Burge R, et al. Examining the treatment gap and risk of subsequent fractures among females with a fragility fracture in the US Medicare population. *Osteoporos Int*. 2017 Aug;28(8):2485-94. doi: 10.1007/s00198-017-4072-6. PMID: 28536737.
404. Mikuls TR, Saag KG, George V, et al. Racial disparities in the receipt of osteoporosis related healthcare among community-dwelling older women with arthritis and previous fracture. *J Rheumatol* 2005;32(5):870-5.
405. Solomon DH, Brookhart MA, Gandhi TK, et al. Adherence with osteoporosis practice guidelines: a multilevel analysis of patient, physician, and practice setting characteristics. *Am J Med*. 2004 Dec;117(12):919-24. doi: 10.1016/j.amjmed.2004.06.040.

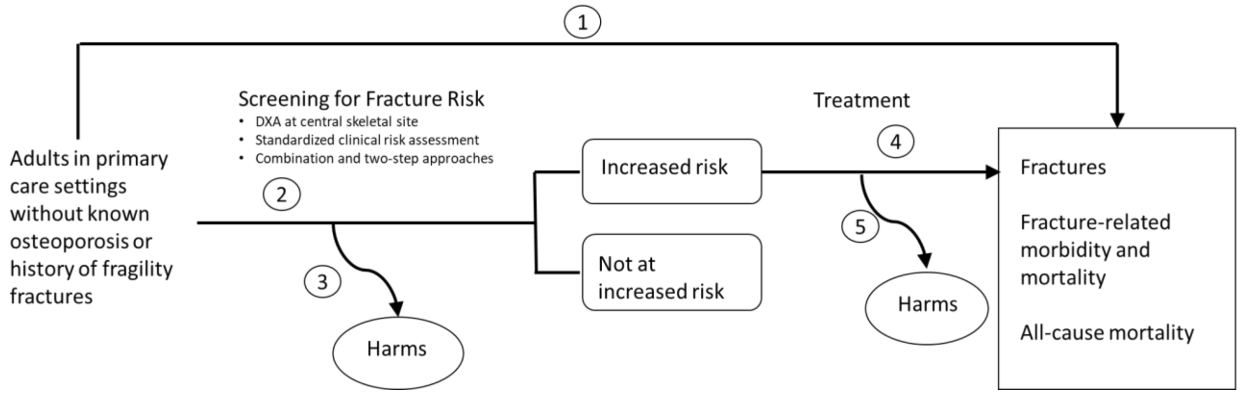
406. Curtis JR, Westfall AO, Allison JJ, et al. Longitudinal patterns in the prevention of osteoporosis in glucocorticoid-treated patients. *Arthritis & Rheumatism*. 2005 2005/08;52(8):2485-94. doi: 10.1002/art.21194.
407. Ruiz-Esteves KN, Teysir J, Schatoff D, et al. Disparities in osteoporosis care among postmenopausal women in the United States. *Maturitas*. 2022 Feb;156:25-9. doi: 10.1016/j.maturitas.2021.10.010. PMID: 35033230.
408. Burgess DJ, Fu SS, van Ryn M. Why do providers contribute to disparities and what can be done about it? *J Gen Intern Med*. 2004 Nov;19(11):1154-9. doi: 10.1111/j.1525-1497.2004.30227.x. PMID: 15566446.
409. Sale JEM, Bogoch E, Hawker G, et al. Patient perceptions of provider barriers to post-fracture secondary prevention. *Osteoporosis International*. 2014 2014/11/01;25(11):2581-9. doi: 10.1007/s00198-014-2804-4.
410. Neuman MD, Kennelly AM, Tosi LL. Breakout session: sex/gender and racial/ethnic disparities in the care of osteoporosis and fragility fractures. *Clinical Orthopaedics & Related Research*. 2011 2011/07;469(7):1936-40. doi: 10.1007/s11999-011-1859-1.
411. van Ryn M, Burke J. The effect of patient race and socio-economic status on physicians' perceptions of patients. *Social Science*. 2000 2000:16.
412. Mossey JM. Defining racial and ethnic disparities in pain management. *Clinical Orthopaedics & Related Research*. 2011 2011/07;469(7):1859-70. doi: 10.1007/s11999-011-1770-9.
413. Ashana DC, D'Arcangelo N, Gazarian PK, et al. "Don't talk to them about goals of care": understanding disparities in advance care planning. *Journals of Gerontology. Series A: Biological Sciences & Medical Sciences*. 2021 2021/03/29;77(2):339-46. doi: 10.1093/gerona/glab091.
414. Zallman L, Ma J, Xiao L, Lasser KE. Quality of US primary care delivered by resident and staff physicians. *Journal of General Internal Medicine*. 2010 2010/11;25(11):1193-7. doi: 10.1007/s11606-010-1456-0.
415. Wieland ML, Jaeger TM, Bundrick JB, et al. Resident physician perspectives on outpatient continuity of care. *Journal of Graduate Medical Education*. 2013 2013/12/01;5(4):668-73. doi: 10.4300/JGME-05-04-40.
416. Essien UR, He W, Ray A, et al. Disparities in quality of primary care by resident and staff physicians: is there a conflict between training and equity? *Journal of General Internal Medicine*. 2019 2019/07;34(7):1184-91. doi: 10.1007/s11606-019-04960-5.
417. Unson CG, Siccion E, Gaztambide J, et al. Nonadherence and osteoporosis treatment preferences of older women: a qualitative study. *J Womens Health*. 2003 2003/12;12(10):1037-45. doi: 10.1089/154099903322643965.
418. Bailey ZD, Krieger N, Agénor M, et al. Structural racism and health inequities in the USA: evidence and interventions. *The Lancet*. 2017 2017/04/08;389(10077):1453-63. doi: 10.1016/S0140-6736(17)30569-X.
419. Wright NC, Melton ME, Sohail M, et al. Race plays a role in the knowledge, attitudes and beliefs of women with osteoporosis. *J Racial Ethn Health Disparities*. 2019 2019/08;6(4):707-18. doi: 10.1007/s40615-019-00569-w.
420. U.S. Preventive Services Task Force. Policies and Procedure Manual, Appendix VI. Criteria for Assessing Internal Validity of Individual Studies. Rockville, MD: U.S. Preventive Services Task Force; 2017. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-appendix-vi-criteria-assessing-internal-validity-individual-studies. Accessed 18 January 2023.
421. Harris RP, Helfand M, Woolf SH, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3S):21-35. doi: https://doi.org/10.1016/S0749-3797(01)00261-6.
422. Chanplakorn P, Lertudomphonwanit T, Daraphongsataporn N, et al. Development of prediction model for osteoporotic vertebral compression fracture screening without using clinical risk factors, compared with FRAX and other previous models. *Arch Osteoporos*. 2021 Jun 3;16(1):84. doi: 10.1007/s11657-021-00957-y. PMID: 34085107.
423. Agarwal A, Baleanu F, Moreau M, et al. External validation of FRISBEE 5-year fracture prediction models: a registry-based cohort study. *Arch Osteoporos*. 2022 Dec 23;18(1):13. doi: 10.1007/s11657-022-01205-7. PMID: 36564674.
424. Takano Y, Furune S, Miyai Y, et al. Drug-induced thrombocytopenia associated with trastuzumab in a patient with HER2-positive recurrent gastric cancer. *International Cancer Conference Journal*. 2022;11(1):67-70. doi: 10.1007/s13691-021-00520-z.

425. Unnanuntana A, Ashfaq K, Ton QV, et al. The effect of long-term alendronate treatment on cortical thickness of the proximal femur. *Clinical Orthopaedics & Related Research*. 2012 Jan;470(1):291-8. doi: 10.1007/s11999-011-1985-9 [doi]. PMID: 21773861.
426. Crandall CJ. Risk assessment tools for osteoporosis screening in postmenopausal women: a systematic review. *Curr Osteoporos Rep*. 2015 Oct;13(5):287-301. doi: 10.1007/s11914-015-0282-z. PMID: 26233285.
427. Chao AS, Chen FP, Lin YC, et al. Application of the World Health Organization Fracture Risk Assessment Tool to predict need for dual-energy X-ray absorptiometry scanning in postmenopausal women. *Taiwan J Obstet Gynecol*. 2015 Dec;54(6):722-5. doi: 10.1016/j.tjog.2015.10.005. PMID: 26700992.
428. Kirilova E, Kirilov N, Bischoff F. Predictive ability of the osteoporosis self-assessment tool for assessing the risk of osteoporosis. *Revmatologiia (Bulgaria)*. 2019;27(3):3-9.
429. Fogelman I, Ribot C, Smith R, et al. Risedronate reverses bone loss in postmenopausal women with low bone mass: results from a multinational, double-blind, placebo-controlled trial. BMD-MN Study Group. *Journal of Clinical Endocrinology & Metabolism*. 2000 May;85(5):1895-900. doi: 10.1210/jcem.85.5.6603. PMID: 10843171.
430. Abrahamsen B, Pazianas M, Eiken P, et al. Esophageal and gastric cancer incidence and mortality in alendronate users. *Journal of Bone & Mineral Research*. 2012 Mar;27(3):679-86. doi: 10.1002/jbmr.1481. PMID: 22113985.
431. Black DM, Geiger EJ, Eastell R, et al. Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. *N Engl J Med*. 2020 Aug 20;383(8):743-53. doi: 10.1056/NEJMoa1916525. PMID: 32813950.
432. Cardwell CR, Abnet CC, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of esophageal cancer. *JAMA*. 2010 Aug 11;304(6):657-63. doi: 10.1001/jama.2010.1098. PMID: 20699457.
433. Chiang CH, Huang CC, Chan WL, et al. Oral alendronate use and risk of cancer in postmenopausal women with osteoporosis: a nationwide study. *Journal of Bone & Mineral Research*. 2012 Sep;27(9):1951-8. doi: 10.1002/jbmr.1645. PMID: 22532232.
434. Choi D, Choi S, Chang J, Park SM. Exposure to oral bisphosphonates and risk of gastrointestinal cancer. *Osteoporos Int*. 2020 Apr;31(4):775-82. doi: 10.1007/s00198-020-05327-x. PMID: 32034451.
435. Kim SH, Lee YK, Kim TY, et al. Incidence of and risk for osteonecrosis of the jaw in Korean osteoporosis patients treated with bisphosphonates: a nationwide cohort-study. *Bone*. 2021 Feb;143:115650. doi: 10.1016/j.bone.2020.115650. PMID: 32956854.
436. Lee WY, Sun LM, Lin MC, et al. A higher dosage of oral alendronate will increase the subsequent cancer risk of osteoporosis patients in Taiwan: a population-based cohort study. *PLoS One*. 2012;7(12):e53032. doi: 10.1371/journal.pone.0053032. PMID: 23300854.
437. Nordström P, Bergman J, Ballin M, et al. Bone-specific drugs and osteonecrosis of sites other than the jaw: a nationwide cohort study. *Journal of Bone & Mineral Research*. 2020 Sep;35(9):1703-10. doi: 10.1002/jbmr.4040. PMID: 32379370.
438. Passarelli MN, Newcomb PA, LaCroix AZ, et al. Oral bisphosphonate use and colorectal cancer incidence in the women's health initiative. *Journal of Bone & Mineral Research*. 2013 Sep;28(9):2043-8. doi: 10.1002/jbmr.1930. PMID: 23519920.
439. Rodríguez AJ, Ernst MT, Nybo M, et al. Oral bisphosphonate use reduces cardiovascular events in a cohort of Danish patients referred for bone mineral density. *Journal of Clinical Endocrinology & Metabolism*. 2020 Oct 1;105(10). doi: 10.1210/clinem/dgaa481. PMID: 32717068.
440. Thadani SR, Ristow B, Blackwell T, et al. Relationship of bisphosphonate therapy and atrial fibrillation/flutter: outcomes of Sleep Disorders in Older Men (MrOS Sleep) study. *Chest*. 2016 May;149(5):1173-80. doi: 10.1016/j.chest.2015.11.022. PMID: 26836889.
441. Vestergaard P. Occurrence of gastrointestinal cancer in users of bisphosphonates and other antiresorptive drugs against osteoporosis. *Calcif Tissue Int*. 2011 Dec;89(6):434-41. doi: 10.1007/s00223-011-9539-4. PMID: 22002678.
442. Wang JC, Chien WC, Chung CH, et al. Adverse cardiovascular effects of nitrogen-containing bisphosphonates in patients with osteoporosis: a nationwide population-based retrospective study. *Int J Cardiol*. 2016 Jul 15;215:232-7. doi: 10.1016/j.ijcard.2016.04.088. PMID: 27128537.
443. Yang HY, Huang JH, Chiu HW, et al. Vitamin D and bisphosphonates therapies for osteoporosis are associated with different risks of atrial fibrillation in women: a nationwide population-based analysis. *Medicine (Baltimore)*. 2018 Oct;97(43):e12947. doi: 10.1097/md.0000000000012947. PMID: 30412111.

444. Yuh DY, Chang TH, Huang RY, et al. The national-scale cohort study on bisphosphonate-related osteonecrosis of the jaw in Taiwan. *J Dent*. 2014 Oct;42(10):1343-52. doi: 10.1016/j.jdent.2014.05.001. PMID: 24907558.
445. Thomsen K, Ryg J, Matzen L, et al. Choice of osteoporosis guideline has important implications for the treatment decision in elderly women referred to a fall clinic. *Dan Med J*. 2014 Dec;61(12):A4980. PMID: 25441734.
446. Tarantino U, Rao C, Tempesta V, et al. Hip fractures in the elderly: the role of cortical bone. *Injury*. 2016 Oct;47 Suppl 4:S107-s11. doi: 10.1016/j.injury.2016.07.058. PMID: 27558243.
447. Baccaro LF, De Souza Santos Machado V, Costa-Paiva L, et al. Factors associated with osteoporosis in Brazilian women: a population-based household survey. *Arch Osteoporos*. 2013;8(1-2).
448. Torgerson DJ, Donaldson C, Russell IT, Reid DM. Hormone replacement therapy: compliance and cost after screening for osteoporosis. *European Journal of Obstetrics, Gynecology, & Reproductive Biology*. 1995 Mar;59(1):57-60. doi: 10.1016/0028-2243(94)02002-v. PMID: 7781863.
449. Koh LKH, Ben Sedrine W, Torralba TP, et al. A simple tool to identify Asian women at increased risk of osteoporosis. *Osteoporosis International*. 2001;12(8):699-705. doi: DOI 10.1007/s001980170070. PMID: WOS:000171206900011.
450. Bazzocchi A, Ponti F, Diano D, et al. Trabecular bone score in healthy ageing. *Br J Radiol*. 2015 Aug;88(1052):20140865. doi: 10.1259/bjr.20140865 [doi]. PMID: 26148778.
451. Frost SA, Nguyen ND, Center JR, et al. Timing of repeat BMD measurements: development of an absolute risk-based prognostic model. *Journal of Bone & Mineral Research*. 2009 Nov;24(11):1800-7. doi: 10.1359/jbmr.090514. PMID: 19419321.
452. Gourlay ML, Overman RA, Fine JP, et al. Baseline age and time to major fracture in younger postmenopausal women. *Menopause*. 2015 Jun;22(6):589-97. doi: 10.1097/GME.0000000000000356 [doi]. PMID: 25349960.
453. Looker AC, Sarafrazi Isfahani N, Fan B, Shepherd JA. FRAX-based estimates of 10-year probability of hip and major osteoporotic fracture among adults aged 40 and over: United States, 2013 and 2014. *Natl Health Stat Report*. 2017 Mar(103):1-16. PMID: 28459415.
454. Chotiyarnwong P, McCloskey EV, Harvey NC, et al. Is it time to consider population screening for fracture risk in postmenopausal women? A position paper from the International Osteoporosis Foundation Epidemiology/Quality of Life Working Group. *Arch Osteoporos*. 2022 Jun 28;17(1):87. doi: 10.1007/s11657-022-01117-6. PMID: 35763133.
455. US Census Bureau. American Community Survey 2016–2020 5-Year Data Release. Washington, DC: U.S. Census Bureau; 2022.
456. Bolland MJ, Grey A. Disparate outcomes from applying U.K. and U.S. osteoporosis treatment guidelines. *Journal of Clinical Endocrinology & Metabolism*. 2010 Apr;95(4):1856-60. doi: 10.1210/jc.2009-2157. PMID: 20147581.
457. McCloskey E, Kanis JA, Johansson H, et al. FRAX-based assessment and intervention thresholds--an exploration of thresholds in women aged 50 years and older in the UK. *Osteoporos Int*. 2015 Aug;26(8):2091-9. doi: 10.1007/s00198-015-3176-0. PMID: 26077380.
458. Kanis JA, Johansson H, Harvey NC, et al. An assessment of intervention thresholds for very high fracture risk applied to the NOGG guidelines: a report for the National Osteoporosis Guideline Group (NOGG). *Osteoporos Int*. 2021 Apr 4. doi: 10.1007/s00198-021-05942-2. PMID: 33813622.
459. Shane E, Burr D, Abrahamsen B, et al. Atypical subtrochanteric and diaphyseal femoral fractures: second report of a task force of the American Society for Bone and Mineral Research. *Journal of Bone & Mineral Research*. 2014 Jan;29(1):1-23. doi: 10.1002/jbmr.1998. PMID: 23712442.
460. Adler RA, El-Hajj Fuleihan G, Bauer DC, et al. Managing osteoporosis in patients on long-term bisphosphonate treatment: report of a task force of the American Society for Bone and Mineral Research. *Journal of Bone & Mineral Research*. 2016 Jan;31(1):16-35. doi: 10.1002/jbmr.2708. PMID: 26350171.
461. Lu L, Lu L, Zhang J, Li J. Potential risks of rare serious adverse effects related to long-term use of bisphosphonates: an overview of systematic reviews. *Journal of Clinical Pharmacy & Therapeutics*. 2020 Feb;45(1):45-51. doi: 10.1111/jcpt.13056. PMID: 31608489.
462. Dennison EM, Cooper C, Kanis JA, et al. Fracture risk following intermission of osteoporosis therapy. *Osteoporos Int*. 2019 Sep;30(9):1733-43. doi: 10.1007/s00198-019-05002-w. PMID: 31175404.
463. Fink HA, MacDonald R, Forte ML, et al. Long-term drug therapy and drug discontinuations and holidays for osteoporosis fracture prevention: a systematic review. *Ann Intern Med*. 2019 Jul 2;171(1):37-50. doi: 10.7326/m19-0533. PMID: 31009947.

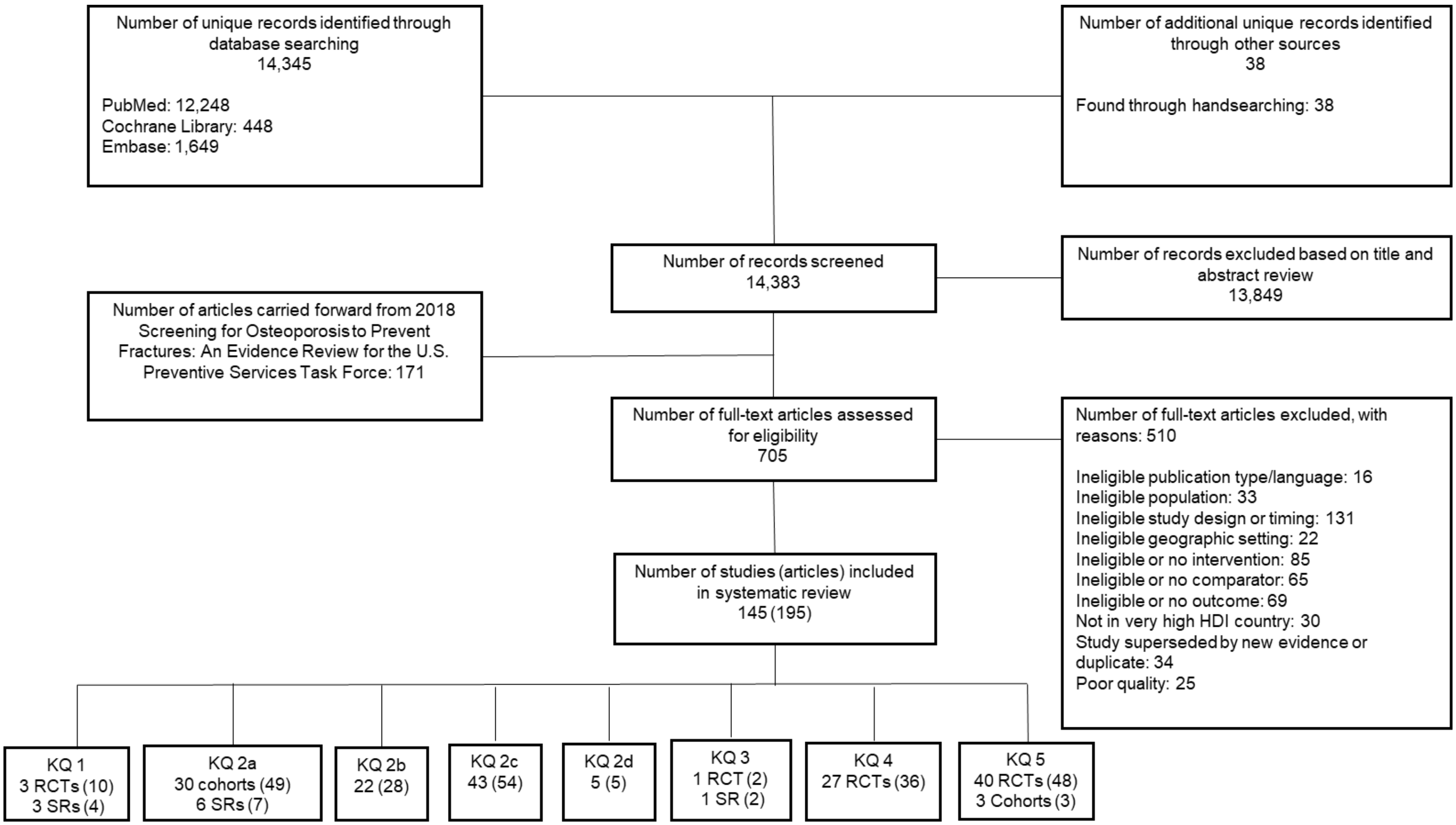
464. Won SH, Ahn S, Ji E, et al. Hip and subtrochanteric or femoral shaft fractures after bisphosphonate use in Korean women, using Korean national sample cohort. *J Korean Med Sci*. 2020 Jun 29;35(25):e193. doi: 10.3346/jkms.2020.35.e193. PMID: 32597040.
465. Lo JC, Grimsrud CD, Ott SM, et al. Atypical femur fracture incidence in women increases with duration of bisphosphonate exposure. *Osteoporos Int*. 2019 Dec;30(12):2515-20. doi: 10.1007/s00198-019-05112-5. PMID: 31555883.
466. Lo JC, Neugebauer RS, Ettinger B, et al. Risk of complete atypical femur fracture with oral bisphosphonate exposure beyond three years. *BMC Musculoskelet Disord*. 2020 Dec 3;21(1):801. doi: 10.1186/s12891-020-03672-w. PMID: 33272248.
467. Bone HG, Wagman RB, Brandi ML, et al. 10 years of denosumab treatment in postmenopausal women with osteoporosis: results from the phase 3 randomised FREEDOM trial and open-label extension. *Lancet Diabetes Endocrinol*. 2017 Jul;5(7):513-23. doi: 10.1016/s2213-8587(17)30138-9. PMID: 28546097.
468. Ruggiero SL, Dodson TB, Fantasia J, et al. American Association of Oral and Maxillofacial Surgeons position paper on medication-related osteonecrosis of the jaw--2014 update. *Journal of Oral & Maxillofacial Surgery*. 2014 Oct;72(10):1938-56. doi: 10.1016/j.joms.2014.04.031. PMID: 25234529.
469. Anastasilakis AD, Pepe J, Napoli N, et al. Osteonecrosis of the jaw and antiresorptive agents in benign and malignant diseases: a critical review organized by the ECTS. *Journal of Clinical Endocrinology & Metabolism*. 2022 Apr 19;107(5):1441-60. doi: 10.1210/clinem/dgab888. PMID: 34922381.
470. Everts-Graber J, Reichenbach S, Gahl B, et al. Risk factors for vertebral fractures and bone loss after denosumab discontinuation: a real-world observational study. *Bone*. 2021 Mar;144:115830. doi: 10.1016/j.bone.2020.115830. PMID: 33359006.
471. Tanaka S, Mizutani H, Tsuruya E, et al. Long-term safety and effectiveness of denosumab in Japanese patients with osteoporosis: 3-year post-marketing surveillance study. *Journal of Bone & Mineral Metabolism*. 2021 May;39(3):463-73. doi: 10.1007/s00774-020-01180-4. PMID: 33387064.
472. Tsourdi E, Zillikens MC, Meier C, et al. Fracture risk and management of discontinuation of denosumab therapy: a systematic review and position statement by ECTS. *Journal of Clinical Endocrinology & Metabolism*. 2020 Oct 26. doi: 10.1210/clinem/dgaa756. PMID: 33103722.
473. Brown JP, Roux C, Torring O, et al. Discontinuation of denosumab and associated fracture incidence: analysis from the Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months (FREEDOM) trial. *Journal of Bone & Mineral Research*. 2013 Apr;28(4):746-52. doi: 10.1002/jbmr.1808 [doi]. PMID: 23109251.
474. Cummings SR, Ferrari S, Eastell R, et al. Vertebral fractures after discontinuation of denosumab: a post hoc analysis of the randomized placebo-controlled FREEDOM trial and its extension. *Journal of Bone & Mineral Research*. 2018 Feb;33(2):190-8. doi: 10.1002/jbmr.3337. PMID: 29105841.
475. McClung MR, Wagman RB, Miller PD, et al. Observations following discontinuation of long-term denosumab therapy. *Osteoporos Int*. 2017 May;28(5):1723-32. doi: 10.1007/s00198-017-3919-1. PMID: 28144701.
476. Anastasilakis AD, Polyzos SA, Makras P, et al. Clinical features of 24 patients with rebound-associated vertebral fractures after denosumab discontinuation: systematic review and additional cases. *Journal of Bone & Mineral Research*. 2017 Jun;32(6):1291-6. doi: 10.1002/jbmr.3110. PMID: 28240371.
477. Burckhardt P, Faouzi M, Buclin T, Lamy O. Fractures after denosumab discontinuation: a retrospective study of 797 cases. *Journal of Bone & Mineral Research*. 2021 May 19;36:1717-28. doi: 10.1002/jbmr.4335. PMID: 34009703.
478. Lyu H, Yoshida K, Zhao SS, et al. Delayed denosumab injections and fracture risk among patients with osteoporosis: a population-based cohort study. *Ann Intern Med*. 2020 Oct 6;173(7):516-26. doi: 10.7326/m20-0882. PMID: 32716706.
479. Tripto-Shkolnik L, Fund N, Rouach V, et al. Fracture incidence after denosumab discontinuation: real-world data from a large healthcare provider. *Bone.* 2020 Jan;130:115150. doi: 10.1016/j.bone.2019.115150. PMID: 31715340.

Figure 1. Analytic Framework



Abbreviations: DXA=dual-energy X-ray absorptiometry.

Figure 2. Literature Flow Diagram



Some studies (articles) are included in more than one KQ.

Abbreviations: HDI=human development index; KQ=key question’ RCT=randomized, controlled trial; SR=systematic review; U.S.=United States.

Figure 3. Randomized, Controlled Trials of Screening vs. Usual Care—Fracture and Mortality Outcomes (KQ 1)

Fracture and Study	Age	PriorFx	Mean FRAX MOF/Hip Risk	Length of F/U	Screened n with Events/Total N(%)	Control n with Events/Total N(%)	Risk Ratio (95% CI)	
All								
SOS	75	43%	24.5/11.5	3.7 y	626/5,516 (11.3%)	632/5,405 (11.7%)	0.97 (0.87, 1.08)	
SCOOP	76	22%	19.3/85	5 y	951/6,233 (15.3%)	1,002/6,250 (16.0%)	0.95 (0.88, 1.03)	
Subgroup, DL (I^2 = 0.0%, p = 0.771)							0.96 (0.90, 1.02)	Absolute effect: 6 fewer per 1,000 (from 14 fewer to 3 more)
Osteoporotic								
SOS	75	43%	24.5/11.5	3.7 y	547/5,516 (9.9%)	578/5,405 (10.7%)	0.93 (0.83, 1.04)	
ROSE	70	10.2%	20/6.7	5 y	996/9,279 (10.7%)	1,025/9,326 (11.0%)	0.98 (0.90, 1.06)	
SCOOP*	76	22%	19.3/85	5 y	805/6233 (12.9%)	852/6250 (13.6%)	0.95 (0.87, 1.04)	
Subgroup, DL (I^2 = 0.0%, p = 0.746)							0.95 (0.91, 1.01)	Absolute effect: 6 fewer per 1,000 (from 11 fewer to 1 more)
MOF								
SCOOP*	76	22%	19.3/85	5 y	805/6233 (12.9%)	852/6250 (13.6%)	0.95 (0.87, 1.04)	
ROSE	70	10.2%	20/6.7	5 y	725/9,279 (7.8%)	786/9,326 (8.4%)	0.93 (0.84, 1.02)	
SOS	75	43%	24.5/11.5	3.7 y	427/5,516 (7.7%)	452/5,405 (8.3%)	0.93 (0.82, 1.05)	
Subgroup, DL (I^2 = 0.0%, p = 0.934)							0.94 (0.88, 0.99)	Absolute effect: 6 fewer per 1,000 (from 12 fewer to 1 fewer)
Hip								
SCOOP	76	22%	19.3/85	5 y	164/6,233 (2.6%)	218/6,250 (3.5%)	0.75 (0.62, 0.92)	
SOS	75	43%	24.5/11.5	3.7 y	133/5,516 (2.4%)	143/5,405 (2.6%)	0.91 (0.72, 1.15)	
ROSE	70	10.2%	20/6.7	5 y	169/9,279 (1.8%)	202/9,326 (2.2%)	0.84 (0.69, 1.03)	
Subgroup, DL (I^2 = 0.0%, p = 0.470)							0.83 (0.73, 0.93)	Absolute effect: 5 fewer per 1,000 (from 7 fewer to 2 fewer)
Mortality								
SCOOP	76	22%	19.3/85	5 y	550/6,233 (8.8%)	525/6,250 (8.4%)	1.05 (0.94, 1.18)	
ROSE	71	12%	20/6.7	5 y	1,968/17,072 (11.5%)	2,038/17,157 (11.9%)	0.97 (0.92, 1.03)	
SOS	75	43%	24.5/11.5	3.7 y	499/5,516 (9.0%)	479/5,405 (8.9%)	1.02 (0.91, 1.15)	
Subgroup, DL (I^2 = 0.0%, p = 0.421)							0.99 (0.95, 1.04)	Absolute effect: 1 fewer per 1,000 (from 5 fewer to 4 more)

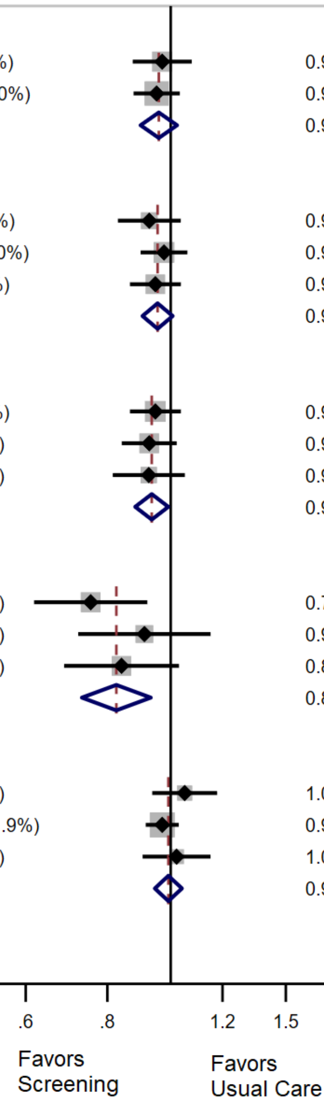


Figure 3. Randomized, Controlled Trials of Screening vs. Usual Care—Fracture and Mortality Outcomes (KQ 1)

Note: This analysis used the first per-protocol data from the ROSE trial for the fracture outcomes because these data reflect a similar study design as the intention to treat (ITT) data reported in SCOOP and SOS. **Appendix E.1 Figure 1** provides a sensitivity analysis using the ITT data from the ROSE trial for the fracture outcomes. The data for mortality is the ITT population for ROSE because per-protocol data for ROSE was not reported.

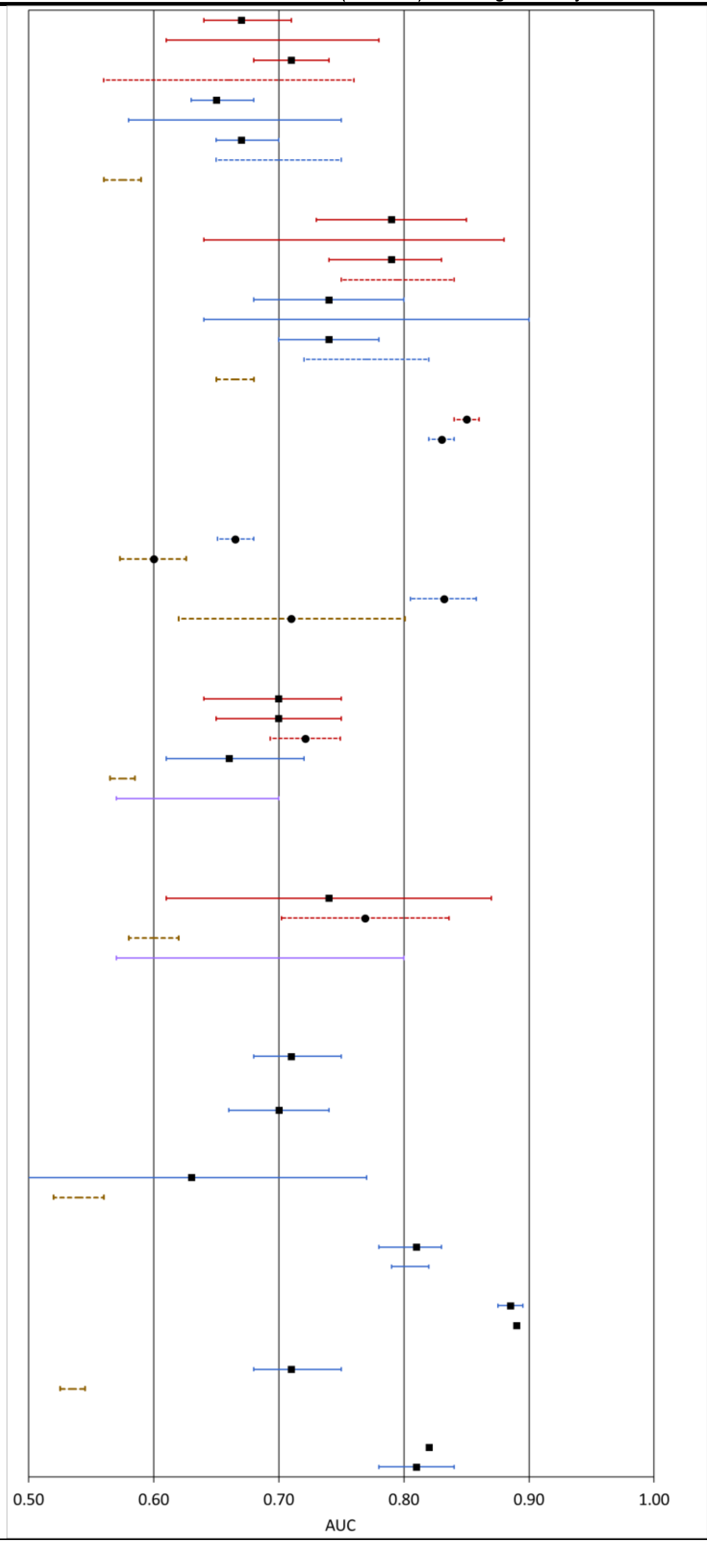
[*] SCOOP reported an outcome entitled "Osteoporotic Fractures," which was defined as clinical fractures excluding hand, foot, skull, or cervical vertebrae. This definition differs from the definition of MOF used by the other two studies (hip, clinical vertebral, distal forearm, and humerus); as such, we have included SCOOP "Osteoporosis" outcome in the estimate for both "Osteoporotic Fractures" and for "MOF" in this figure. The RR estimate for MOF without SCOOP included is 0.93 (95% CI, 0.86 to 1.00); Absolute Effect: 6 fewer (from 12 fewer to 0 fewer). It is also not clear that fractures associated with trauma were excluded from SCOOP.

Abbreviations: ARD=absolute risk difference; AUC=area under the curve; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FRAX=Fracture Risk Assessment Tool; F/U=followup; KQ=key question; MOF=major osteoporotic fracture; N/n=number; NR=not reported; PriorFx=prior fracture; ROSE=Risk-stratified Osteoporosis Strategy Evaluation; SCOOP=Screening in the Community to Reduce Fractures in Older Women; SOS=Stichting Artsen Laboratorium en Trombosedienst (SALT) Osteoporosis Study; vs.=versus; y=year.

Figure 4a. Accuracy of Risk Instruments for Predicting Major Osteoporotic Fractures and Hip Fractures in Women (KQ 2a)

No. Studies in SR (No. Participants)	No. Primary Studies (No. Participants)
FRAX MOF with BMD	
5 (14,244)[134]; 18 (NR)[137]; 23 (NR)[138]	6 (155,093)[142, 172, 175, 176, 179, 312]
FRAX MOF without BMD	
7 (24,726)[134]; 13 (NR)[137]; 19 (NR)[138]	3 (243,359)[142, 172, 176] 1 (63,621)[139, 140, 143] ages 50-64
FRAX Hip with BMD	
5 (115, 611)[134]; 12 (NR)[137]; 12(NR)[138]	5 (152,812)[142, 172, 176, 179, 312]
FRAX Hip without BMD	
9 (131,244)[134]; 10 (NR)[137]; 17 (NR)[138]	3 (242,676)[142, 172, 176] 1 (63,723)[140] ages 50-64
FRC Hip with BMD	
None	1 (94,489)[73, 165]
FRC Hip without BMD	
None	1 (94,489)[73, 165]
FREM MOF	
None	1 (34,149)[163] age≥65 1 (33,781)[163] age<65
FREM Hip	
None	1 (34,149)[163] age≥65 1 (33,781)[163] age<65
Garvan MOF with BMD	
3 (6,932)[134]; 1 (NR)[135]	1 (3,030)[179]
Garvan MOF without BMD	
1 (NR)[135]	1 (63,621)[140] ages 50-64
Garvan MOF (BMD not specified)	
7 (NR)[137]	None
Garvan Hip with BMD	
2 (5,574)[134]	1 (3,030)[179]
Garvan Hip without BMD	
None	1 (62,723)[140] ages 50-64
Garvan Hip (BMD not specified)	
3 (NR)[137]	None
ORAI	
1 (NR)[135]	None
OSIRIS	
1 (NR)[135]	None
OST MOF	
2 (NR)[135]	2 (58,915)[139, 143, 159] ages 50-64
QFracture MOF	
3 (1,778,570)[134]; 3 (NR)[137]	None
QFracture Hip	
3(1,779,154)[134]; 3 (NR)[137]	None
SCORE	
1 (NR)[135]	1 (62,492)[139] ages 50-64
WHI Hip	
1 (NR)[137]; 2 (NR)[135]	None

Area Under the Curve
Range for >1 Study or
Point Estimate (95% CI) for Single Study



■—SR; ●---- Primary studies; with BMD, no BMD, age <65 no BMD; BMD NR, Range represented when no symbol, otherwise symbol represents pooled (SR) or individual (primary study) estimate and 95% CI.

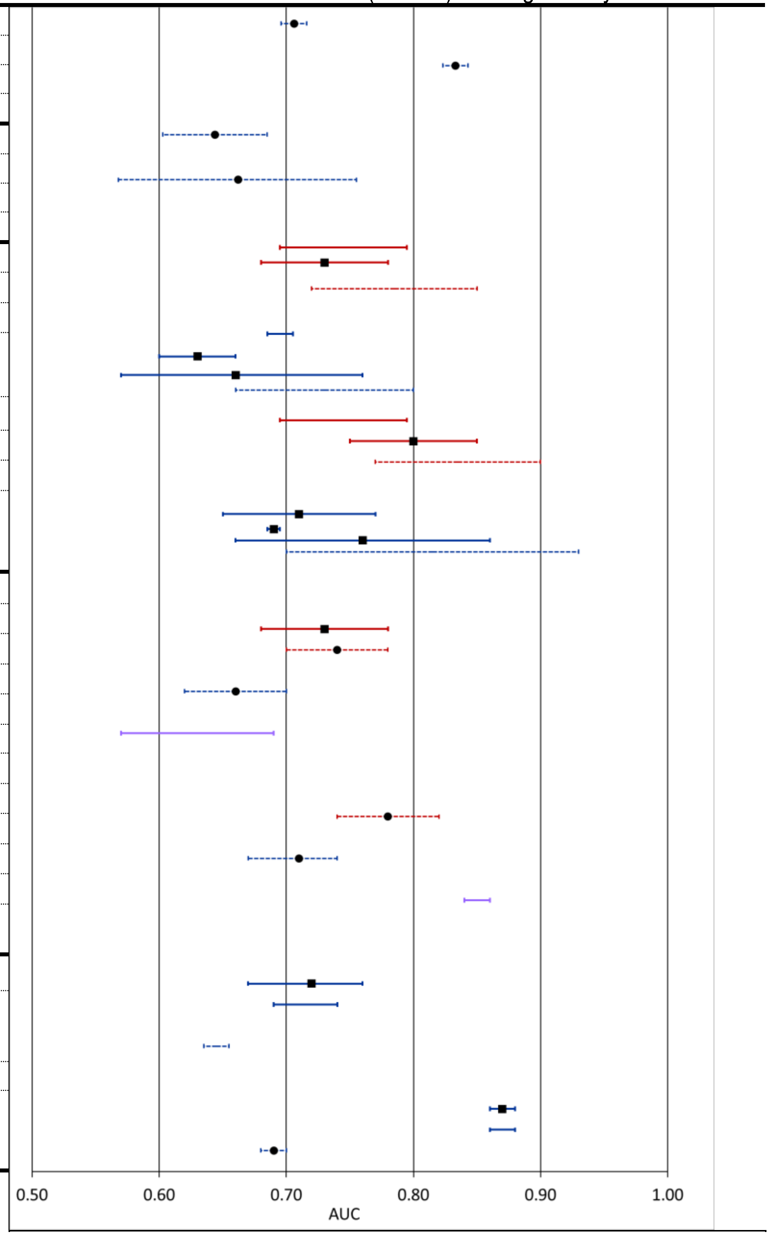
Figure 4a. Accuracy of Risk Instruments for Predicting Major Osteoporotic Fractures and Hip Fractures in Women (KQ 2a)

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; FREM=Fracture Risk Evaluation Model; KQ=key question; MOF=major osteoporotic fracture; NR=not reported; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=OSteoporosis Index of RISk; OST=Osteoporosis Self-Assessment Tool; SCORE=Simple Calculated Osteoporosis Risk Evaluation; SR=systematic reviews; WHI=Women's Health Initiative.

Figure 4b. Accuracy of Risk Instruments for Predicting Major Osteoporotic Fractures and Hip Fractures in Men (KQ 2a)

No. Studies in SR (No. Participants)	No. Primary Studies (No. Participants)
EPIC MOF	
None	1 (1,823,217)[180]
EPIC Hip	
None	1 (1,823,217)[180]
FREM MOF	
None	1 (6,898)[163]
FREM Hip	
None	1 (6,898)[163]
FRAX MOF with BMD	
8 (NR)[137]; 5 (NR)[138]	2 (5,883)[146, 176]
FRAX MOF without BMD	
2 (11,199)[134]	2 (5,883)[146, 176]
2 (NR)[137]; 5 (NR)[138]	
FRAX Hip with BMD	
4 (NR)[137]; 3 (NR)[138]	2(5,883)[146, 176]
FRAX Hip without BMD	
2 (11,199)[134]	2 (5,883)[146, 176]
1 (NR)[137]; 6 (NR)[138]	
Garvan MOF with BMD	
2 (5,010)[134]	1 (5,200)[146]
Garvan MOF without BMD	
None	1 (5,200)[146]
Garvan MOF BMD NR	
3 (NR)[137]	None
Garvan Hip with BMD	
None	1 (5,200)[146]
Garvan Hip without BMD	
None	1 (5,200)[146]
Garvan Hip BMD NR	
1 (NR)[137]	None
QFracture MOF	
3 (NR)[137]	1 (5,200)[146]
2 (1,741,983)[134]	
QFracture Hip	
2 (1,741,983)[134]	1 (5,200)[146]
3 (NR)[137]	

Area Under the Curve
Range for >1 Study or
Point Estimate (95% CI) for Single Study

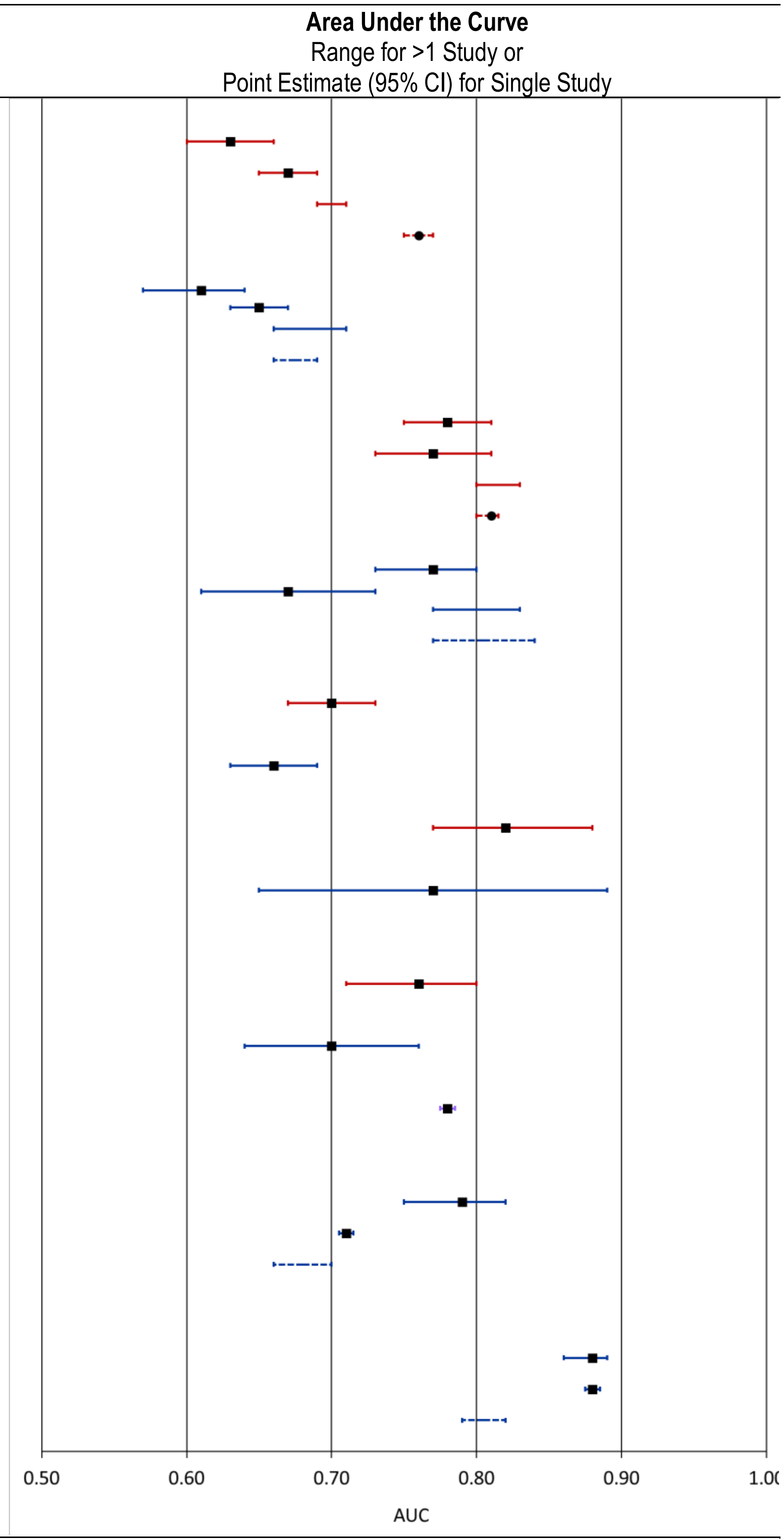


■—SR; ●------- Primary studies; with BMD, no BMD, BMD NR; Range represented when no symbol, otherwise symbol represents pooled (SR) or individual (primary study) estimate and 95% CI.

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; EPIC=Escala de Prediccion de fracturas Implementable en historia Clínica electronica - fracture prediction scale implementable in electronic medical record; FRAX=Fracture Risk Assessment Instrument; FREM=Fracture Risk Evaluation Model; KQ=key question; MOF=major osteoporotic fracture; NR=not reported; SR=systematic reviews.

Figure 4c. Accuracy of Risk Instruments for Predicting Major Osteoporotic Fractures and Hip Fractures in Mixed-Sex Populations (KQ 2a)

Instrument	No. Studies in SR (No. Participants)	No. Primary Studies (No. Participants)
FRAX MOF with BMD	3 (276,786)[134] 20 (NR)[135] 3 (NR)[137]	1 (115,206)[178]
FRAX MOF without BMD	3 (276,786)[134] 22 (NR)[135] 4 (NR)[137]	2 (33,711)[177, 311]
FRAX Hip with BMD	3 (276,786)[134] 17 (NR)[135] 3 (NR)[137]	1 (115,206)[178]
FRAX Hip without BMD	3 (276,786)[134] 23 (NR)[135] 3 (NR)[137]	3 (34,522)[173, 181, 311]
FRC MOF with BMD	1 (NR)[135]	None
FRC MOF without BMD	1 (NR)[135]	None
FRC Hip with BMD	2 (NR)[135]	None
FRC Hip without BMD	2 (NR)[135]	None
Garvan Hip with BMD	5 (NR)[135]	None
Garvan Hip without BMD	2 (NR)[135]	None
Garvan Hip BMD NR	1 (NR)[137]	None
QFracture MOF	2 (NR)[135] 1 (NR)[137]	2 (33,711)[177, 311]
QFracture Hip	2 (NR)[135] 1 (NR)[137]	2(25,440)[173, 311]



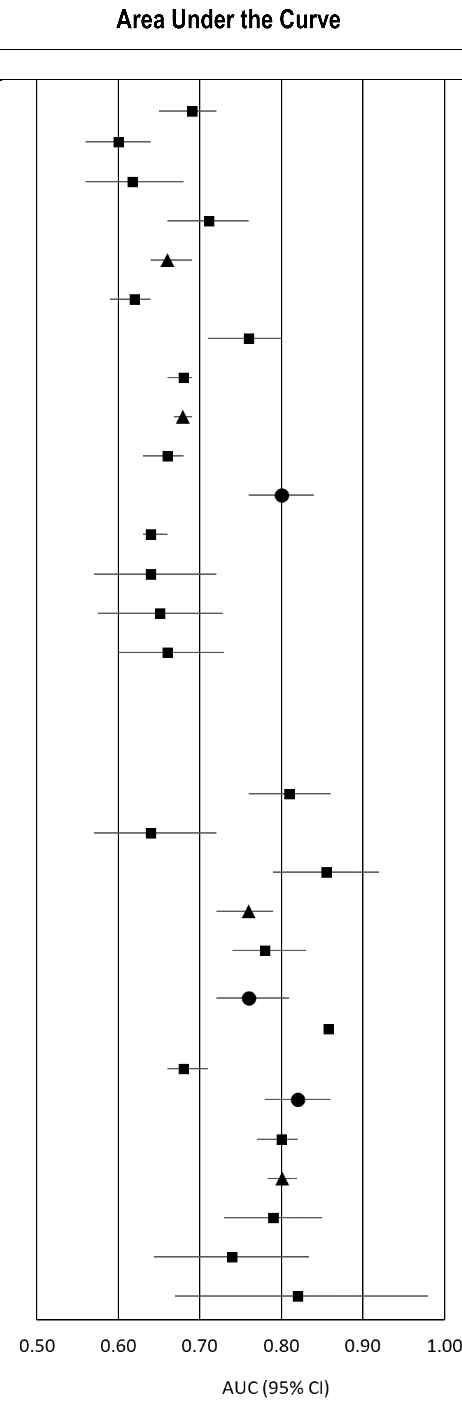
■—SR ●------- Primary studies; with BMD, no BMD, BMD NR; Range represented when no symbol, otherwise symbol represents pooled (SR) or individual (primary study) estimate and 95% CI.

Figure 4c. Accuracy of Risk Instruments for Predicting Major Osteoporotic Fractures and Hip Fractures in Mixed-Sex Populations (KQ 2a)

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; FRAX=Fracture Risk Assessment Instrument; FRC=Fracture Risk Calculator; KQ=key question; MOF=major osteoporotic fracture; NR=not reported; SR=systematic reviews.

Figure 5. Accuracy of Bone Mineral Density for Predicting Major Osteoporotic Fractures and Hip Fractures (KQ 2b)

Study	Sex	Total N	N With Fracture	Estimate (95% CI)
MOF				
Baleanu, 2021[179]	Women	3,030	281	0.69 (0.65 to 0.71)
Bolland, 2011[166]	Women	1,422	279	0.60 (0.56 to 0.62)
Chapurlat, 2020[175]	Women	2,100	61	0.62 (0.56 to 0.68)
Cheung, 2012[150]	Women	2,266	106	0.71 (0.66 to 0.76)
Fraser, 2011[154]	Mixed	6,697	695	0.66 (0.64 to 0.69)
Goldshtein, 2018[172]	Women	16,578	2,263	0.62 (0.59 to 0.64)
Gourlay, 2017[146]	Women	4,994	326	0.76 (0.71 to 0.80)
MBR-Hans, 2011[183]	Women	29,407	1,668	0.68 (0.66 to 0.69)
MBR-Leslie, 2010[157]	Mixed	39,603	2,543	0.68 (0.67 to 0.69)
Marques, 2017[176]	Women	1,943	145	0.66 (0.63 to 0.68)
	Men	683	33	0.80 (0.76 to 0.84)
Stewart, 2006[190]	Women	3,883	128	0.64 (0.63 to 0.66)
Tamaki, 2011[151]	Women	815	43	0.64 (0.57 to 0.72)
Tanaka, 2010[156]	Women	400	60	0.65 (0.58 to 0.73)
Tremollieres, 2010[189]	Women	2,196	145	0.66 (0.60 to 0.73)
Hip Fracture				
Baleanu, 2021[179]	Women	3,030	47	0.81 (0.76 to 0.86)
Bolland, 2011[166]	Women	1,422	57	0.64 (0.57 to 0.72)
Cheung, 2012[150]	Women	2,266	21	0.86 (0.79 to 0.92)
Fraser, 2011[154]	Mixed	6,697	175	0.76 (0.72 to 0.79)
Goldshtein, 2018[172]	Women	16,578	481	0.78 (0.74 to 0.83)
Gourlay, 2017[146]	Men	4,994	175	0.76 (0.72 to 0.81)
Iki, 2021[194]	Women	1,331	68	0.86 (NR)
Marques, 2017[176]	Women	1,943	20	0.68 (0.66 to 0.71)
	Men	683	8	0.82 (0.78 to 0.86)
MBR-Hans, 2011[183]	Women	29,407	293	0.80 (0.77 to 0.82)
MBR-Leslie, 2010[157]	Mixed	39,603	549	0.80 (0.78 to 0.82)
Robbins, 2007[187]	Women	10,750	80	0.79 (0.73 to 0.85)
Sund, 2014[185]	Women	2,755	21	0.74 (0.64 to 0.83)
Tamaki, 2011[151]	Women	815	4	0.82 (0.67 to 0.98)



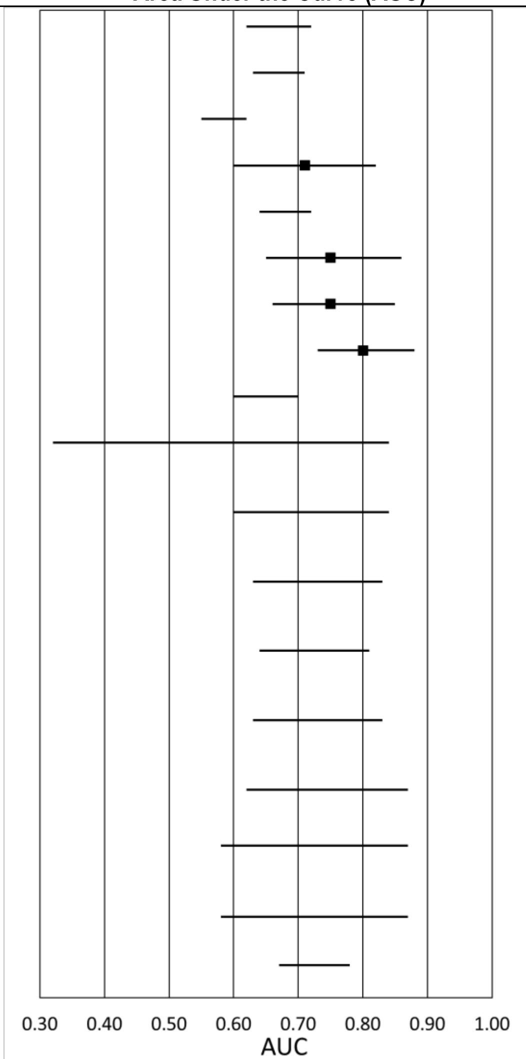
■=women; ●= men; ▲=mixed population of men and women

Abbreviations: AUC=area under the curve; CI=confidence interval; KQ=key question; MBR=Manitoba BMD Registry; MOF=major osteoporotic fracture; N/n=number; NR=not reported.

Figure 6a. Accuracy of Risk Assessment Instruments for Identifying Osteoporosis in Women (KQ 2c)

Instrument (Cutoff for Sn and Sp)	No. of Cohorts (Total N)	AUC Range or estimate (■ 95% CI) for single study	Sensitivity	Specificity
ABONE (varied*)	4[206, 207, 233, 242] (4,203)	0.62 to 0.72†	66% to 100%	10% to 60%
AMMEB (≥10)	2[199, 225] (1,520)	0.63 to 0.71	NR	NR
FRAX MOF (USPSTF Threshold‡)	3[141, 195, 220, 234, 236] (9,333); *Ages 50-64*	0.55 to 0.62	5% to 49%	63% to 96%
FRAX MOF ≥20%	1[233](367); Age ≥60	0.71 (0.60 to 0.82)	17%	96%
FRAX MOF (none)	2 [143, 200](22,922); *Ages 50-64*	0.64 to 0.72	NA	NA
FRAX Hip(>3%)	1[233] (367); *Age ≥60*	0.75 (0.65 to 0.86)	83%	54%
Garvan OF(≥20%)	1[233] (367); *Age ≥60*	0.75 (0.66 to 0.85)	55%	73%
Garvan Hip (>3%)	1[233] (367); *Age ≥60*	0.80 (0.73 to 0.88)	28%	95%
NOF (≥1 risk)	4[199, 206, 225, 226] (4,087)	0.60 to 0.70	96% and 100% (k=2)	10% and 18% (k=2)
ORAI (varied§)	19[199, 203, 206-210, 214, 217, 223-226, 228, 230, 233, 234, 236, 242] (24,277)	0.32 to 0.84†	43% to 100% (k=15)	0% to 100% (k=15)
ORAI (varied§)	5[209, 226, 228, 234, 236] (6,981) *Age <65*	0.60 to 0.84	44% to 99%	46% to 77%
OSIRIS (varied‖)	7[203, 208, 210, 214, 217, 233, 242] (6,987)	0.63 to 0.83	58% to 100% (k=6)	6% to 69% (k=6)
OST (varied¶)	14[199, 203, 205, 208-210, 214, 217, 223, 225, 228, 230, 231, 242, 243](35,812)	0.64 to 0.81 (k=10)#	29% to 95% (k=10)	37% to 92% (k=10)
OST (varied**)	6[143, 159, 195, 200, 209, 228, 234, 236] (29,701); *Age <65*	0.63 to 0.83	47% to 89% (k=5)	45% to 81% (k=5)
OSTA (varied††)	7[204, 207, 213, 215, 227, 233, 245] (3,967)	0.62 to 0.87 (k=5)†	41% to 100%	24% to 67%
SCORE (varied‡‡)	16[195, 203, 206-211, 217, 222, 224, 226, 228, 230, 233, 234, 236, 242, 245] (24,311)	0.58 to 0.87†	50% to 100% (k=12)†	15% to 93% (k=12)†
SCORE (varied§§)	6[195, 209, 226, 228, 234, 236](9,838); *Age <65*	0.58 to 0.87	62% to 100% (k=5)	34% to 71%) (k=5)
SOFSURF (varied‖‖)	3[208, 227, 231](1,720)	0.72 (0.67 to 0.78) (k=1)	72% to 92%	36% to 67%

Area Under the Curve (AUC)



Note: This plot depicts the range of AUC estimates (line with no symbol) from across 2 or more studies OR a single study point estimate and 95% CI when only one study reported an AUC estimate. The number of studies (k) in AUC, sensitivity, and specificity columns is provided when the estimate reported was from a fewer number of studies than

Figure 6a. Accuracy of Risk Assessment Instruments for Identifying Osteoporosis in Women (KQ 2c)

what is reported in the second column. Not all studies reported all three outcomes. Unless otherwise indicated, populations generally included postmenopausal women (>45, >50, >55, >60 years), but in some cases women as young as 40 years without regard to menopausal status were enrolled.

* Thresholds evaluated included ≥1.5, ≥2, ≥3.

† Does not include one study that was an extreme outlier.[245]

ǂ MOF risk ≥8.4 percent (2018 USPSTF recommendation) or 9.3 percent (2011 USPSTF recommendation).

§ The most common threshold was ≥9, but threshold varied from 8 to 20; the sensitivities and specificities reflect estimates from across all score thresholds.

‖ Score thresholds for sensitivity and specificity varied from <-3 to <1.5.

¶ The most common threshold was <2, but thresholds evaluated included ≤1, <0, ≤-1, <-2.9, and some studies did not report threshold because they only reported AUC.

Excluding two outliers (AUC 0.32[199] and 0.22).[225]

** The most common threshold was <2, but also included studies that did not report threshold (AUC only) and threshold of ≤1.

†† The most common threshold was ≤-1, but also included <0, <-1, ≤-2.

ǂǂ The most common threshold evaluated was ≥6, but other thresholds included >7, ≥7, ≥8, ≥11, ≥12, ≥20.75, and some studies where no threshold (AUC only) was reported were also evaluated.

§§ The most common threshold evaluated was ≥6, but >6, >7, and ≥7 were also evaluated.

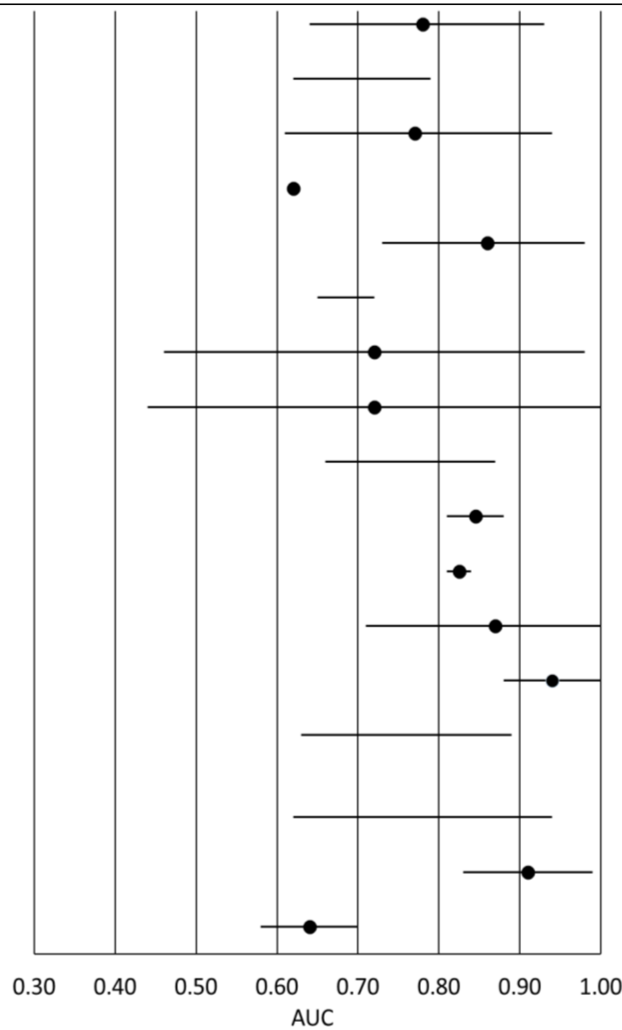
‖‖ Thresholds evaluated included ≥1, >1.7, ≥0.

Abbreviations: ABONE=Age, Bone, No Estrogen; AMMEB=Age, years after Menopause, age at MEnarche; AUC=area under the curve; CI=confidence interval; FRAX=Fracture Risk Assessment Tool; KQ=key question; MOF=major osteoporotic fracture; OF=osteoporotic fractures; N/n=number; NR=not reported; NOF=National Osteoporosis Foundation; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=OSteoporosis Index of RISk; OST=Osteoporosis Self-Assessment Tool; OSTA=OST for Asians; SCORE=Simple Calculated Osteoporosis Risk Estimation; Sn=sensitivity; SOFSURF=Study of Osteoporotic Fractures Research Group Study Utilizing Risk Factors; Sp=specificity; USPSTF=U.S. Preventive Services Task Force.

Figure 6b. Accuracy of Risk Assessment Instruments for Identifying Osteoporosis in Men (KQ 2c)

Instrument (Cutoff for Sn and Sp)	No. of Cohorts (Total N)	AUC Range or estimate (● 95% CI) for single study	Sensitivity	Specificity
ABONE (≥2)	1[233] (186)	0.78 (0.64 to 0.93)	100%	28%
FRAX MOF (USPSTF Threshold*)	2[232, 235] (5,541)	0.62 to 0.79	39% to 59%	59% to 89%
FRAX MOF ≥20%	1[233] (186)	0.77 (0.61 to 0.94)	0%	99%
FRAX MOF (other)	1[235] (4,043)	0.62 (NR)	53% to 81%	33% to 65%
FRAX Hip (>3%)	1[233] (186)	0.86 (0.73 to 0.98)	80%	71%
FRAX MOF (≥20% MOF or ≥3% Hip)	2[237, 238] (1,189)	0.65 to 0.72	27% to 69%	54% to 88%
Garvan OF	1[233] (186)	0.72 (0.46 to 0.98)	20%	96%
Garvan Hip	1[233] (186)	0.72 (0.44 to 1.0)	60%	79%
MORES (≥6)	2[197, 202, 232] (3,290)	0.66 to 0.87	58% to 96%	61% to 70%
MOST (≤26 or ≤21†)	1[218] (4,658)	0.81 to 0.88	87% to 89%	50% to 59%
MSCORE or reduced MSCORE (≥9)	1[216] (197)	0.81 to 0.84	85% to 88%‡	57% to 58%‡
ORAI (≥9)	1[233] (186)	0.87 (0.71 to 1.0)	100%	19%
OSIRIS (≤1)	1[233] (186)	0.94 (0.88 to 100)	100%	29%
OST (varied§)	10[198, 201, 216, 218, 219, 221, 235, 237, 240, 243] (9,887)	0.63 to 0.89	40% to 93% (k=8)	25% to 95% (k=8)
OSTA (varied‖)	3[212, 229, 233, 244] (4,171)	0.62 to 0.94	56% to 100%	33% to 68%
SCORE (≥6)	1[233] (186)	0.91 (0.83 to 0.99)	100%	45%
VA-FARA (≥3% hip or ≥20% MOF)	1[237] (463)	0.64 (0.58 to 0.70)	64%	58%

Area Under the Curve



Note: This plot depicts the range of AUC estimates (line with no symbol) from across 2 or more studies OR a single study point estimate and 95% CI when only one study reported an AUC estimate. The number of studies (k) in AUC, sensitivity, and specificity columns is provided when the estimate reported was from a fewer number of studies than what is reported in the second column. Not all studies reported all three outcomes.

Figure 6b. Accuracy of Risk Assessment Instruments for Identifying Osteoporosis in Men (KQ 2c)

[*] MOF risk ≥8.4 percent (2018 USPSTF recommendation) or 9.3 percent (2011 USPSTF recommendation).

[†] Threshold ≤26 for U.S. participants; ≤21 for Hong Kong participants.

[‡] In a separate cohort of 134 African Americans derived from a convenience sample, the sensitivity was either 93 percent or 100 percent and the specificity was either 73 percent or 79 percent depending on whether a Caucasian or African American reference range was used to calculate T-scores. See **Appendix D Table 13** for details.

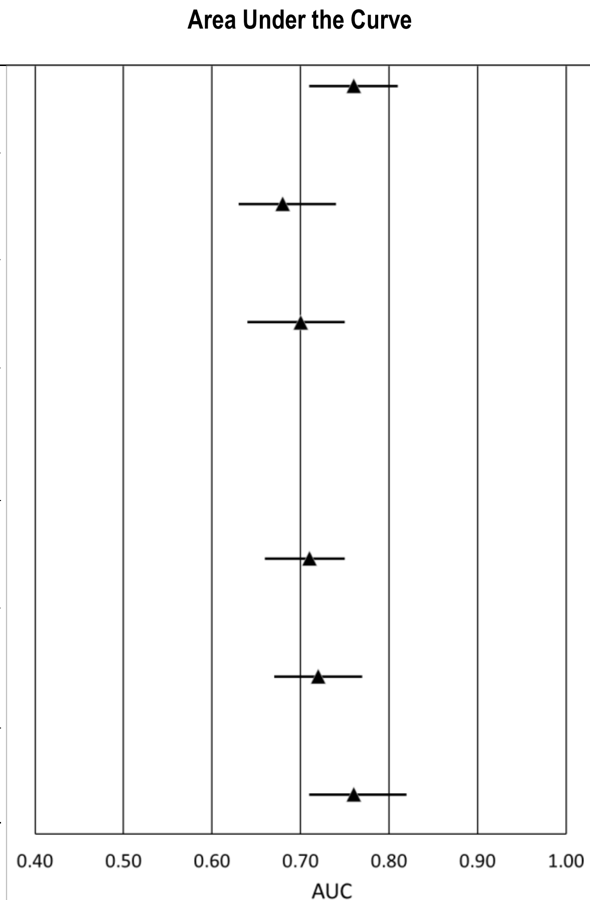
[§] The most common threshold evaluated was <2; however, the following thresholds were also evaluated: ≤6, <3, <1, <0.99, <0, ≤0, <-1, ≤-2, and several studies that did not report thresholds because they only reported AUC.

[‖] The most common threshold evaluated was ≤-1, but <2, ≤1, ≤0, and <0.5 were also evaluated.

Abbreviations: ABONE=Age, Bone, No Estrogen; AUC=area under the curve; CI=confidence interval; FRAX=Fracture Risk Assessment Tool; KQ=key question; MOF=major osteoporotic fracture; MORES=Male Osteoporosis Risk Estimation Score; MOST=Male Osteoporosis Screening Tool; MSCORE=Male Simple Calculated Osteoporosis Risk Estimation; OF=osteoporotic fractures; N/n=number; NR=not reported; ORAI=Osteoporosis Risk Assessment Instrument; OST=Osteoporosis Self-Assessment Tool; OSTA=OST for Asians; SCORE=Simple Calculated Osteoporosis Risk Estimation; Sn=sensitivity; Sp=specificity; USPSTF=U.S. Preventive Services Task Force; VA-FARA=Veterans Affairs Fracture Absolute Risk Assessment.

Figure 6c. Accuracy of Risk Assessment Instruments for Identifying Osteoporosis in Mixed-Sex Populations (KQ 2c)

Instrument (Cutoff for Sn and Sp)	No. of Cohorts (Total N/% Female)	AUC Estimate (▲ 95% CI) for single study	Sensitivity	Specificity
FRAX MOF ≥20%	1[239] (531/45%)	0.76 (0.71 to 0.81)	NR	NR
FRAX MOF (other)	1[196](626/45%)	0.68 (0.63 to 0.74)	90%	35%
FRAX Hip (>3%)	1[196](626/45%)	0.70 (0.64 to 0.75)	92%	37%
FRAX MOF (≥20% MOF ≥3% Hip)	1[241] (45/87%)	NR	100%	91%
Garvan OF*	1[239] (531/45%)	0.71 (0.66 to 0.75)	69%	95%
Garvan Hip*	1[239] (531/45%)	0.72 (0.67 to 0.77)	72%	90%
OST (<2)	1[196] (626/45%)	0.76 (0.71 to 0.82)	91%	40%



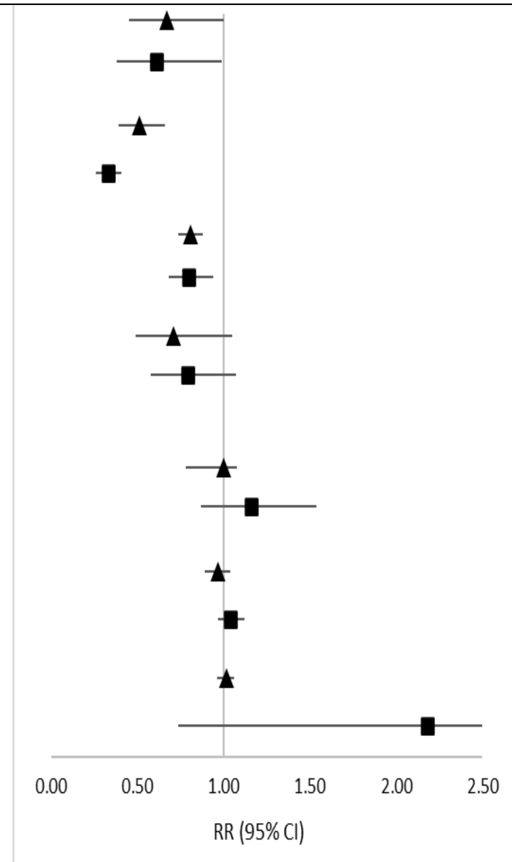
* Used empirically derived, age-based thresholds.

Abbreviations: CI=confidence interval; FRAX=Fracture Risk Assessment Tool; k=number of studies; KQ=key question; MOF=major osteoporotic fracture; OF=osteoporotic fractures; OST=Osteoporosis Self-Assessment Tool; N =number; NR=not reported; Sn=sensitivity; Sp=specificity.

Figure 7. Results of Randomized, Placebo-Controlled Trials of Treatment for Osteoporosis, Fractures, Mortality, and Harms (KQs 4 and 5)

Outcomes	Drug	No. Studies	Total N	n with Event	Pooled RR (95% CI)	ARD per 1,000 (95% CI)
Benefits (KQ 4)						
Hip	Bisphosphonates	6	12,055	98	0.67 (0.45 to 1.00)	3 fewer (5 fewer to 0 fewer)
	Denosumab	2	8,050	69	0.61 (0.38 to 0.99)*	4 fewer (7 fewer to 0 fewer)
Vertebral	Bisphosphonates	10	9,015	250	0.51 (0.39 to 0.66)†	18 fewer (23 fewer to 13 fewer)
	Denosumab	3	8,179	352	0.33 (0.26 to 0.41)*	44 fewer (49 fewer to 39 fewer)
Nonvertebral	Bisphosphonates	13	20,929	1,954	0.81 (0.74 to 0.88)	28 fewer (38 fewer to 18 fewer)
	Denosumab	3	8,382	543	0.80 (0.68 to 0.94)*	14 fewer (23 fewer to 4 fewer)
Mortality	Bisphosphonates	6	3,714	104	0.71 (0.49 to 1.05)	10 fewer (17 fewer to 2 more)
	Denosumab	5	8,828	164	0.79 (0.58 to 1.07)*	4 fewer (9 fewer to 1 more)
Harms (KQ 5)						
Discontinuations due to AE	Bisphosphonates	24	18,617	2,116	1.00 (0.92 to 1.08)	0 fewer (9 fewer to 9 more)
	Denosumab	5	8,826	193	1.16 (0.87 to 1.54)	3 more (3 fewer to 11 more)
Serious AE	Bisphosphonates	21	13,878	2,719	0.97 (0.91 to 1.04)	6 fewer (18 fewer to 8 more)
	Denosumab	6	8,934	2,081	1.04 (0.97 to 1.12)	9 more (7 fewer to 28 more)
Upper GI AE	Bisphosphonates	26	22,172	5,873	1.02 (0.98 to 1.06)	5 more (5 fewer to 16 more)
GI AE	Denosumab	4	932	15	2.18 (0.74 to 6.46)	14 more (3 fewer to 66 more)

▲=Bisphosphonate; ■=Denosumab



* Although multiple studies reported, the evidence base is dominated by one large (N=7,808) study.

† We conducted a sensitivity analysis limiting to studies reporting clinical vertebral fractures (4 studies) and the pooled RR was 0.44 (95% CI, 0.24 to 0.79; 2,373 participants, I^2=0%).

Abbreviations: AE=adverse event; ARD=absolute risk difference; CI=confidence interval; GI=gastrointestinal; KQ=key question; N/n=number; RR=relative risk.

Table 1. Summary of Randomized, Controlled Trials of Screening for Fracture Risk or Osteoporosis (KQs 1 and 3)

Author, Year Trial Name, Registry No.	Recruitment Setting	Mean Age (SD)	N (%) Female	Intervention Groups (N Randomized)	Summary of Results	Study Quality
Rubin et al, 2017[126, 127] ROSE, NCT01388244	Civic registries in southern Denmark	Median 71 [IQR 68, 76]	34,229 (100)	**Screening:** FRAX without BMD assessment with invitation to DXA and VFA if 10-year FRAX MOF risk ≥15%; results sent to the participant and PCP with treatment recommendations based on national guidelines (17,072) **Routine care:** no contact after completion of baseline data collection, usual care guided by PCP (17,157)	aSHR (95% CI) at median followup 5.0 years • MOF (primary study endpoint): 0.99 (0.92 to 1.06) • Hip fracture: 1.00 (0.89 to 1.13) • All osteoporotic fractures excluding some sites:* 1.00 (0.95 to 1.06) • Mortality NR	Fair
Shepstone et al, 2018[120, 121] SCOOP, ISRCTN 55814835	General practice clinics in the U.K.	**Screening:** 75.5 (4.2) Routine Care: 75.5 (4.1)	12,483 (100)	**Screening:** FRAX without BMD assessment; if high-risk based on 10-year FRAX hip risk ≥age-specific threshold, then invitation to DXA; if below threshold, then letter sent to participants and PCPs confirming low risk status; DXA results sent to participant and PCP with participant's revised FRAX risk (including BMD information), age-specific treatment thresholds, and recommendation to discuss treatment if above threshold (6,233) **Routine care:** letter informing PCP of patient's participation in the study; usual care guided by PCP (6,250)	aHR (95% CI) at 5 years followup • All clinical fractures excluding some sites† without regard to trauma: 0.94 (0.86 to 1.03) • Hip fracture: 0.72 (0.59 to 0.89) • All clinical fractures including all sites: 0.94 (0.86 to 1.03) • All-cause mortality: 1.05 (0.93 to 1.19)	Fair
Merlijn et al, 2019[124, 125] SOS NTR2430	General practice registeries in the Netherlands; only women with 1 or more clinical risks‡ were recruited	75.0 (6.7)	11,032 (100)	**Screening:** FRAX without BMD assessment, DXA, VFA, fall risk assessment, and blood chemistries to exclude secondary osteoporosis; women with treatment indications based on results (FRAX with BMD risk above age-dependent threshold, T-score <-2, or prevalent vertebral fracture) had referral to PCP for personalized treatment advice including medication, evaluation for secondary osteoporosis, fall prevention, and calcium/vitamin D supplementation; PCPs were provided group education on the study protocol and treatment options (N=5,516) **Routine care:** wait-list placement for screening intervention; notification to participant and PCP of indication for DXA or VFA if clinical risks present based on existing Dutch guidelines, usual care guided by PCP (N=5,405)	aHR (95% CI) at mean followup 3.7 years • Any clinical fracture: 0.97 (0.87 to 1.08) • Hip fracture: 0.91 (0.71 to 1.15) • Major osteoporotic fracture: 0.91 (0.80 to 1.04) • Any osteoporotic fracture: 0.91 (0.81 to 1.03) • All-cause mortality: 1.03 (0.91 to 1.17)	Fair

Table 1. Summary of Randomized, Controlled Trials of Screening for Fracture Risk or Osteoporosis (KQs 1 and 3)

* Excluding fingers, toe, skull, and face.

† Excluding hands, feet, nose, skull, or cervical vertebrae.

‡ Clinical risk factors: previous fracture after age 50, parental hip fracture, BMI <19 kg/m^2, rheumatoid arthritis, menopause <45 years, malabsorption syndrome, chronic liver disease, type 1 diabetes, immobility.

Abbreviations: aHR=adjusted hazard ratio; aSHR=adjusted subhazard ratio; BMD=bone mineral density; BMI=body mass index; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; FRAX=Fracture Risk Assessment Tool; IQR=interquartile range; ISRCTN=International Standard Randomised Controlled Trial Registry; KQ=key question; MOF=major osteoporotic fracture; N=number of participants; NCT=National Clinical Trial; NR=not reported; NTR=Dutch National Trial Register; PCP=primary care provider; ROSE=Risk-stratified Osteoporosis Strategy Evaluation; SCOOP=Screening in the Community to Reduce Fractures in Older Women; SD=standard deviation; SOS=Stichting Artsen Laboratorium enTrombosedienst (SALT) Osteoporosis Study; U.K.=United Kingdom; VFA=vertebral fracture assessment.

Table 2. Randomized, Placebo-Controlled Trials of Treatment for Osteoporosis (KQs 4 and 5)

Author, Year	Study Quality	Total N	% Female	Mean Age (SD)	Race/Ethnicity	% With Prior Fracture*	T-Score Inclusion Criteria	Dose and Duration	Key Question
Alendronate									
Adachi et al, 2009[296]	Fair	438	100%	65.5 (NR)	89% White, 8% Hispanic, 3% Asian, 1% Black	6.8%	<-2.0	10 mg per day; 3 months	KQ 5
Ascott-Evans et al, 2003[252]	Fair	144	100%	57.3 (6.6)	91.7% White, 8.3% other	0%	LS <-1.5 and >-3.5	10 mg per day; 1 year	KQ 4, KQ 5
Bell et al, 2002[276]	Fair	65	100%	66 (NR)	100% African American	NR	LS <-1.75	10 mg per day; 2 years	KQ 4, KQ 5
Bone et al, 1997[277]	Fair	359	100%	71 (NR)	97% White	34% to 42%	≤-2.0	1, 2.5, or 5 mg per day; 2 years	KQ 4, KQ 5
Chesnut et al, 1995[253]	Fair	188	100%	62.9 (6.1)	97.9% White, 2.1% Asian	0%	NR; mean T-score -1.1	Various;[†] 2 years	KQ 4, KQ 5
Cryer et al, 2005[297]	Fair	454	100%	65 (10)	91% White, 2% Black, 1% Asian, 5% Hispanic, 1% Native American, 1% other	NR	Any site <-2.0 and >-3.5	70 mg weekly; 6 months	KQ 5
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255] FIT	Good	4,432[‡]	100%	67.6 (6.2)	97% White	0%[‡]	FN <-1.6	5 mg per day for 2 years, then 10 mg per day for 1 year; 3 years	KQ 4, KQ 5
Devogelaer et al, 1996[302]	Fair	516	100%	62 (NR)	NR	NR	LS ≤-2.5	5, 10, 20[§] mg per day; 3 years	KQ 5
Eisman et al, 2004[298]	Good	449	93%-96%	63.6 (NR)	65.7% White, 18% Asian, 12% Hispanic, 5% other	NR	NR; mean T-score NR	70 mg weekly; 3 months	KQ 5
Greenspan et al, 2002[299]	Fair	450	92%	67 (NR)	96% White	NR	NR; mean T-score NR	70 mg weekly; 3 months	KQ 5
Greenspan et al, 2003[300]	Good	186	100%	71.5 (NR)	NR	0%	NR; mean T-score -1.7	10 mg per day; 3 years	KQ 5
Hosking et al, 2003[267]	Fair	549[ǁ]	100%	69 (NR)	99.5% Caucasian	48.5%	LS or TH <-2.5 or both <-2.0	70 mg weekly; 1 year	KQ 4, KQ 5
Johnell et al, 2002[294]	Fair	331	100%	63.6 (NR)	95% White	NR	FN <-2.0	10 mg per day; 1 year	KQ 5

Table 2. Randomized, Placebo-Controlled Trials of Treatment for Osteoporosis (KQs 4 and 5)

Author, Year	Study Quality	Total N	% Female	Mean Age (SD)	Race/Ethnicity	% With Prior Fracture*	T-Score Inclusion Criteria	Dose and Duration	Key Question
Liberman et al, 1995[256]	Fair	994	100%	64 (NR)	87.4% White, 0.4% Black, 12.2% other	21%	LS <-2.5	5 or 10 mg per day; 3 years 20 mg per day for 2 years followed by 5 mg/day for 1 year	KQ 4, KQ 5
Pols et al, 1999[259]	Fair	1,908	100%	62.8 (7.5)	94% White	NR	NR; mean T-score	10 mg per day; 1 year	KQ 4, KQ 5
Tucci et al, 1996[268]	Fair	478	100%	64 (NR)	91% White, 8% Asian	NR	LS <-2.5	5 mg, 10 mg, or 20 mg per day for 2 years followed by 5 mg per day; 3 years	KQ 4, KQ 5
Ibandronate									
Chapurlat et al, 2013[288]	Fair	148	100%	62.7 (5.0)	NR	NR	LS or TH <-1.0 and >-2.5	150 mg per month; 2 years	KQ 5
McClung et al, 2009[275]	Fair	160	100%	53 (NR)	NR	0%	LS <-1.0 and >-2.5 with TH or FN >-2.5	150 mg per month; 1 year	KQ 4, KQ 5
McClung et al, 2004[291]	Fair	653	100%	58.2 (8.6)	NR	0%	LS <-1.0 and >-2.5	0.5 mg, 1.0 mg, or 2.5 mg per day; 2 years	KQ 5
Ravn et al, 1996[264]	Fair	180	100%	65 (NR)	100% White	0%	NR; mean T-score 1.7	0.25 mg, 0.50 mg, 1.0 mg, 2.5 mg, or 5.0 mg per day; 1 year	KQ 4, KQ 5
Reginster et al, 2005[266]	Fair	144	100%	65.7 (NR)	NR	NR	NR; mean T-score -0.3 to -1.9	Various;¶ 3 months	KQ 4, KQ 5
Riis et al, 2001[265]	Fair	240	100%	66.8 (4.9)	NR	NR	LS or FN <-2.5	2.5 mg per day or intermittent cyclic dose; 2 years	KQ 4, KQ 5
Tanko et al, 2003[29]	Fair	630	100%	55 (NR)	NR	0%	LS ≥-2.5	5 mg, 10 mg, or 20 mg weekly; 2 years	KQ 5
Thiebaud et al, 1997[293]	Fair	126	100%	64 (NR)	NR	0%	LS <-2.5	0.25, 0.5 mg, 1.0, or 2.0 mg every 3 months; 1 year	KQ 5
Risedronate									
Hosking et al, 2003[267]	Fair	549‖	100%	69 (NR)	99.5% Caucasian	48.5%	LS or TH <-2.5 or both <-2.0	5 mg daily; 3 months	KQ 4, KQ 5
McClung et al, 2001[257]	Fair	9,331	100%	NR, all age 70 or older	98% White	39% to 44%	FN <-4 or <-3 with risk factor for hip fracture	2.5 or 5 mg per day; years	KQ 4, KQ 5
Mortensen et al, 1998[258]	Fair	111	100%	52.1 (3.9)	100% White	0%	Z-score >-2.0; mean T-score -1.1	5 mg cyclic or 5 mg per day; 2 years	KQ 4, KQ 5

Table 2. Randomized, Placebo-Controlled Trials of Treatment for Osteoporosis (KQs 4 and 5)

Author, Year	Study Quality	Total N	% Female	Mean Age (SD)	Race/Ethnicity	% With Prior Fracture*	T-Score Inclusion Criteria	Dose and Duration	Key Question
Shiraki et al, 2003[2]	Fair	211	99%	60.3 (NR)	100% Japanese	Mean prevalent vertebral fractures: 0.3 (SD 0.8)	LS <-2.5 without vertebral fracture; <-1.5 with vertebral fracture	1 mg, 2.5 mg, or 5 mg per day; 8 months	KQ 5
Valimaki et al, 2007[261]	Fair	170	100%	65.9 (6.8)	100% White	NR	LS >-2.5 and <-1 and proximal femur ≤-1	5 mg per day; 2 years	KQ 4, KQ 5
Zoledronic Acid									
Boonen et al, 2012[251]	Good	1,199	0%	Median 66	94% White, 1% Black, 1% Asian, 0.5% other	32%	TH or FN ≤-1.5	5 mg every year; 2 years	KQ 4, KQ 5
Grey et al, 2010[262] Grey et al, 2009[263]	Fair	50	100%	62 (8)	NR	42%	LS or TH <-1 and >-2	5 mg; single dose with 3 year followup	KQ 5
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	Fair	180	100%	66 (9)	NR	14% to 21%	LS or TH <-1 and >-2.5	1 mg, 2.5 mg, 5 mg; single dose	KQ 4, KQ 5
McClung et al, 2009[287]	Fair	581	100%	59.6 to 60.5	NR	0%	LS -1.0 and -2.5 and FN >-2.5	5 mg single dose or 5 mg yearly for 2 years; 2 years	KQ 5
Reid et al, 2002[260]	Fair	351	100%	65 (7)	95% White	0%	LS <-2.0	Various;[#] 1 year	KQ 4, KQ 5
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]	Good	2,000	100%	71 (5.1)	95% European, 0.02% Maori, 0.01% Pacific Islander, 0.02% East Asian, 0.005% Indian, 0.002% other	23.7%	TH or FN -1.0 to -2.5	5 mg every 18 months; 6 years	KQ 4, KQ 5
Denosumab									
Bone et al, 2008[285]	Fair	332	100%	59.4 (7.5)	NR	0%	LS or TH between -1 and -2.5	60 mg every 6 months; 3 years	KQ 4, KQ 5
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[28] McCloskey et al, 2012[282] Palacios et al, 2015[283] FREEDOM	Fair	7,808	100%	72.3 (5.2)	NR	50%	LS or TH <-2.5 but >-4.0	60 mg every 6 months; 3 years	KQ 4, KQ 5
Koh et al, 2016[286]	Fair	135	100%	67.0 (4.9)	NR	23% to 30%	TH or LS <-2.5 and ≥-4.0	60 mg; single dose with 6 month followup	KQ 4, KQ 5
Lewiecki et al, 2007[284] McClung et al, 2006[304]	Fair	365	100%	62.5 (8.1)	86.2% White, 9.5% Hispanic, 2.9% Black, 1.5% other	0%	LS -1.8 to -4.0 or FN -1.8 to -3.5	Various;** 2 years	KQ 4, KQ 5

Table 2. Randomized, Placebo-Controlled Trials of Treatment for Osteoporosis (KQs 4 and 5)

Author, Year	Study Quality	Total N	% Female	Mean Age (SD)	Race/Ethnicity	% With Prior Fracture*	T-Score Inclusion Criteria	Dose and Duration	Key Question
Nakamura et al, 2012[279]	Fair	226	100%	65.1 (6.8)	100% Japanese	34%	LS -2.5 to -4.0 or FN or TH -2.5 to -3.5	Various;[††] 1 year	KQ 4, KQ 5
Orwoll et al, 2012[27] ADAMO	Fair	242	0%	65.0 (9.8)	94.2% White	39.3%	LS or FN -2.0 to -3.5;[‡‡] or LS or FN 1.0 to -3.5[‡‡] with prior MOF	60 mg every 6 months; 2 years	KQ 4, KQ 5

* Studies define this in varying ways: any fracture, fracture after age 50, fragility fracture, and vertebral fracture only.

† 5 mg/day or 10 mg/day or 40 mg/day for 3 months, then 2.5 mg/day for 21 months; 20 mg/day for 1 year, then placebo for 1 year; 40 mg/day for 1 year, then placebo for 1 year.

ǂ Only the portion of the enrolled population without prior vertebral fracture was used for this review.

§ Dosage was 20 mg for first 2 years and lowered to 5 mg in the final year.

‖ Includes the alendronate, risedronate, and placebo arms.

¶ 50 mg per month; 50 mg for the first month, then 100 mg for months 2–3; 100 mg per month; 150 mg per month.

\# 0.25 mg every 3 months, 0.5 mg every 3 months, 1 mg every 3 months, 4 mg every 1 year, 2 mg every 6 months.

** 6 mg, 14 mg, or 30 mg every 3 months; 14 mg, 60 mg, 100 mg, or 210 mg every 6 months.

†† 14 mg, 60 mg, or 100 mg every 6 months.

ǂǂ T-scores based on male reference range.

Abbreviations: ADAMO= Study to Compare the Efficacy and Safety of DenosumAb Versus Placebo in Males With Osteoporosis; FIT=Fracture Intervention Trial; FN=femoral neck; FREEDOM=Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months KQ=key question; LS=lumbar spine; MOF=major osteoporotic fracture; N/n= number; NR=not reported; SD=standard deviation; TH=total hip.

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
1 Benefits of Screening	Fractures	3 RCTs[120-129] (42,009 using ROSE per protocol 1 population) 3 SRs[120, 124, 126, 130-133]	***Hip fractures*** pooled RR, 0.83 (95% CI, 0.73 to 0.93); ARD, 5 fewer per 1,000 (95% CI, from 7 fewer to 2 fewer) ***MOF*** pooled RR, 0.94 (95% CI, 0.88 to 0.99); ARD, 6 fewer per 1,000 (95% CI, from 12 fewer to 1 fewer) ***Osteoporotic fractures*** pooled RR, 0.95 (95% CI, 0.91 to 1.01); ARD, 6 fewer per 1,000 (95% CI, from 11 fewer to 1 more) Estimates from SRs consistent.	Consistent, precise for hip and MOF, imprecise for osteoporotic fracture	Modest screening uptake and adherence to treatment among those treated; contamination in control groups; followup for only 3.7 to 5 years	Moderate* for benefit on MOF and hip fracture; Low† for benefit on osteoporotic fractures	Two-stage screening used by 2 studies; European women age 60 years or older at high baseline fracture risk; extensive screening battery (imaging, labs, falls assessment) used in 1 study
	Mortality	3 RCTs[120-129 120, 124, 126] (57,633) 1 SR[131, 132]	Pooled RR, 0.99 (95% CI, 0.95 to 1.04) ARD, 1 fewer per 1,000 (95% CI, from 5 fewer to 4 more) Estimates from SR consistent.	Consistent, imprecise	Same as above	Low† for no effect	Same as above
2a Predictive Accuracy of Risk Assessment Instruments	Calibration (MOF and Hip Fracture)	Two SRs[131, 132, 137] and 25 cohorts reported in 40 articles[72, 73, 139, 141-158, 160, 161, 163-173, 176, 177, 181, 311-313] (Unable to estimate precisely due to overlap in reporting for some cohorts)	Reported for 6 instruments: FRAX, FREM, FRC, Garvan, OST, QFracture FRAX (28 articles from 20 unique cohorts): reasonably calibrated in some cohorts and poorly calibrated in other cohorts. Too few studies and outcomes reported for instruments other than FRAX.	Varied by instrument	All studies high risk of bias	Low‡ for FRAX for poor to modest calibration Insufficient§ for FRC, FREM, Garvan, OST, QFracture	Studies included postmenopausal women and men

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
2a Predictive Accuracy of Risk Assessment Instruments (continued)	Discrimination (MOF and Hip Fracture)	Four SRs[134-137] 16 cohorts published in 25 articles[73, 139-143, 146, 159, 160, 163, 165, 172-181, 311, 312] (Unable to estimate precisely due to overlap in reporting for some cohorts)	Reported for 11 instruments: EPIC, FRAX, FRC, FREM Garvan Fracture Risk Calculator, ORAI, OSIRIS, OST, QFracture, SCORE, WHI Prediction Model. AUC range Younger women (<65): 0.52 to 0.71 Women: 0.63 to 0.89 Men: 0.63 to 0.93 Mixed sex: 0.61 to 0.88 FRAX, FRC, and Garvan instruments with BMD had higher AUCs compared with same instrument without BMD. AUCs higher for prediction of hip fracture compared with MOF for FRAX, FRC and QFracture, and Garvan.	Varied by instrument	All studies high risk of bias for development cohorts and for external validation cohorts	Low[‖] for FRAX, FRC, Garvan, QFracture for poor to modest discrimination Insufficient[§] for EPIC, FREM, OST, SCORE, WHI	Studies included postmenopausal women and men, but not for all instruments
2b Predictive Accuracy of BMD	Calibration (MOF and Hip Fracture)	4 articles from 4 unique cohorts[146, 150, 185, 192] (18,145)	Consistent calibration measures not reported across studies; calibration poor in some studies and good in others for prediction of MOF or hip fracture.	Inconsistent; unable to judge precision	Not the primary aim of any study; not enough fracture events in some studies, particularly for hip fractures	Insufficient[¶]	Cohorts include both men and women; persons with known osteoporosis or on treatment excluded from some cohorts; BMD typically measured at FN
2b Predictive Accuracy of BMD (continued)	Discrimination (MOF and Hip Fracture)	18 articles from 16 unique cohorts[146, 150, 151, 154, 156, 157, 166, 172, 175, 176, 179, 183, 185, 187, 189, 190, 193, 194](101,446)	AUC range MOF: 0.60 to 0.80 (13 cohorts; 15 estimates) Hip: 0.64 to 0.86 (12 cohorts; 14 estimates) Threshold T-score <-2.5 Sn MOF: 17.5% to 51.3% (5 studies) Sn Hip: 25.0% to 66.7% (5 studies) Sp MOF: 70.9% to 95.4% (3 studies) Sp Hip: 88.6% to 94.0% (4 studies)	Inconsistent, precise	10 analyses were high ROB; predictive accuracy of BMD not the primary aim of any study	Low[‖] for poor to modest discrimination	Same as above

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
2c Diagnostic Accuracy (Women and Men)	FRAX/ Discrimination	MOF risk 15 studies from 12 unique cohorts[141, 143, 195, 196, 200, 220, 232-239, 24] (37,756/85% women) Hip fx risk 3 studies from 3 unique cohorts[196, 233, 239] (1,710/52% women)	MOF (9.3% or 8.4% risk threshold) Women ages 50–64 (3 estimates) AUC: 0.55 to 0.62 Sn: 5% to 49%; Sp: 63% to 96% Men (2 estimates) AUC: 0.62 to 0.79 Sn: 39% to 59%; Sp: 59% to 89% MOF (>20% risk threshold) Women age ≥60 (1 estimate) AUC: 0.71 (95% CI, 0.60 to 0.82) Sn: 17%; Sp: 96% Men (1 estimate) Sn: 0%; Sp: 99% Mixed sex (1 estimate) AUC: 0.76 (95% CI, 0.71 to 0.81) MOF (various thresholds or no threshold) Women ages 50–64 (2 estimates) AUC: 0.64 to 0.72 Men (1 estimate) AUC: 0.62 Mixed sex (1 estimate) AUC: 0.68 (95% CI, 0.63 to 0.72) Hip (>3% risk threshold) Women age ≥60 (1 estimate) AUC: 0.75 (95% CI, 0.65 to 0.86) Sn: 83%; Sp: 54% Men (1 estimate) AUC: 0.86 (95% CI, 0.73 to 0.98) Sn: 80%; Sp: 71% Mixed sex (1 estimate) AUC: 0.70 (95% CI, 0.64 to 0.75)	Inconsistent; precise	Heterogeneity in BMD sites measured; all but 1 fair quality because of unclear methods for patient selection and risk for selection bias, lack of blinding of index or reference test results, unclea BMD reference range used for T-score, unclear interva between risk assessment and BMD measurement	Low[‖] for poor to modest discrimination	Men and post-menopausal women from community or clinic-based populations; FRAX risk assessment without BMD input into calculation; some studies used EHR data to determine FRAX risks

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
2c Diagnostic Accuracy	OST/ Discrimination	31 studies from 29 cohorts [143, 159, 195, 196, 198-201, 203, 205, 208-210, 214, 216-219, 221, 223, 225, 228, 230, 231, 234-237, 240, 242, 24] (80,592/82% women)	AUC (95% CI) or range: Women (20 estimates): 0.32 to 0.89 Women ages 50 to 64 (3 estimates): 0.63 to 0.75 Men (10 estimates): 0.63 to 0.89 Mixed sex (1 estimate): 0.76 (0.71 to 0.82) At a score threshold of <2: Women (11 estimates) Sn range: 53% to 95% Sp range: 37% to 72% Women ages 50 to 64 (3 estimates) Sn range: 56% to 79% Sp range: 56% to 70% Men (7 estimates) Sn range: 62% to 89% Sp range: 36% to 74%	Inconsistent; precise	All but 1 fair quality; similar limitations as for FRAX above	Low[‖] for poor to modest discrimination	Men and post-menopausal women from community or clinic-based populations
2c Diagnostic Accuracy (Women)	Other risk assessments/ Discrimination	29 studies from 26 cohorts[195, 199, 203, 204, 206-211, 213-215, 217, 222-228, 230, 231, 233, 234, 236, 239, 242, 24] (30,621)	AUC range: 0.32 to 0.87 (25 estimates) Across various thresholds: Sn: 28% to 100% (24 estimates) Sp: 5% to 100% (24 estimates)	Inconsistent; precision varies by instrument	All fair quality, similar limitations as for FRAX above	Low[‖] for poor to modest discrimination (ABONE, NOF, ORAI, OSIRIS, OSTA, SCORE) Insufficient[¶] (AMMEB, Garvan FRC, SOFSURF)	Postmenopausal women from community and clinic-based populations

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
2c Diagnostic Accuracy (Men)	Other risk Assessments/ Discrimination	21 studies[196-198, 201, 202, 212, 216, 218, 219, 221, 229, 232, 233, 235, 237, 238, 240, 241, 243, 244 196-198, 201, 20 212, 216, 218, 219, 221, 229, 232, 233, 235, 237, 238, 240, 241, 243, 244] (24,258)	AUC range: 0.64 to 0.88 in the studies exclusively enrolling men and evaluating instruments developed specifically for men; AUC range: 0.62 to 0.94 from the male population component of the studies with mixed populations.	Inconsistent, precision varies by instrument	All but 1 study fair quality; similar limitations as for FRAX above	Low[‖] for poor to modest discrimination (FRAX, MORES MOST, OST, OSTA) Insufficient[¶] (ABONE, Garvan FRC, MSCORE, ORAI, OSIRIS, SCORE, VA-FARA)	Men mostly from clinic-based populations
2d Repeat Screening	BMD at baseline and repeat BMD	5 studies[246-250] (19,957)	Predictive accuracy of repeat BMD at 4 to 8 years after initial BMD was similar to predictive accuracy of initial BMD for predicting MOF and hip fractures over followup of 8 to 11 years after repeat BMD.	Consistent; precise	Two studies were poor quality; 3 were fair quality; indirect evidence	Moderate[#] for nc added value of repeat DXA	1 study exclusively in men; 1 study with 40% men; mean age 60 to 75 years across studies
3 Harms of Screening	Anxiety	1 RCT[120] (12,483)	No difference in anxiety between screening and control participants over 5 years (P=0.515).	Single study, consistency unknown; Precision unknown	Fair-quality pragmatic trial; only modest uptake and adherence of intervention	Insufficient[§]	Two-stage screening approach in U.K. women ages 70 to 85 years
3 Harms of screening (continued)	Overdiagnosis	1 SR[131] (NA)	Based on data from 2 included RCTs, overdiagnosis estimated to range from 11.8% to 24.1%.	Single review, consistency unknown; Precision unknown	Good quality SR; however, included RCTs are fair quality; method for estimating overdiagnosis for being labeled as “high risk” is evolving	Insufficient (based on extrapolations)	Two-stage screening in U.K. women ages 70 to 85 years in 1 study; Dutch women age 60 years or older at high baseline fracture risk and used extensive screening battery (imaging, labs, falls assessment) in other study

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
4 Benefits of Treatment	Bis-phosphonates Vertebral Fx (clinical and radiographic)	10 RCTs[251-254, 25 258, 260, 261, 269, 277] (9,015)	Pooled RR, 0.51 (95% CI, 0.39 to 0.66); ARD, 18 fewer per 1,000 (95% CI, from 23 fewer to 13 fewer)	Consistent, precise	Most studies fair quality; evidence dominated by 3 of the larger studies; 5 studies had zero events in at least 1 study arm	Moderate* for benefit	Only 1 study in men; the rest were in mostly White post-menopausal women with low bone mass or osteoporosis
4 Benefits of Treatment	Bis-phosphonates Nonvertebral Fx	13 RCTs[251, 252, 254, 256-261, 268, 269, 272, 277] (20,929)	Pooled RR, 0.81 (95% CI, 0.74 to 0.88); ARD, 28 fewer per 1,000 (95% CI, from 38 fewer to 18 fewer)	Consistent, precise	Most studies fair quality, evidence dominated by 6 larger studies; 2 studies had zero events in at least 1 arm	Moderate* for benefit	Only 1 study in men; the rest were in mostly White post-menopausal women with low bone mass or osteoporosis
4 Benefits of Treatment (continued)	Bis-phosphonates Hip Fx	6 RCTs[254, 256-259, 269] (12,055)	Pooled RR, 0.67 (95% CI, 0.45 to 1.0); ARD, 3 fewer per 1,000 (95% CI, from 5 fewe to 0 fewer)	Consistent, imprecise	Most studies fair quality; none were powered to evaluate hip fractures; 1 study had zero events in at least 1 arm	Low† for benefit	All studies in mostly White postmenopausal women with low bone mass or osteoporosis
	Bis-phosphonates Mortality	6 RCTs[251, 264-266, 269, 275] (3,714)	Pooled RR, 0.71 (95% CI, 0.49 to 1.05); ARD, 10 fewer per 1,000 (95% CI, from 17 fewer to 2 more)	Consistent, imprecise	Same as above	Low† for benefit	Only 1 study in men, the rest were in mostly White post-menopausal women with low bone mass or osteoporosis

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
4 Benefits of Treatment (continued)	Denosumab Vertebral Fx	4 RCTs[278-280, 285] (8,179)	Evidence base dominated by FREEDOM study (n=7,808 women), RR, 0.32 (95% CI, 0.26 to 0.41), ARD, 48 fewer per 1,000 participants (95% CI, from 52 fewer to 42 fewer) All other studies with 0 to 1 event per arm; pooled RR across all 4 RCTs, 0.33 (95% CI, 0.26 to 0.41); ARD, 44 fewer per 1,000 persons (95% CI, from 49 fewer to 39 fewer)	Consistent, precise	All studies fair quality; evidence dominated by 1 study; outcome included both clinical and asymptomatic morphometric fractures	Moderate* for benefit	Post-menopausal women with osteoporosis or low bone mass; 1 study was only in men but had only 1 fracture event
	Denosumab Nonvertebral Fx	3 RCTs[278, 280, 285] (8,382)	Evidence base dominated by FREEDOM study (n=7,808 women), RR, 0.80 (95% CI, 0.67 to 0.95); ARD, 15 fewer per 1,000 participants (95% CI, from 24 fewer to 4 fewer); Across all 3 RCTs, pooled RR, 0.80 (95% CI, 0.68 to 0.94), ARD, 14 fewer per 1,000 (95% CI, from 23 fewer to 4 fewer)	Consistent, imprecise	Fair quality studies; evidence dominated by 1 large study	Low† for benefit	Postmenopausal women with osteoporosis or low bone mass; 1 trial was only in men but had only 3 events
4 Benefits of Treatment (continued)	Densoumab Hip Fx	2 RCTs[278, 280] (8,050)	Evidence base dominated by FREEDOM study (n=7,808 women), 0.7% vs. 1.1%; RR, 0.60 (95% CI, 0.37 to 0.97), ARD, 4 fewer per 1,000 (95% CI, from 7 fewer to 0 fewer) 0 events in the other trial involving 242 men Across both studies, pooled RR, 0.61 (95% CI, 0.38 to 0.99), ARD 4 fewer per 1,000 (95% CI, from 7 fewer to 0 fewer)	Consistent, imprecise	Fair quality; large trial with uncertainties in random-ization allocation concealment, blinding, and attrition; no events in the other trial	Low† for benefit	Postmenopausal women with osteoporosis or low bone mass; smaller trial was only in men but had no fracture events
	Denosumab Mortality	5 RCTs[278, 280, 284-286] (8,828)	Pooled RR, 0.79 (95% CI, 0.58 to 1.07); ARD, 4 fewer per 1,000 (95% CI, from 9 fewer to 1 more)	Consistent, imprecise	Fair quality; some un-certainties in randomization for 3 studies, allocation concealment in 4 studies, and attrition and blinding in 2 studies	Low† for benefit	Postmenopausal women with osteoporosis or low bone mass; 1 trial only in men but had only 2 events

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
5 Harms of Treatment	Bis-phosphonates Discontinuations due to AEs	27 RCTs[252-254, 256-261, 264, 266-268, 27, 275-277, 288, 291-294, 29, 299, 302] (18,617)	Based on 24 RCTs: Pooled RR, 1.00 (95% CI, 0.92 to 1.08); ARD, 0 fewer per 1,000 (95% CI, from 9 fewe to 9 more)	Consistent, precise	Most studies fair quality, none powered for this outcome	Moderate* for nc effect	Mostly White postmenopausal women with low bone mass or osteoporosis
	Bis-phosphonates SAEs	22 RCTs[251, 257, 259-261, 264, 266-268, 272, 275, 276, 287, 288, 291-293, 295-297, 299, 302] (13,878)	Based on 21 RCT comparisons: Pooled RR, 0.97 (95% CI, 0.91 to 1.04); ARD, 6 fewer per 1,000 (95% CI, from 18 fewer to 8 more)	Consistent, precise	Most studies fair quality, none powered for this outcome, not long enough to detect rare harms	Moderate† for nc effect	Only 1 study exclusively in men, the rest were in mostly White post-menopausal women with low bone mass or osteoporosis
5 Harms of Treatment (continued)	Bis-phosphonates Upper GI AEs	26 RCTs[252, 256-25, 261, 264, 266-269, 272, 275, 276, 289, 291-300, 30] (22,280)	Pooled RR, 1.02 (95% CI, 0.98 to 1.06); ARD, 5 more per 1,000 (95% CI, from 5 fewer to 16 more)	Consistent, precise	Most studies fair quality, none powered for this outcome	Moderate* for nc effect	Mostly White postmenopausal women with low bone mass or osteoporosis
	Denosumab Discontinuations due to AEs	5 RCTs[278, 280, 284-286] (8,826)	Pooled RR, 1.16 (95% CI, 0.87 to 1.54); ARD, 3 more per 1,000 (95% CI, from 3 fewer to 11 more)	Consistent, imprecise	Fair quality; some un-certainties in random-ization for 3 studies, allocation concealment ir 4 studies, and attrition and blinding in 2 studies.	Low† for no effect	Postmenopausal women with osteoporosis or low bone mass

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

Key Question	Intervention or Test/ Outcome	No. of Studies (No. of Participants)	Summary of Findings	Consistency and Precision	Limitations	Strength of Evidence	Applicability
	Denosumab Serious AEs	6 RCTs[278-280, 284-286] (8,934)	Pooled RR, 1.04 (95% CI, 0.97 to 1.12); ARD, 9 more per 1,000 (95% CI, from 7 fewer to 28 more)	Consistent, imprecise	Fair quality; some uncertainty for allocation concealment in all studies, random-ization in 4 studies, and attrition and masking in 2 studies; not large enough or long enough to detect rare harms	Low† for no effect	Postmenopausal women with osteoporosis or low bone mass
5 Harms of treatment (continued)	Denosumab Upper GI AEs	4 RCTs [279, 280, 284-286] (932)	Pooled RR, 2.18 (95% CI, 0.74 to 6.46); ARD, 14 more per 1,000 (95% CI, from 3 fewer to 66 more)	Consistent, imprecise	Fair quality; some uncertainty for allocation concealment in all studies, randomiza-tion in 3 studies, and attrition and masking in 1 study	Low† for harm	Postmenopausal women with osteoporosis or low bone mass

* Rated down 1 level for study limitations.

† Rated down 1 level for imprecision and 1 level for study limitations.

‡ Rated down 1 level for inconsistency and 1 level for study limitations.

§ Not enough data to evaluate SOE.

‖ Downgraded 1 level for study limitations and 1 level for inconsistency.

¶ Downgraded 1 level for study limitations, 1 level for inconsistency, and 1 level for imprecision.

Downgraded 1 level for study limitations, including indirectness as these study designs did not directly compare a strategy of repeat screening with single screening.

Abbreviations: ABONE=Age, Bone, No Estrogen instrument; AE=adverse event; AMMEB=Age, years after Menopause, age at Menarche Index; ARD=absolute risk difference; AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; EHR=electronic health record; FN=femoral neck; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; FREEDOM=Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months; Fx=fracture; GI=gastrointestinal; KQ=key question; MOF=major osteoporotic fracture; MORES=Male Osteoporosis Risk Estimation Score; MOST=Male Osteoporosis Screening Tool;

Table 3. Summary of Evidence: Screening for Osteoporosis to Prevent Fracture

MSCORE=Male Simple Calculated Osteoporosis Risk Estimation; NA=not available; NOF=National Osteoporosis Foundation tool; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=OSteoporosis Index of RISk; OST=Osteoporosis Self-Assessment Tool; OSTA=OST for Asians; RCT=randomized, controlled trial; ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation; RR=relative risk; SAE=serious adverse event; SCORE=Simple Calculated Osteoporosis Risk Estimation; Sn=sensitivity; SOFSURF=Study of Osteoporotic Fractures Research Group Study Utilizing Risk Factors; Sp=specificity; U.S. Preventive Services Task Force; VA-FARA=Veterans Affairs Fracture Absolute Risk Assessment; vs.=versus; WHI, Womens Health Initiative Prediction Model.

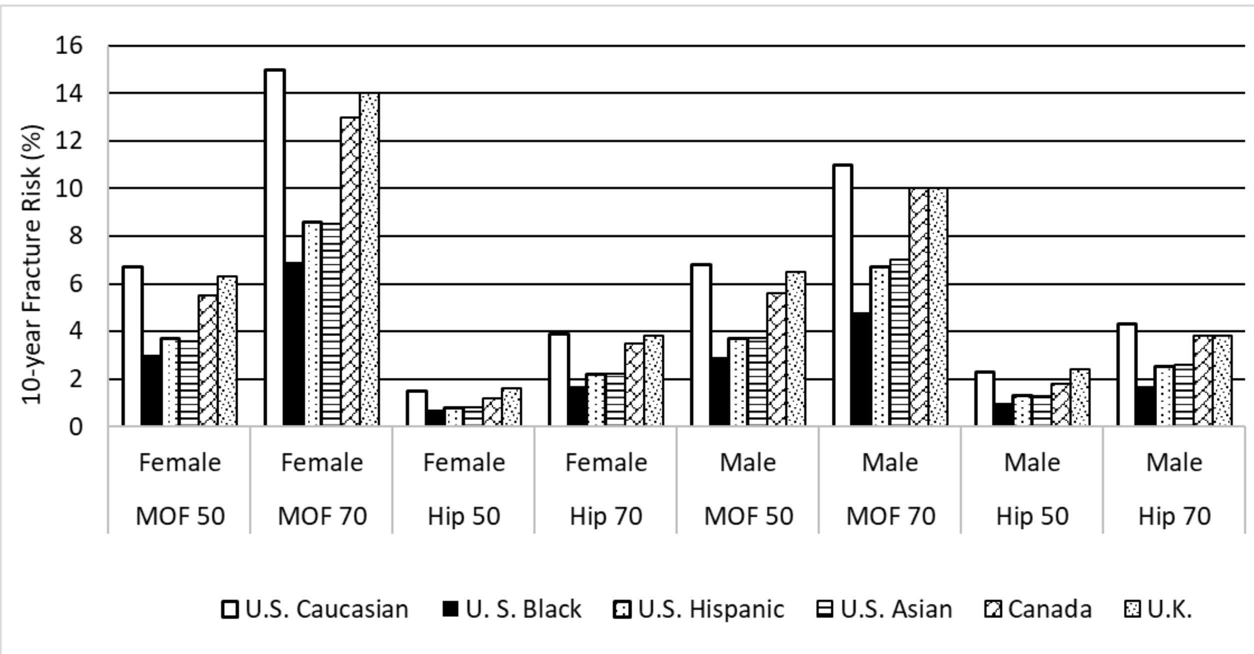
Prevalence and Burden

Appendix A Table 1. Age-Standardized Hip Fracture Incidence (per 1,000 Person-Years) in a Large Cohort (N=1,841,263) of Medicare Advantage Enrollees[35]

Year	Age Strata	Overall	Males	Females
2007	50 to 64	5.90	4.32	7.39
	≥65	20.51	12.00	27.49
2013	50 to 64	5.67	4.20	7.09
	≥65	17.01	10.72	22.08
2017	50 to 64	6.03	4.33	7.73
	≥65	19.35	12.04	24.92

Abbreviations: N=number.

Appendix A Figure 1. FRAX Estimates of the 10-Year Risk for Major Osteoporotic Fracture and Hip Fracture at a Bone Mineral Density T-Score of -2.5 and Body Mass Index of 25 at Ages 50 and 70 Years With No Other Clinical Risks[322]



Abbreviations: FRAX=Fracture Risk Assessment Tool; MOF=major osteoporotic fracture defined as fracture of the hip, spine (clinical), wrist, or humerus; U.K.=United Kingdom; U.S.=United States.

Risk Assessment Instruments

Appendix A Table 2. Risk Assessment Instruments for Identifying Osteoporosis or Predicting Fracture

Purpose/Risk Factor	ABONE	AMMEB	FRAX	FRC	FRISC	FRISK	Garvan FRC	MORES	MOST	MSCORE	NOF	ORAI	OSIRIS	OST	OSTA	QFracture	SCORE	SOF	SOFSURF	wHI
Osteoporosis Identification (OI) or Fracture Risk Prediction (FRP)	OI	OI	FRP	FRP	FRP	FRP	FRP	OI	OI	OI	OI	OI	OI	OI	OI	FRP	OI	OI	OI	FRP
Age	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sex	F	F	F, M	F, M	F	F	F, M	M	M	M	F	F	F	F	F	F, M	F	F	F	F
Race/ethnicity			X[*]	X											Asian only	X	X	X		X
Weight or BMI	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BMD			X[†]	X[†]	X[‡]	X[†]	X[†]													X[†]
Prior fragility fracture			X	X		X	X			X	X		X			X	X	X	X	X
Maternal/parental history of fracture			X	X							X					X		X		X
History of falls						X	X									X				
Current smoking			X	X							X					X			X	X
Alcohol consumption			X	X												X				
Secondary osteoporosis			X	X																
Corticosteroid use			X	X												X		X		X
Anticonvulsant use																X				
Antidepressant/benzodiazepine use																X				
Estrogen Related																				
Menopausal status					X															
Menopausal symptoms																X				
Hormone therapy	X											X	X			X	X	X		
Age at menarche		X																		
Years postmenopausal		X																		
Medical Conditions																				
Asthma																X				
Back pain					X															
Cancer																X				
Cardiovascular disease																X				
COPD								X		X						X				
Chronic liver disease																X				
Chronic renal disease																X				
Dementia					X											X		X		
Diabetes (type 1)																Type I				X
Endocrine problems																X				
Epilepsy																X				

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Purpose/Risk Factor	ABONE	AMMEB	FRAX	FRC	FRISC	FRISK	Garvan FRC	MORES	MOST	MSCORE	NOF	ORAI	OSIRIS	OST	OSTA	QFracture	SCORE	SOF	SOFSURF	wHI
Gastrectomy										X										
Gastrointestinal malabsorption																X				
Rheumatoid arthritis			X	X												X	X			
Parkinson's disease																X				
Other Factors																				
Global health																				X
Heart rate >80																		X		
Living in a care or nursing home																X				
Needs help getting up																		X		
On feet more than 4 hours per day																		X		
Quantitative ultrasound index									X											
Walk for exercise																		X		

* Separate risk calculators are available for U.S. Caucasian, Black, Hispanic, and Asian persons.

† Can be used with or without BMD either at femoral neck, lumbar spine, or both depending on instrument.

ǂ Must include BMD at lumbar spine.

Note: this table does not include the Fracture Risk Evaluation Model, which uses age, sex, and 38 clinical risk factors for women and 43 risk factors for men.

Abbreviations: ABONE=assessing age, body size, and estrogen use; AMMEB=Age, years after Menopause, age at Menarche; BMI=body mass index; BMD=bone mineral density; COPD=chronic obstructive pulmonary disease; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; FRISC=Fracture and Immobilization Score; FRISK=Fracture Risk Score; FRC=Fracture Risk Calculator; FRP=Fracture Risk Prediction; MORES=Multiple Outcomes of Raloxifene Trial; MOST=Male Osteoporosis Screening Tool; MSCORE=Male Simple Calculated Osteoporosis Risk Estimation; NOF=National Osteoporosis Foundation; OI=Osteoporosis Identification; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=Osteoporosis Index of Risk; OST=osteoporosis self-assessment tool; OSTA=Osteoporosis Self-assessment Tool for Asians; SCORE=Simple Calculated Osteoporosis Risk Estimate; SOF=Study of Osteoporotic Fractures; SOFSURF=Study of Osteoporotic Fractures Simple Useful Risk Factors; U.S.=United States; WHI=Women's Health Initiative.

Additional Information: Fracture Risk Assessment Tool

The most studied fracture risk assessment instrument is Fracture Risk Assessment Tool (FRAX), released in 2008 and developed by the University of Sheffield in the United Kingdom during the time the University hosted the World Health Organization's (WHO's) Collaborating Centre for Metabolic Bone Disease (1991 to 2010).[12, 71] FRAX predicts the 10-year probability of hip fracture or major osteoporotic fractures (MOFs) (hip, spine, wrist, shoulder) for persons ages 40 to 90 years using demographic and clinical factors alone or in combination with bone mineral density (BMD) measured at the femoral neck (FN).[89] Risks predicted by FRAX without BMD are not as accurate when compared with FRAX with the use of BMD; however, risks predicted by FRAX without BMD are similar to risks predicted by BMD alone.[323]

FRAX was derived from nine cohorts in Europe, the United States, Japan, and Canada and further validated in an additional 11 cohort studies.[12, 71] As of spring 2021, 73 different country-specific versions of FRAX were available that have been calibrated using country-specific fracture incidence and mortality data (which is considered a competing risk in the model).[78] For the United States, four different versions of FRAX are available that have been calibrated based on racial- and ethnic-specific fracture incidence data, including unique versions for non-Hispanic Caucasians, non-Hispanic Blacks, Hispanics, and non-Hispanic Asians.[69] We note that the group labels used to describe the race-specific FRAX calculators may not be consistent with current preferred terminology for various racial and ethnic groups.

As of 2016, FRAX was incorporated into 120 guidelines worldwide and added into dual-energy X-ray absorptiometry (DXA) software following regulatory approval by the U.S. Food and Drug Administration (FDA) and has been incorporated into clinical decision support tools within electronic health record systems.[89] The most commonly cited limitations of the FRAX instrument include use of binary exposure to glucocorticoids and alcohol use (yes/no vs. quantified dose exposure), lack of use of lumbar spine (LS) BMD or trabecular bone score, no information collected about history of falls, frailty, and lack of medical conditions such as diabetes that may portend an increased risk.[69, 84, 324] Falls and propensity to fall become increasingly important risk factors with advancing age. Further, FRAX has only been validated for use with FN BMD, and using FRAX in persons with low BMD at the LS but relatively normal BMD at the FN may underestimate fracture risk.[84] Because hip fracture incidence in the United States is lower in most non-White racial and ethnic groups, predicted fracture risk estimates for persons in these racial and ethnic groups will always be lower than for White persons of the same age, sex, weight, BMD, and clinical risks used in the FRAX model. See **Appendix A Contextual Question 2** for additional information about the use of race and ethnicity in FRAX.

USPSTF's Prior Recommendations Related to Use of FRAX

In the last two updates to its recommendations, the USPSTF has recommended BMD testing in all women age 65 years or older but only recommended BMD testing for women ages 50 to 64 years who are at increased risk of osteoporosis, as determined by a formal clinical risk assessment tool (topline recommendation).[1] In the clinical considerations, the USPSTF suggests that one approach in women younger than age 65 years is to screen individuals when their risk of a 10-year MOF is equivalent to that of a 65-year-old White woman with no other clinical risks.[1]

In the 2011 recommendation, this risk was 9.3 percent based on risk for a White women with BMI of 25 kg/m^2. In the 2018 recommendation, this risk was 8.4 percent based on the risk for a White woman of average height and weight in the United States, which was a BMI slightly higher than 25 kg/m^2.

The usefulness of the USPSTF's approach in younger women is unclear. Identifying persons with a T-score less than -2.5 is important because that is the population for whom trial evidence supports treatment. However, tools other than FRAX that were developed specially to identify osteoporosis are simpler and have higher diagnostic accuracy than FRAX.[139] Several studies have retrospectively applied the USPSTF FRAX criteria to a sample of women to evaluate accuracy for identifying osteoporosis;[141, 195, 220, 234, 236] these are included in **key question (KQ) 2c** of this update.

Although the use of FRAX does have limitations, the field has evolved toward trying to identify those at risk for fracture and not just those with osteoporosis because most fragility fractures occur in persons with T-scores greater than -2.5. Age is a large driver of fracture risk relative to the T-score in older populations, and because fracture risk has greater between-country variability than BMD does, some researchers argue that treatment decisions should probably be based on fracture risk and not BMD alone.[89] For example, in a 65-year-old female, the 10-year MOF risk of 20 percent (the treatment threshold in the United States) corresponds to a FN T-score of -4.6 in Venezuela but only -2.0 in Iceland.[89] For these reasons, some experts and organizations recommend fracture risk assessment as the initial screening approach for all ages, with subsequent BMD testing for persons at an intermediate or higher risk (see **Appendix A Table 4** in the subsequent section). Randomized, controlled trials (RCTs) are now available (see **KQ 1** of main report) that compare a screening strategy that uses FRAX risk calculation followed by BMD in selected patients who score above a certain risk threshold; however, no published studies have been designed to evaluate a treatment strategy based on FRAX, although some treatment trials may now report baseline characteristics related to fracture risk and provide results stratified by fracture risk.[89]

Pharmacologic Treatment of Osteoporosis

Appendix A Table 3. FDA-Approved Pharmacologic Agents for the Prevention or Treatment of Osteoporosis Included in this Review

Drug Generic (Brand Name[s])	FDA-Approved Indications Related to Osteoporosis	Dose and Route of Administration for Osteoporosis	Date First Approved for Osteoporosis
Alendronate (Binosto, Fosamax, Fosamax plus D and generics)	Binosto, Fosamax, Fosamax plus D • Treatment of osteoporosis in postmenopausal women • Treatment to increase bone mass in men with osteoporosis Fosamax only • Prevention of osteoporosis in postmenopausal women • Treatment of glucocorticoid-induced osteoporosis	Fosamax • Treatment: 10 mg daily or 70 mg (tablet or oral solution) once weekly • Prevention: 5 mg daily or 35 mg once weekly • Glucocorticoid-induced osteoporosis: 5 mg daily or 10 mg daily Fosamax plus D • 70-mg alendronate/2,800 or 5,600 IU vitamin D3 once weekly Binosto (no generics available) • 70-mg effervescent tablet once weekly	09/29/1995
Zoledronic acid (Reclast and generics)	• Treatment and prevention of postmenopausal osteoporosis • Treatment to increase bone mass in men with osteoporosis • Treatment and prevention of glucocorticoid-induced osteoporosis	• Infusion given intravenously over no less than 15 minutes • Treatment in women and men or treatment and prevention of glucocorticoid-induced osteoporosis: 5 mg once a year • Prevention: 5 mg once every 2 years	4/16/2007
Risedronate (Actonel, Actonel with calcium, Atelvia, and generics)	Actonel, Actonel with calcium • Treatment and prevention of postmenopausal osteoporosis Actonel only • Treatment to increase bone mass in men with osteoporosis • Treatment and prevention of glucocorticoid-induced osteoporosis Atelvia • Treatment of postmenopausal osteoporosis	Actonel • Prevention or treatment in women and men: 5 mg daily, 35 mg once a week • Prevention or treatment in women: 75 mg on 2 consecutive days each month, 150 mg once a month • Prevention or treatment of glucocorticoid-induced osteoporosis: 5 mg daily Actonel with calcium • One 35-mg tablet orally, taken once a week followed by one 1,250-mg calcium carbonate tablet (500-mg elemental calcium) orally, taken with food daily on each of the remaining 6 days of the week Atelvia • One 35-mg delayed-release tablet once a week	3/27/1998
Ibandronate (Boniva and generics)	Boniva • Treatment and prevention of postmenopausal osteoporosis	Boniva • One 150-mg tablet once monthly on the same day each month	5/16/2003

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Drug Generic (Brand Name[s])	FDA-Approved Indications Related to Osteoporosis	Dose and Route of Administration for Osteoporosis	Date First Approved for Osteoporosis
Denosumab (Prolia)	• Treatment of postmenopausal women with osteoporosis at high risk for fracture • Treatment to increase bone mass in men with osteoporosis at high risk for fracture • Treatment of glucocorticoid-induced osteoporosis in men and women at high risk for fracture • Treatment to increase bone mass in men at high risk for fracture receiving androgen deprivation therapy for nonmetastatic prostate cancer • Treatment to increase bone mass in women at high risk for fracture receiving adjuvant aromatase inhibitor therapy for breast cancer	60-mg subcutaneous injection every 6 months	6/2/2010

Abbreviation: FDA=Food and Drug Administration.

Recommendations and Guidelines for Screening From Professional Organizations

Recommendations for screening developed by various organizations and specialty societies share commonalities but also have significant differences (**Appendix A Table 4**). In general, most guidelines focus on postmenopausal women and use the WHO standard for defining osteoporosis. One important difference among guidelines is that some recommend screening for fracture risk via fracture risk assessment tools such as FRAX, while some recommend screening for osteoporosis via BMD measured through DXA. Current guidelines from several organizations recommend a combination of fracture risk assessment and DXA screening.

The most recent guideline recommending a combination approach is the 2023 Canadian Task Force on Preventive Health Care (CTFPHC) recommendation for screening to prevent primary fragility fractures.[90] The CTFPHC recommends screening women age 65 years or older with the Canadian FRAX tool to facilitate shared decision making about pharmacotherapy. If pharmacotherapy is considered, it then recommends ordering DXA testing in order to facilitate re-estimation of fracture risk with a BMD input. The CTFPHC recommends against screening in men age 40 or older and in women younger than age 65 years.

Other examples include the 2020 American Association of Clinical Endocrinologists (AACE) Guideline, which recommends evaluating all women age 50 years or older for fracture risk and consider BMD measurement based on clinical fracture risk profile.[91] The AACE guidelines state that osteoporosis should be diagnosed based on a T-score of -2.5 or lower in the LS, FN, total hip (TH), and/or distal third of the radius in the absence of a prevalent fracture, or in patients with a T-score between -1.0 and -2.5 and increased fracture risk using FRAX country-specific thresholds. Similarly, the 2017 National Institutes for Health and Care Excellence (NICE) (United Kingdom) recommended fracture risk–based screening for all women age 65 years or older and all men age 75 years or older (i.e., using FRAX or the QFracture), followed by BMD screening if indicated.[325] The NICE guidelines also recommended screening in women younger than age 65 years and men younger than age 75 years in the presence of fracture risk factors. In 2019, the Endocrine Society updated its guidelines for postmenopausal women and noted that screening should be determined using country-specific clinical fracture risk assessment tools (e.g., FRAX) and patient preference, though the guidelines for women are ambiguous with respect to whether BMD should be used to determine fracture risk.[326] The Osteoporosis Canada 2023 Guideline Update Group recommends assessing all men and postmenopausal women age 50 or older with the Canadian FRAX or the Canadian Association of Radiologists and Osteoporosis Canada 10-year fracture risk assessment tool, CAROC. Then, based on age, prior fractures, and risk profile criteria, obtain BMD and calculate 10-year fracture risk with BMD.[327]

Other guidelines focus on osteoporosis screening via DXA measurement of BMD. The International Society for Clinical Densitometry (ISCD) 2019 guidelines recommend central skeletal site BMD screening in all women age 65 years or older and all men age 70 years or older.[11] It also recommends BMD screening for postmenopausal women younger than age 65 years and men younger than age 70 years who have risk factors for osteoporosis.[11] The American College of Obstetricians and Gynecologists recommends BMD screening with DXA beginning at age 65 years in all women and selective screening with BMD in women younger than age 65 years who have an elevated risk of osteoporosis based on a formal clinical risk assessment tool, with repeat screening no sooner than 2 years after initial screening for those with a BMD near a treatment threshold at the time of initial screening.[328] The National Osteoporosis Foundation's (NOF's), now the Bone Health and Osteoporosis Foundation, most recent clinical guideline (2022) recommends BMD evaluation in all women age 65 years or older and all men age 70 years or older. It also recommends BMD testing in postmenopausal women, women in menopausal transition, men ages 50 to 69 years with clinical risk factors, and adults with fractures at age 50 years or older.[84]

An outlier in recommending against screening, the United Kingdom National Screening Committee, reviewed three recent RCTs[120, 124, 126] on screening for fracture risk (SCOOP, ROSE, SOS) and did not find the evidence compelling enough to recommend a screening program and continues to favor case finding. Other guidelines remain mostly unchanged from the last time the USPSTF reviewed this topic (2018), including those from the American College of Preventive Medicine and the American College of Radiology.

Appendix A Table 4. Recommendations for Fracture or Osteoporosis Screening by Organization

Organization, Year	Population	Recommendations
American Association of Clinical Endocrinology, 2020[91]	Postmenopausal women	Screening • Evaluate all postmenopausal women age 50 years or older for osteoporosis risk • Include a detailed history, physical examination, and clinical fracture risk assessment with FRAX or other risk assessment tool in the initial evaluation for osteoporosis • Consider BMD testing based on clinical fracture risk profile Treatment • When BMD is measured, use DXA measurement (spine and hip, 1/3 radius if indicated) • Osteoporosis should be diagnosed based on presence of fragility fractures even in the absence of metabolic bone disorders or a normal T-score or on a T-score of -2.5 or lower in the lumbar spine, femoral neck, total hip, or 1/3 radius even in the absence of a prevalent fracture, and the diagnosis persists even if subsequent measures improve • Osteoporosis may also be diagnosed in patients with a T-score between -1.0 and -2.5 and increased fracture risk using FRAX (Fracture Risk Assessment Tool) country-specific thresholds • Approved agents with efficacy to reduce hip, nonvertebral, and spine fractures, including alendronate, denosumab, risedronate, and zoledronate, are appropriate as initial therapy for most osteoporotic patients with high fracture risk • Abaloparatide, denosumab, romosozumab, teriparatide, and zoledronate should be considered for patients unable to use oral therapy and as initial therapy for patients at very high fracture risk • Ibandronate and raloxifene may be appropriate initial therapy in some cases for patients requiring drugs with spine-specific efficacy
American Academy of Family Physicians, 2021[329]	Postmenopausal women Men	Same recommendations as the 2018 USPSTF recommendations: • Women age 65 years or older (B) • In younger women whose fracture risk is equal to or greater than that of a 65-year-old White woman who has no additional risk factors (B) • Insufficient evidence to assess the balance of benefits and harms of screening for osteoporosis in men
American College of Obsetrics and Gynecology, 2021[328]	Women	Screening by DXA: • Postmenopausal patients age 65 years or older • Younger postmenopausal patients if they are at elevated risk of osteoporosis based on a formal clinical risk assessment tool • Repeat screening no sooner than 2 years after initial screening for postmenopausal patients with BMD near treatment thresholds at the time of initial screening

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Organization, Year	Population	Recommendations
American College of Physicians, 2023[318]	Postmenopausal women Men	Screening: No specific guideline related to screening Treatment • Bisphosphonates for initial pharmacologic treatment in postmenopausal females (high certainty) and males (low certainty) diagnosed with primary osteoporosis • Denosumab for second-line treatment in postmenopausal females (moderate certainty) and males (low certainty) diagnosed with primary osteoporosis • Romosozumab or rPTH followed by bisphosphonate for females with very high risk of fracture (conditional recommendation) • Individualized approach regarding whether to start bisphosphonate treatment in females older than age 65 years with low bone mass (osteopenia) (low certainty)
American College of Preventive Medicine, 2009[330]	Women age 65 years or older Men age 70 years or older	Screening: Recommend DXA • All women age 65 years or older and men age 70 years or older and not more frequently than every 2 years • Younger postmenopausal women and men ages 50 to 69 years should undergo screening if they have at least one major or two minor risk factors for osteoporosis • Osteoporosis risk assessment tools that estimate absolute fracture risk can be useful supplements to BMD testing, improving the sensitivity and specificity of either approach (BMD or risk assessment) alone; risk assessment can also be used if BMD testing is not readily available or feasible
American College of Radiology, 2017[63]	Asymptomatic BMD screening for individuals with established or clinically suspected low BMD	• All women age 65 years or older and men age 70 years or older (asymptomatic screening) • Women younger than age 65 years who have additional risk for osteoporosis, based on medical history and other findings. Additional risk factors for osteoporosis include: a. Estrogen deficiency b. A history of maternal hip fracture that occurred after the age of 50 years c. Low body mass (<127 lb or 57.6 kg) d. History of amenorrhea (>1 year before age 42 years) • Women younger than age 65 years or men younger than age 70 years who have additional risk factors, including: a. Current use of cigarettes b. Loss of height, thoracic kyphosis
The Bone Health and Osteoporosis Foundation, 2022[84]	Men age 50 years or older and postmenopausal women	Screening with DXA • Women age 65 years or older and men age 70 years or older • Postmenopausal women and men ages 50 to 69 years with clinical risk factors • Adults with fractures at age 50 years or older Treatment • Consider for postmenopausal women and men age 50 years or older with T-scores -2.5 or worse • Consider for postmenopausal women and men age 50 years or older with T-scores between -1.0 and -2.5 and a 10-year FRAX probability of major osteoporosis-related fracture ≥20%
Canadian Task Force on Preventive Health Care, 2023[90]	Men and women age 40 or older	Screening with DXA • Screen women age 65 years or older with Canadian version of FRAX to facilitate shared decision making about pharmacotherapy; if pharmacotherapy is considered, obtain DXA and re-estimate FRAX with BMD input • Recommend against screening in women younger than age 65 years and men

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Organization, Year	Population	Recommendations
Endocrine Society, 2012[331] 2019[326] 2020[332]	Screening for higher-risk men Screening and treatment for postmenopausal women	Screening in men: Recommend BMD screening by central DXA in: • Men age 70 years or older • Men ages 50 to 69 years with risk factors (e.g., low body weight, prior fracture as an adult, smoking) Screening and treatment in postmenopausal women • The risk of future fractures in postmenopausal women should be determined using country-specific assessment tools to guide decision making. The guidelines are ambiguous with respect to whether BMD should be evaluated to determine fracture risk. • Patient preferences should be incorporated into treatment planning. • Nutritional and lifestyle interventions and fall prevention should accompany all pharmacologic regimens to reduce fracture risk. • Multiple pharmacologic therapies are capable of reducing fracture rates in postmenopausal women at risk with acceptable risk-benefit and safety profiles.
International Society of Clinical Densitometry, 2019[11]	Men and postmenopausal women	BMD screening • Women age 65 years or older • Postmenopausal women younger than age 65 years with risk factors for low bone mass • Women during the menopausal transition with clinical risk factors for fracture, such as low body weight, prior fracture, or high-risk medication use • Men age 70 years or older • Men younger than age 70 years with clinical risk factors for low bone mass Diagnosis • Recommends the WHO international reference standard for osteoporosis diagnosis: a T-score of -2.5 or less at the femoral neck. The reference standard from which the T-score is calculated is the female, White, ages 20 to 29 years, NHANES III database. • Osteoporosis may be diagnosed in postmenopausal women and in men age 50 years or older if the T-score of the lumbar spine, total hip, or femoral neck is -2.5 or less. In certain circumstances, the 1/3 radius may be used.
National Institute for Health and Care Excellence (U.K.), 2017[325]	Persons presenting in any healthcare setting	Screening • In all women age 65 years or older and all men age 75 years or older • In women younger than age 65 years and men younger than age 75 years in the presence of risk factors, for example: a. History of falls b. Family history of hip fracture c. Low BMI (<18.5 kg/m^2) d. Smoking e. Alcohol intake more than 14 units per week (women) or more than 21 units per week (men) • Do not routinely assess fracture risk in people younger than age 50 years unless they have major risk factors (e.g., current or frequent recent use of oral or systemic glucocorticoids, untreated premature menopause, or previous fragility fracture), because they are unlikely to be at high risk. • Estimate absolute risk when screening. Consider measuring BMD with DXA in people whose absolute fracture risk (via FRAX or QFracture) is in the region of an intervention threshold for a proposed treatment and recalculate FRAX with BMD value.

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Organization, Year	Population	Recommendations
National Osteoporosis Guideline Group (U.K.), 2022[333]	Postmenopausal women; men age 50 years or older	• FRAX assessment should be done in any postmenopausal woman or in any man age 50 years or older with a clinical risk factor for fragility fracture to guide BMD measurement.
North American Menopause Society, 2021[334]	Postmenopausal women	Screening with DXA • Postmenopausal women with history of fracture after menopause • All women with medical causes of bone loss (e.g., steroid use, hyperparathyroidism), regardless of age • Consider for postmenopausal women younger than age 65 years with specified risk factors (see below) a. Discontinued estrogen with additional risk factors for fracture b. Thinness (body weight <127 lb [57.7 kg] or BMI <21 kg/m^2) c. History of hip fracture in a parent c. Current smoker d. Alcohol intake of more than two units per day (one unit is 12 oz of beer, 4 oz of wine, or 1 oz of liquor) e. Long-term use of medications associated with bone loss Treatment • A variety of nonpharmacologic treatments reviewed such as nutrition, mineral and vitamin use, exercise, fall prevention, and smoking cessation; routine use of calcium and vitamin D supplements is not recommended except when daily targets are not achieved from dietary sources • All postmenopausal women who experienced vertebral or hip fracture • All postmenopausal women who have BMD values consistent with osteoporosis (i.e., T-scores equal to or worse than -2.5) at the lumbar spine, femoral neck, or total hip region • All postmenopausal women who have T-scores from -1.0 to -2.5 and any of the following: increased fracture risk based on country-specific FRAX threshold; history of proximal humerus, pelvis, or distal forearm fracture; or history of multiple fractures at other sites excluding face, feet, and hands • Several pharmacologic options are available for osteoporosis therapy, including bisphosphonates, the selective estrogen receptor modulator (SERM; also known as estrogen agonist/antagonist) raloxifene, PTH, estrogens, and calcitonin. Bisphosphonates are the first-line drugs for treating postmenopausal women with osteoporosis.

Organization, Year	Population	Recommendations

Osteoporosis Canada 2023 Guideline Update Group, 2023[327]	Men and postmenopausal women age 50 years or older	Screening
		• Assess all men and postmenopausal women age 50 years or older for fracture risk factors using the Canada-specific FRAX tool, or alternatively the CAROC.
		• Obtain BMD and calculate 10-year fracture risk with BMD using FRAX or CAROC for all postmenopausal women and men ages 50 to 64 years with previous fracture or 2 or more risk factors, ages 65 to 69 years with 1 risk factor, or age 70 years or older with no risk factors.
		• Reassess BMD and fracture risk in 3 years for patients who initiated pharmacotherapy. Reassess BMD and fracture risk for those not taking pharmacotherapy based on 10-year fracture risk:
		  • <10%, reassess in 5 to 10 years
		  • 10% to 15%, reassess in 5 years
		  • ≥15%, reassess in 3 years
		
		Treatment
		• Recommend pharmacotherapy to postmenopausal women or men age 50 years or older with previous hip or spine fracture or 2 or more fracture events, those with high absolute risk (≥20% probability for major osteoporotic fracture over 10 years), and those with T-score ≤-2.5% and age 70 years or older.
		• Suggest pharmacologic therapy to patients with 15% to 19.9% probability for major osteoporotic fracture over 10 years and those with T-score ≤-2.5 and age less than 70 years.
		• A variety of nonpharmacologic treatments were reviewed, including nutrition and exercise for fall and fracture prevention.

United Kingdom National Screening Committee, 2019[335]	Postmenopausal women	After a review of the SALT-SOS, ROSE, and SCOOP trials, a systematic screening program for osteoporosis is not recommended in the United Kingdom. However, hip fracture is an important outcome, and future work should focus attention on this area.

Appendix A. Additional Background and Contextual Questions 2, 3 and 4

Organization, Year	Population	Recommendations
World Health Organization, 2007[5, 12]	Men and women ages 40 to 90 years	Screening • Population-based (i.e., public health) prevention programs are appropriate for all Member States and should include attention to nutritional factors, particularly calcium and vitamin D. Cigarette smoking, prevention of excessive alcohol consumption, and the avoidance of immobility are also recommended as public health measures. • In Member States without access to densitometry, case-finding strategies can be pursued with use of clinical risk factors alone. The performance characteristics of the FRAX model are at least as good as those provided by peripheral assessment of BMD. • In Member States where BMD is universally recommended (e.g., at age 65 years or older in North America), the stratification of risk can be improved by considering clinical risk factors in conjunction with BMD. This is particularly valuable in the context of younger individuals for hip fracture prediction. • In Member States with limited access to DXA, clinical risk factors can be used to stratify target populations to those at very high risk in whom a BMD test would not alter their risk category, those with very low risk in whom a BMD would not alter their risk category, and those at intermediate risk where a BMD test would be helpful for the characterization of fracture probability. Treatment • The validation of BMD measurements and the increase in epidemiological information permit diagnostic criteria for osteoporosis to be more precisely defined than previously. The international reference standard for the description of osteoporosis in postmenopausal women and in men age 50 years or older is a femoral neck BMD of 2.5 SD or more below the young female adult mean, using normative data from the NHANES reference database on Caucasian women ages 20 to 29 years. • Although the reference standard for the description of osteoporosis is BMD at the femoral neck, other central sites (e.g., lumbar spine, total hip) can be used for diagnosis in clinical practice. • T-scores should be reserved for diagnostic use in postmenopausal women and men age 50 years or older. With other technologies, and other populations, measurement values should be expressed as Z-scores, units of measurement, or preferably in units of fracture risk.

Abbreviations: BMD=bone mineral density; BMI=body mass index; CAROC=Canadian Association of Radiologists and Osteoporosis Canada 10-year fracture risk assessment tool; DXA=dual-energy X-ray absorptiometry; FDA=U.S. Food and Drug Administration; FRAX=Fracture Risk Assessment Tool; IU/day=international unit per day; NHANES=National Health and Nutrition Examination Survey; PTH=parathyroid hormone: Promising Developments in Osteoporosis Treatment; ROSE=Risk-Stratified Osteoporosis Strategy Evaluation study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study; SD=standard deviation; SERM=selective estrogen receptor modulator; UKNSC=United Kingdom National Screening Committee; USPSTF=U.S. Preventive Services Task Force.

Contextual Questions 2–4

CQ 2. How Do Various Risk Assessment Tools Use Race and Ethnicity in Osteoporosis or Fracture Risk Calculations?

Summary

Although several fracture risk estimators have been developed, only two that are commonly used in clinical practice incorporate race or ethnicity: the FRAX, calibrated for use internationally, and QFracture, developed in the United Kingdom.[77] Several other fracture risk assessment models have been developed—the Women's Health Initiative (WHI) model,[187] the Established Populations for the Epidemiologic Study of the Elderly (EPESE) model,[336] the American Bone Health Fracture Risk Calculator (FRC),[73, 337] and the Study of Osteoporotic Fracture (SOF)-based screening tool[52]—but these models are not commonly used. The only osteoporosis risk assessment tool that incorporates race or ethnicity into the assessment is the Simple Calculated Osteoporosis Risk Estimation (SCORE) tool.[76]

Fracture Risk Assessments

FRAX

FRAX, the most widely used fracture risk assessment tool, is a tool that was developed for use internationally with country-specific estimates derived through calibrating fracture risk to country-specific fracture incidence and mortality data.[71, 338, 339] FRAX was originally calibrated to the U.S. White population using population-based data from Olmsted County, MN, prior to the availability of the Nationwide Inpatient Sample (NIS), a large U.S. hospital discharge database; however, the data from Olmsted County have been subsequently shown to be similar to NIS data.[12, 71, 340-342] For Black, Asian, and Hispanic populations, race-specific FRAX calculators were created by applying the ratio of race- and sex-specific hip fracture incidence rates (0.43 and 0.53 for Black women and men, 0.53 and 0.58 for Hispanic women and men, and 0.50 and 0.64 for Asian women and men) derived from multiple epidemiologic studies to the calculators developed for the U.S. White population.[80]

Because hip fracture incidence in the United States is lower in these racial and ethnic groups, predicted risk estimates for persons in these racial and ethnic groups will always be lower than for White persons of the same age, sex, weight, BMD, and clinical risks used in the FRAX model (**see Appendix F.2, CQ5 for further details**). In the wake of recent attention to racial bias in clinical algorithms, some have raised questions regarding the validity of race-specific FRAX calculators. The relationships between age and clinical risk factors (including BMD) with fracture incidence are the same across all racial groups in FRAX;[80] however, the predicted risk for persons of different races or ethnicities occurs because of calibration of the race-specific calculators, which use race-specific hip fracture incidence data. Of note, although FRAX is available for countries with multiracial populations such as Canada and the United Kingdom, the only countries with race and ethnicity-specific FRAX models are Singapore, South Africa, and the United States.[343]

Whether these differences in absolute fracture risk reflect bias in the FRAX prediction model or whether these differences simply reflect the end result of using a model calibrated to race-specific incidence data is a matter of debate. Some experts state the lower absolute risks produced by FRAX for Black, Asian, and Hispanic populations simply reflect the underlying epidemiology of fractures in those populations.[344] Other experts have acknowledged the limitations of race in FRAX, where it likely serves as a proxy for environmental factors; does not account for multiracial people; and minimizes the diversity within racial groups, which, for BMD, may be even greater than the diversity between racial groups.[324] These acknowledged limitations are all consistent with the USPSTF's current perspective on race as a social, not a biologic, construct.[345] However, experts also note some biological differences within and between populations that may explain some of the observed variability in fracture incidence.[343] This is discussed further in **Contextual Question 3** below.

Because treatment recommendations that incorporate predicted fracture risk in the United States are based on fixed predicted risk thresholds (e.g., FRAX ≥20% MOF risk or ≥3% hip fracture risk) that are not specific to race, Black, indigenous, and persons of color populations may be less likely to be identified as high risk and offered treatment compared with White persons of the same age, BMD, and clinical risk profile.[324, 344] Similarly, other conditions that increase fracture risk and that disproportionately affect persons of color (e.g., diabetes) may result in biased underestimates of risk.[324, 344] This in turn may lead to less treatment for at-risk individuals belonging to "low-risk" racial groups if these underestimates result in misclassification below the fixed risk thresholds used to recommend treatment. Studies evaluating the sequelae of fracture have found greater postfracture morbidity and mortality in Black women than in White women.[346] For these reasons, many recommend avoiding strict application of treatment thresholds at the individual level to account for additional risks that are not taken into account by the FRAX model.[343, 344]

Other acknowledged limitations of the current versions of race-specific U.S. FRAX models are that they do not use the most currently available data for race-specific fracture incidence (estimates are from cohort studies in the 1980s and 1990s) and the mortality data used as a competing risk are from 2004 and have not been updated, which would perpetuate underestimates of fracture risk that don't reflect recent gains in life expectancy across racial and ethnic groups.[324] Further, because FRAX incorporates age-, race-, and sex-specific mortality data as a competing risk,[340] residual differences in life expectancy by race and ethnicity may reflect the impact of structural racism on health and this may result in continued underestimates of fracture risk in non-White populations.[324]

QFracture

In contrast to FRAX, QFracture uses what its developers define as ethnicity as a variable in its sex-specific equations estimating fracture risk.[147] The ethnic groups used in QFracture differ from those in FRAX, suggesting that conceptualizations of race and ethnicity and their relevance to disease risk differ between societies. The ethnicities used in QFracture are White, Indian, Pakistani, Bangladeshi, other Asian, Black African, Black Caribbean, Chinese, and other including "mixed." Notably, Hispanics are not included, likely because Hispanic is an ethnic group created in the United States and is not recognized as an ethnic group elsewhere.[347] In the 2012 version of QFracture, White women (and those with unknown racial category) have the highest predicted risk of fracture, while Black Caribbean (hazard ratio [HR], 0.23 relative to White women), Bangladeshi (HR, 0.44), Pakistani (HR, 0.46), and Black African (HR, 0.48) women have the lowest predicted fracture rates.[147] Among men, White persons (and those with unknown racial category) have the highest predicted risk of fracture, while Bangladeshi (HR, 0.29 relative to White men), Black Caribbean (HR, 0.38), Black African (HR, 0.52), and persons from “other” ethnic groups (HR, 0.57) have the lowest predicted fracture rates.[147] QFracture has been updated with 2016 data not described in the literature, but a review of the tool suggests similar associations between ethnicity and fracture persist.[348] Nevertheless, Black persons are less than 3.5 percent of the U.K. population and are not a representative sample of Caribbean and African persons.[349-351] Little data exist regarding the distribution of nationalities among the U.K. Black population, but data suggest that Black Africans primarily comprise Nigerians, Ghanaians, and Somalis,[351] and Black Caribbeans have majority Jamaican ancestry.[352]

Other Models

Other models that include race in fracture risk estimation include the WHI,[187] the EPESE models,[336] the American Bone Health FRC,[73, 337] and the SOF-based screening tool.[52] In the WHI and the EPSE models, race is dichotomized as White vs. non-White and is used as one of eight and 11 fracture risk factors, respectively, in multivariable models predicting fracture risk. The WHI model, which was developed in the United States, includes Hispanic persons in the non-White group, so White is presumed to mean non-Hispanic White (NHW) in that model. Hispanic ethnicity was not discussed in the EPESE model. The coefficients in the multivariable analyses used to create both models were then translated into a point system for which White persons receive more points, indicating greater fracture risk. The SOF tool includes 12 risk factors that were found to be associated with hip fracture in multivariate models. The presence of each factor is assigned 1 point; three additional factors can result in a point being subtracted from the overall score, and African American race is one of those three factors. We did not identify any studies evaluating the EPESE or SOF tool that met eligibility criteria for inclusion in **KQ 2** of this update.

Osteoporosis Risk Assessments

The SCORE tool was developed initially to identify a patient’s risk of osteoporosis (i.e., T-score <-2.5) as opposed to fracture risk.[76] Race is dichotomized in SCORE; however, the categories used by SCORE are Black vs. non-Black. Race is one of six factors used in this additive model.[76] The SCORE equation was developed from multivariable linear regressions estimating BMD. The coefficients from the model were then translated into a point system representing osteoporosis risk. Individuals with a SCORE value greater than 6 are at moderate to high risk of osteoporosis. Of note, identifying as non-Black adds 5 points to SCORE, whereas identifying as Black adds 0 points.

CQ 3. What Is the Incidence of Fractures Among Persons of Different Races and Ethnicities in the United States in the Last 10 to 15 Years, and What Factors Might Explain Differences in Incidence Among Different Races and Ethnicities?

Summary

The few studies documenting differences in fracture incidence have found that non-Hispanic Black (NHB) and Asian Americans have lower rates of fracture compared with NHW and Hispanic Americans. Racial differences in fracture incidence have been attributed to differences in bone quality, bone morphology, and fall frequency. Studies have reported that NHB Americans have higher BMD than other racial groups, and Asian Americans (who have lower fracture risk) have been primarily found to have lower BMD than NHW Americans. However, studies comparing NHW Americans with NHB and Asian U.S. subpopulations (i.e., Black immigrants and Asian ethnic subgroups) and studies comparing White and Black persons outside the United States have mixed findings. Studies evaluating racial difference in bone quality (architecture, hip axis length) and fall frequency were also inconclusive.

Fracture Incidence

U.S. studies evaluating fracture incidence among persons of different races and ethnicities have been primarily conducted with White persons as the comparator group; therefore, our discussion here reflects that approach. To our knowledge, only three studies using data from the last 15 years have been conducted to evaluate racial differences in fracture incidence. Two were clinic-based cohort studies[57, 58] and one used administrative data.[59] Liu et al used 2010 Kaiser Permanente Northern California data to calculate age-adjusted hip fracture incidence rates (per 100,000) that were highest among NHW men (137) and lowest among Asian men (45), with Hispanic (98) and NHB (80) men in between (see **Appendix A Table 5**).[57] A study using 2012 Kaiser data found White women had the highest age-adjusted incidence of hip fracture (288 per 100,000), followed by Hispanic (198), Asian (148), and NHB women (87).[58] Authors also found that Asian women had the highest rate of femoral diaphyseal fractures (27 per 100,000), followed by NHB (10), Hispanic (6), and NHW (5) women. Finally, the incidences of osteoporotic fracture (per 10,000) among Medicare fee-for-service 2016 beneficiaries were 389, 381.9, 271.8, 259.6, and 166.8 for Native American, NHW, Hispanic, Asian, and Black persons, respectively.[59] After adjusting for age and sex, the order of fracture incidence by race remained unchanged.

Appendix A Table 5. Fracture Incidence (per 100,000) by Study Year, Fracture Type, Race and Ethnicity, and Sex[57-59]

	2010*	2012*	2012*	2016
Fracture type	***Hip***	***Hip***	***Femoral diaphysis***	***Several†***
Sex	***Male***	***Female***	***Female***	***Male and Female***
Asian	45	148	27	2,596
Hispanic	98	198	6	2,718
NHB	80	87	10	1,668
NHW	137	288	5	3,819
North American Native				3,890

* Age-adjusted.

† Hip, distal femur shaft/distal femur, pelvis/sacrum, tibia/fibula, radius/ulna, clavicle, spine, rib.

Abbreviations: NHB=non-Hispanic Black; NHW=non-Hispanic White.

Racial or ethnic differences in fracture incidence have primarily been attributed to differences in BMD, bone microarchitecture, hip geometry, and fall frequency, each of which we discuss in the following section.

Bone Mineral Density

Studies of men and women have consistently demonstrated that NHB Americans have greater age-adjusted BMD than NHW[353-358] and Hispanic Americans.[353, 355, 356, 359] Differences in body size (height, weight, or both) explain some differences but residual differences after adjusting for body size are unexplained.

Studies comparing White and Black Americans without consideration of Hispanic ethnicity have shown Black Americans to have greater age-adjusted BMD than White Americans.[358, 360-364] A study of older adolescent American girls found Black Americans to have greater BMD than White and Asian Americans.[365] Black American men and women have also been shown to have slower rates of BMD decline with age than White Americans.[360, 366]

Comparisons between NHW, Hispanic, and Asian Americans have had varied findings. Studies of BMD in Hispanic and NHW Americans have found mixed results, including lower,[367] greater,[353] and similar[354, 356, 368] BMD in Hispanic Americans compared with NHW Americans. One study also found faster rates of bone density decline in Hispanic Americans compared with NHB and NHW Americans.[353] Asian Americans have primarily been found to have lower BMD than White Americans, although this difference has been explained by differences in body size.[60, 354, 356]

Racial Differences in BMD Among U.S. Subpopulations, U.S. Immigrant, and Non-U.S. Populations

Although NHB persons have greater BMD than other U.S. racial groups, little is known about why this racial difference exists, whether racial differences in BMD exist outside the United States, or whether Black immigrants differ from Black American-born persons. To our knowledge, only two studies have evaluated BMD in Black African immigrants residing in the United States. Gong and colleagues studied 55 male and 88 premenopausal female immigrants from South Sudan.[369] The authors found that the South Sudanese immigrants had lower lumbar BMD but similar hip and total body BMD compared with White and Black American normative values.[369] Melton et al found that compared with White American women, Somali-born women in the United States had a similar BMD at the LS but a greater BMD at the FN.[370]

Studies evaluating BMD in Black persons living outside the United States are few, often small in size, and have mixed findings. Demeke and colleagues found that Somali immigrant women (N=67) living in Sweden had lower BMD (LS and left and right hip) than a Black American reference group and lower LS BMD (but similar left and right hip) than a White American reference group.[371] A comparison between Black Gambian women living in Gambia and White British women living in the United Kingdom age 45 years or older (N=586) found that after adjusting for weight, age, and height, Gambian-born women had lower BMD at the LS.[372] In a study comparing Black Gambian-born U.K. immigrants and White U.K. residents (N=39), Black Gambian-born men had greater BMD at the FN alone but similar BMD at the LS, hip trochanter, radius, and whole body compared with White men.[373] In this same study, Gambian-born women and British women had similar BMD.[373] A study of South African Black and White women (N=294) found similar distal radius and lumbar BMD but greater femoral BMD among Black women.[374]

Likewise, few studies have been conducted evaluating Hispanic subgroups to determine whether patterns seen in the larger Hispanic population are present in Hispanic subpopulations (e.g., Mexican Americans, Puerto Rican Americans). Of note, studies evaluating differences in BMD between Hispanic and White persons in other countries could not be conducted because Hispanics are only defined as a population in the United States. Studies using National Health and Nutrition Examination Survey (NHANES) data have shown that Mexican American men and women had a higher prevalence of osteoporosis[375] and lower lumbar BMD[355] than NHW and NHB men and women. A study by Noel et al—using Boston Puerto Rican Osteoporosis Study and the NHANES 2005–2010 data—found that Puerto Rican men had a higher prevalence of osteoporosis than NHW and NHB men and a similar prevalence of osteoporosis to Mexican American men.[359] Puerto Rican women were found to have similar rates of osteoporosis as NHW and Mexican women but higher rates than NHB women.

Studies of U.S. Asian subpopulations have generally shown lower or similar BMD when compared with White women. For example, a study of FN BMD among older Asian Americans of Filipino, Chinese, or Japanese descent and White American women found that the Asian women had similar BMD, which was lower than those of NHW women.[60] The difference between Asian American and NHW women decreased when BMD was adjusted for height.[60] In a study of premenopausal women ages 42 to 52 years, Finkelstein and colleagues reported that Chinese and Japanese Americans had similar lumbar, spine, and FN BMD as White Americans, when adjusted for age, age at menarche, weight, years of oral contraceptive use, physical activity, number of prior pregnancies, educational level, total calcium intake, cigarette smoking, and alcohol intake, but lower than Black Americans (though this difference was not significant for Chinese American women at the LS).[376] When the analysis was limited to women weighing less than 70 kg, Chinese and Japanese American women had similar LS BMD as Black Americans and slightly greater (though not significantly so) than White Americans. In the subset of women weighing less than 70 kg, Japanese, Chinese, and White American women had similar FN BMD, which was lower than those of Black Americans (though this difference was not significant for Chinese American women). Other studies of Chinese American women have found lower BMD at LS, TH, and FN than White American women.[61, 62] However, a study of Filipina, Hispanic, and NHW American women found that Filipinas had higher total body (but similar hip and LS) BMD than the other two groups.[368]

Bone Microarchitecture

Bone microarchitecture includes cortical and trabecular volumetric BMD, cortical and trabecular area, cortical and trabecular thickness, cortical porosity, cortical perimeter, trabecular separation, and trabecular number. Trabecular bone score is correlated with trabecular microarchitecture.[377] Studies evaluating racial and ethnic differences in bone microarchitecture associations are few and findings were sometimes mixed. Compared with White Americans, Black Americans have microarchitecture favoring reduced[362, 365, 378, 379] and increased fracture risk.[355] A study by Jain and colleagues found that among women younger than age 60 years, White women had higher trabecular bone scores than Black women but similar scores among those age 60 years or older.[380] Black older adolescent girls have also been shown to have better bone microarchitecture than their Asian counterparts.[365] Studies comparing Asian and White American women found Asian women have greater cortical[381, 382] and trabecular[382] thickness than White women, both of which are associated with lower fracture risk.

Hip Geometry

Studies evaluating racial differences in hip geometry have varied findings. Differences in hip axis length have been posited as explaining racial differences in fracture: shorter hip axis length is associated with lower risk of fracture.[383] Some studies on hip axis length have found Black Americans to have shorter hip axis lengths than White Americans;[379, 384, 385] others have found that NHB Americans have hip axis lengths similar to NHW[358, 386] and Mexican Americans.[386] In one study, Asian Americans were found to have a shorter hip axis length[384] than White Americans, but in a different study, Japanese Americans had a similar hip axis length as White Americans after adjusting for height.[387]

Fall Frequency

A possible explanation for racial and ethnic differences in fracture incidence is differences in fall frequency among older adults. In three studies, White adults reported a greater number of falls compared with Black adults, but in two of these studies, differences in fall frequency were not medically or statistically meaningful.[388] Shumway-Cook and colleagues examined falls using Medicare Current Beneficiary Survey data and found that NHW adults were more likely (odds ratio [OR], 1.40 [95% confidence interval (CI),[260] 1.20 to 1.63]) to *report* at least one fall in the prior year than those who identified with other racial groups.[388] However, there was no significant difference in medically injurious falls (falls for which participants sought medical assistance) between NHW adults and the non-White racial group when the analysis was limited to those who had fallen.[388] In contrast, a study of Black and White older adults in the Boston area found that White participants were more likely (risk ratio [RR], 1.77 [95% CI, 1.14 to 2.74]) than Black participants to experience injurious falls.[389] However, not all studies indicated racial differences in fall frequency. Finally, a study by Faulkner and colleagues found that although White women had numerically higher fall rates, these rates were not significantly different than Black women's rates (RR, 1.30 [95% CI, 0.93 to 1.83]).[390]

CQ 4. What Are the Differences in Rates of Screening or Treatment Initiation Among Persons of Different Races and Ethnicities, and What Might Explain These Differences?

Summary

Racial disparities in screening and treatment were found. Black women are less likely to be screened and treated for osteoporosis than White women.[79] Studies comparing Hispanic and

NHW women had mixed results regarding screening but consistently found that Hispanic women were more likely to be treated for osteoporosis. In two different studies, Asian women were found to have similar rates of screening as White women but higher rates of treatment. Differences in screening and treatment could be attributed to patient factors (such as awareness of osteoporosis, competing health issues that require greater attention), clinician factors (e.g., knowledge and bias), and system factors related to differences in where patients access care.

Detailed Information

Racial disparities in osteoporosis screening and treatment exist (see **Appendix A Tables 6** and **7**). In the United States, Black women are less likely than White women to be screened[79, 391-397] and treated for osteoporosis,[97, 398-400] even after diagnosis of fracture.[401-405] They are also less likely than White women to receive preventive antiosteoporosis treatment after steroid initiation.[406] For example, in a 2015 retrospective clinic-based cohort study of women without prior treatment, screening, or diagnosis of osteoporosis (N=50,995), Black women were far less likely than White women (HR, 0.60 [95% CI, 0.54 to 0.65]) to have an incident DXA.[391] Comparisons between White women and other racial groups were mixed and sometimes inconsistent. Compared with White women, studies reported inconsistent findings: depending on the study, Hispanic women had lower,[397] similar,[391] or higher[394] rates of incident DXA. Hispanic women were also found to have higher rates of treatment after fracture[402, 403] and after a diagnosis of osteoporosis[400] than White women. Asian women were also found to have similar rates of incident DXA[391] but higher rates of treatment[400] after a diagnosis of osteoporosis compared with White women.[407]

Appendix A Table 6. Racial Differences in BMD Testing for Osteoporosis in Women Age 60 Years or Older

Race/Ethnicity	HR (95% CI)		OR (95% CI)				Proportion Receiving Testing
White	Referent group[61]	Referent group[64]	5.96 (3.01 to 11.79)[62]	Referent group[65]	Referent group[66]	Referent group[67]	38.4% ($p<0.05$)[63]
Asian	1.04 (0.96 to 1.13)						
Black	0.60 (0.54 to 0.65)	0.695 ($p<0.05$)	Reference group	0.39 (0.22 to 0.68)	0.47 (0.39 to 0.58)	0.52 (0.43 to 0.62)	29.8%
Hispanic	0.93 (0.86 to 1.01)	1.571 ($p<0.05$)				0.66 (0.54 to 0.80)	
Other	0.95 (0.87 to 1.04)						
Unknown	1.01 (0.96 to 1.06)						

Abbreviations: BMD=bone mineral density; CI=confidence interval; HR=hazard ratio; OR=odds ratio.

Appendix A Table 7. Racial Differences in Osteoporosis Treatment

Race/Ethnicity	OR (95% CI)			Marginal Effects*	Proportion Receiving Treatment
White	Referent group[66]	2.23 (1.76 to 2.84)[68]	Referent group[69]	Referent group[71]	89.2% ($p<0.05$)[63]
Asian				0.175 (0.139 to 0.211)	
Black	0.35 (0.30 to 0.41)	Referent group	0.36 (0.31 to 0.42)	-0.024 (-0.046 to -0.002)	79.6%
Hispanic				0.076 (0.051 to 0.103)	
Other		1.76 (1.21 to 2.55)		0.041 (-0.010 to 0.092)	
Unknown				Included in "other"	

* From logistic regression models; represent the change in predicted probability of treatment. Positive values represent higher likelihood of receiving treatment compared with White persons; negative values indicate lower likelihood of receiving treatment compared with White persons.

Abbreviations: CI=confidence interval; OR=odds ratio.

In Burgess and colleagues' review of provider contributions to racial health disparities, the authors described how ecological fallacies, whereby an individual is presumed to represent the racial population to which they belong, can contribute to disparities.[408] As such, the data indicating that White women are at greatest risk of fractures may result in reduced osteoporosis screening and treatment for those who do not share that identity. In fact, in a Canadian qualitative study of adults ages 50 to 79 years with a history of fracture, the authors found that provider understanding about racial differences in bone fragility was a barrier to BMD testing and treatment in a group of adults for which BMD testing would most certainly be indicated.[409] In a breakout session examining barriers to equitable osteoporosis care, participants identified lack of knowledge regarding the need to screen racially minoritized patients as a barrier.[410] Thus, racial disparities in fracture risk at a population level can translate into underscreening and undertreatment among racially and ethnically minoritized people.

Differences in care could also be attributed to provider bias, although we did not find any studies examining bias as it relates to osteoporosis screening or treatment. A study by Van Ryn and colleagues found that physicians held negative views of Black patients compared with White patients.[411] Additionally, studies showing racial disparities in pain management indicated that provider bias has significant impacts on patient care.[412] Racial animosity may unconsciously result in less time spent counseling and educating patients on their risk of osteoporosis and less interest in motivating and encouraging patients to complete screening. A recent study found that provider assumptions about the values held by racially and ethnically minoritized persons presented a barrier to advanced care planning.[413] Likewise, beliefs about patients' values regarding preventive care may also be associated with the extent to which clinicians spend time educating patients on osteoporosis and fracture risk and the effort they invest in ensuring that their patients get screened.

Racial differences in where patients access care may also be a contributor to racial differences in osteoporosis management. Racial and ethnic minorities are more likely than White patients to be seen by resident physicians,[414] who offer little patient continuity.[415] Lack of continuity may result in disengagement in preventive services. Few studies have been conducted evaluating resident and faculty care, with mixed results. One study found that residents and faculty scored similarly on health counseling metrics.[414] A more recent study found that residents' patients fared worse in chronic disease management and cancer screening than those of faculty.[416]

Patient factors also contribute to racial differences in osteoporosis screening and treatment. Solomon and colleagues found that patients who did not identify as White were less likely to adhere to osteoporosis treatment than those who did identify as White.[405] There are many explanations for this finding, for example, differences in care seeking for preventive care in general or the belief by patients themselves that Black women do not get osteoporosis could lead to reduced uptake of preventive treatments. In a qualitative study evaluating osteoporosis treatment preferences and medication adherence, some African American participants reported lack of interest in osteoporosis treatment given their low risk of fracture.[417] In this study, prescription fatigue was also a reason patients described for not taking medications, a problem of greater relevance to populations with a higher burden of disease. Medication cost could also be a factor: lower-income patients reduce pill burden to save money. Racial differences in educational achievement, a function of structural racism[418] that results in economic and educational inequity, likely translate into racial differences in osteoporosis knowledge,[419] which has an impact on treatment adherence.

B.1 Update Search Strategies

PubMed April 1, 2016, through July 14, 2021

Search	Query	Results
#1	"Osteoporosis"[mh] OR "Osteoporotic Fractures"[mh] OR "Fractures, Bone/prevention and control"[mh:noexp] OR "Decalcification, Pathologic"[mh] OR (("Bone Diseases, Metabolic"[mh:noexp] OR "Osteoporosis"[tiab] OR "Osteoporoses"[tiab] OR "osteoporotic"[tiab] OR "osteopenia"[tiab] OR "Age-Related Bone Loss"[tiab] OR "Age-Related Bone Losses"[tiab] OR "Calcaneus"[mh] OR "Menopause"[mh] OR "menopause"[tiab] OR "menopausal"[tiab] OR "postmenopause"[tiab] OR "postmenopausal"[tiab] OR "perimenopause"[tiab] OR "perimenopausal"[tiab] OR "Risk Factors"[Mesh]) AND ("Bone Density"[mh] OR "bone mineral density"[tiab] OR "bone density"[tiab] OR "density of bone"[tiab] OR "density of bones"[tiab] OR "bone loss"[tiab] OR "bone mass"[tiab] OR "brittle bone"[tiab] OR "brittle bones"[tiab] OR "fragile bone"[tiab] OR "fragile bones"[tiab] OR "broken bone"[tiab] OR "broken bones"[tiab] OR "bone health"[tiab] OR "health of bones"[tiab] OR "fractures, bone"[mh] OR "hip fractures"[mh] OR "spinal fractures"[mh] OR "fractures, spontaneous"[mh] OR "femoral fractures"[mh] OR "humeral fractures"[mh] OR "radius fractures"[mh] OR "ulna fractures"[mh] OR "fracture"[tiab] OR "fractures"[tiab] OR "fractured"[tiab] OR "bone turnover"[tiab] OR "bone resorption"[tiab] OR ("bone"[tiab] AND preserve*[tiab]) OR "bone formation"[tiab]))	110,697
#2	#1 AND (English[lang] AND ("2016/04/01"[Date - MeSH] : "3000"[Date - MeSH])) NOT (animals[mh] NOT humans[mh]) NOT (rat[ti] OR rats[ti] OR mouse[ti] OR mice[ti])	24,747
#3	#2 AND ("mass screening"[mh:noexp] OR "Diagnostic Screening Programs"[mh] OR "diagnostic imaging"[Subheading] OR "algorithms"[mh:noexp] OR "Surveys and Questionnaires"[mh] OR "risk assessment"[mh] OR screening[ti] OR screening[ot] OR "Absorptiometry, Photon"[mh] OR "Dual-Energy X-Ray Absorptiometry Scan"[tiab] OR "DXA"[tiab] OR "DEXA"[tiab] OR "Densitometry"[mh] OR "densitometry"[tiab] OR "Age Bulk One or Never Estrogens"[tiab] OR "ABONE"[tiab] OR "body weight criterion"[tiab] OR "BWC"[tiab] OR "Brown''s clinical risk assessment"[tiab] OR "Canadian Risk for Osteoporosis Calculator"[tiab] OR "CAROC"[tiab] OR "fracture absolute risk assessment"[tiab] OR "FARA"[tiab] OR "fracture risk assessment"[tiab] OR "FRAX"[tiab] OR "fracture risk score"[tiab] OR "fracture risk calculator"[tiab] OR "fracture risk tool"[tiab] OR "risk assessment"[ti] OR "risk assessment"[ot] OR "predictive model*"[tiab] OR "prognostic model*"[tiab] OR "Garvan"[tiab] OR "Hong Kong Osteoporosis Study"[tiab] OR "HKOS"[tiab] OR "Male Osteoporosis Risk Estimation Score"[tiab] OR "MORES"[tiab] OR "Osteoporosis Self-assessment Tool"[tiab] OR "OST"[tiab] OR "OSTA"[tiab] OR "OSTAi"[tiab] OR "risk assessment instrument"[tiab] OR "ORAI"[tiab] OR "Osteoporosis Index of Risk"[tiab] OR "OSIRIS"[tiab] OR "Q fracture"[tiab] OR "osteoporosis risk estimate"[tiab] OR "Study of Osteoporotic Fractures"[tiab] OR "SOF"[tiab] OR "SOFSURF"[tiab] OR "Weight-only-EPIDOS"[tiab] OR (("American Society for Bone and Mineral Research"[tiab] OR "ASBMR"[tiab] OR "International Society for Clinical Densitometry"[tiab] OR "ISCD"[tiab] OR "National Osteoporosis Foundation"[tiab] OR "National Osteoporosis Guideline Group"[tiab] OR "NOGG"[tiab] OR "World Health Organization"[tiab]) AND ("guideline"[tiab] OR "guidelines"[tiab])))	8,916
#4	#2 AND ("Diphosphonates"[mh:noexp] OR "Bisphosphonates"[tiab] OR "Bisphosphonate"[tiab] OR "Alendronate"[mh] OR "Alendronate"[tiab] or "alendronic acid"[tiab] OR "Fosamax"[tiab] OR "Binosto"[tiab] OR "Ibandronic Acid"[mh] OR "Ibandronic Acid"[tiab] OR "Ibandronate"[tiab] OR "Boniva"[tiab] OR "Bonviva"[tiab] OR "Bondronat"[tiab] OR "Risedronic Acid"[mh] OR "Risedronic Acid"[tiab] OR "Atelvia"[tiab] OR "Actonel"[tiab] OR "Risedronate"[tiab] OR "Zoledronic Acid"[mh] OR "Zoledronic Acid"[tiab] OR "Zometa"[tiab] OR "Zoledronate"[tiab] OR "Reclast"[tiab] OR "abaloparatide"[Supplementary Concept] OR "abaloparatide"[tiab] OR "Tymlos"[tiab] OR "Teriparatide"[mh] OR "Teriparatide"[tiab] OR "Forteo"[tiab] OR "Parathar"[tiab] OR "romosozumab"[Supplementary Concept] OR "romosozumab"[tiab] OR "evenity"[tiab] OR "sclerostin inhibitor"[tiab] OR "Denosumab"[mh] OR "Denosumab"[tiab] OR "Prolia"[tiab] OR "Xgeva"[tiab] OR "RANK Ligand"[mh] OR "RANK Ligand"[tiab] OR "Osteoprotegerin Ligand"[tiab] OR "TRANCE Protein"[tiab] OR "RANKL Protein"[tiab] OR "Osteoclast Differentiation Factor"[tiab])	3,376

Appendix B. Additional Methods

Search	Query	Results
#5	(#3 OR #4) AND ("randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR "random allocation"[mh] OR "randomized"[tiab] OR "randomized"[tiab] OR "randomization"[tiab] OR "randomization"[tiab] OR "randomly"[tiab] OR "placebos"[mh] OR placebo[tiab] OR placebos[tiab] OR "multicenter study"[pt] OR "comparative study"[pt] OR "comparative study"[tiab] OR "comparative"[ti] OR "clinical study"[pt:noexp] OR "clinical trial"[pt] OR "clinical trials as topic"[mh] OR "clinical protocols"[mh] OR "clinical trial"[tiab] OR "clinical trials"[tiab] OR ((trial[tiab] OR trials[tiab] OR study[tiab] OR studies[tiab]) AND ("control"[tiab] OR "controlled"[tiab] OR "controls"[tiab] OR group[tiab] OR groups[tiab] OR volunteer*[tiab] OR cohort[tiab] OR cohorts[tiab])) OR "single-blind method"[mh] OR single blind*[tiab] OR "double-blind method"[mh] OR double blind*[tiab] OR triple blind*[tiab] OR ((singl*[tiab] OR doubl*[tiab] OR trebl*[tiab] OR tripl*[tiab]) AND (mask*[tiab] OR blind* [tiab])) OR "treatment outcome"[mh] OR "evaluation studies"[pt] OR "evaluation studies as topic"[mh] OR "evaluation study"[tiab] OR "evaluation studies"[tiab] OR "intervention study"[tiab] OR "intervention studies"[tiab] OR systematic[sb] OR "systematic review"[tiab] OR "meta-analysis"[pt] OR meta-analysis[mh] OR "meta-analysis as topic"[mh] OR "meta-analysis"[tiab] OR "meta-analyses"[tiab] OR "meta-synthesis"[tiab] OR "meta-syntheses"[tiab] OR "survival analysis"[mh] OR "systematic overview"[tiab] OR "quantitative review"[tiab] OR "quantitative synthesis"[tiab] OR "quantitative syntheses"[tiab] OR "pooled analysis"[tiab] OR "pooled analyses"[tiab] OR "meta-regression"[tiab] OR "data synthesis"[tiab] OR "data syntheses"[tiab] OR "data extraction"[tiab] OR "data abstraction"[tiab] OR "fixed effect"[tiab] OR "fixed effects"[tiab] OR "indirect comparison"[tiab] OR (("indirect treatment"[tiab] OR "mixed-treatment"[tiab]) AND ("comparison"[tiab] OR "comparisons"[tiab])) OR "comparative efficacy"[tiab] OR "comparative effectiveness"[tiab])	6,683
#6	(#3 OR #4) AND ("observational study"[pt] OR "observational studies as topic"[mh] OR "observation"[mh] OR "observational"[tiab] OR "cohort studies"[mh] OR "cohort"[tiab] OR "cohorts"[tiab] OR "concurrent study"[tiab] OR "concurrent studies"[tiab] OR "incidence study"[tiab] OR "incidence studies"[tiab] OR "follow-up studies"[mh] OR "follow-up"[tiab] OR "followup"[tiab] OR "longitudinal studies"[mh] OR "longitudinal"[tiab] OR "longitudinally"[tiab] OR "prospective studies"[mh] OR "prospective"[tiab] OR "prospectively"[tiab] OR "case-control studies"[mh] OR "case-control"[tiab] OR "case-crossover"[tiab] OR "retrospective studies"[mh] OR "retrospective"[tiab] OR "nonexperimental"[tiab] OR "non-experimental"[tiab] OR "nonrandomized"[tiab] OR "nonrandomized"[tiab] OR "non-randomised"[tiab] OR "nonrandomized"[tiab] OR "adverse effects"[subheading])	6,554
#7	#3 AND ("predictive value of tests"[mh] OR "models, statistical"[mh] OR "logistic models"[mh] OR "logistic models"[mh] OR "sensitivity and specificity"[mh] OR "roc curve"[mh] OR "proportional hazards models"[mh] OR "area under curve"[mh] OR "analysis of variance"[mh] OR "models, statistical"[mh] OR "fracture prediction"[ot] OR "reproducibility of results"[mh] OR "accuracy"[tiab] OR "discrimination"[tiab] OR "discriminant validity"[tiab] OR "goodness-of-fit"[tiab] OR "Hosmer-Lemeshow"[tiab] OR "c-statistic*"[tiab] OR "cstatistic*"[tiab] OR "calibrat*"[tiab] OR (("accurac*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value*"[tiab]) AND "predict*"[tiab]) OR (("accurac*"[tiab] OR "effectiveness"[tiab] OR "efficac*"[tiab] OR "error*"[tiab] OR "perform*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value"[tiab] OR "yield*"[tiab]) AND "diagnostic*"[tiab]) OR "receiver operat*"[tiab])	2,014
#9	#5 OR #6 OR #7	8,910

Cochrane Central April 1, 2016, through July 14, 2021

Search	Query	Results
#1	[mh “Osteoporosis”] OR [mh “Osteoporotic Fractures”] OR [mh ^”Fractures, Bone”/PC] OR [mh “Decalcification, Pathologic”] OR (([mh ^”Bone Diseases, Metabolic”] OR “Osteoporosis”:ti,ab,kw OR “Osteoporoses”:ti,ab,kw OR “osteoporotic”:ti,ab,kw OR “osteopenia”:ti,ab,kw OR “Age-Related Bone Loss”:ti,ab,kw OR “Age-Related Bone Losses”:ti,ab,kw OR [mh “Calcaneus”] OR [mh “Menopause”] OR “menopause”:ti,ab,kw OR “menopausal”:ti,ab,kw OR “postmenopause”:ti,ab,kw OR “postmenopausal”:ti,ab,kw OR “perimenopause”:ti,ab,kw OR “perimenopausal”:ti,ab,kw OR [mh “Risk Factors”]) AND ([mh “Bone Density”] OR “bone mineral density”:ti,ab,kw OR “bone density”:ti,ab,kw OR “density of bone”:ti,ab,kw OR “density of bones”:ti,ab,kw OR “bone loss”:ti,ab,kw OR “bone mass”:ti,ab,kw OR “brittle bone”:ti,ab,kw OR “brittle bones”:ti,ab,kw OR “fragile bone”:ti,ab,kw OR “fragile bones”:ti,ab,kw OR “broken bone”:ti,ab,kw OR “broken bones”:ti,ab,kw OR “bone health”:ti,ab,kw OR “health of bones”:ti,ab,kw OR [mh “fractures, bone”] OR [mh “hip fractures”] OR [mh “spinal fractures”] OR [mh “fractures, spontaneous”] OR [mh “femoral fractures”] OR [mh “humeral fractures”] OR [mh “radius fractures”] OR [mh “ulna fractures”] OR “fracture”:ti,ab,kw OR “fractures”:ti,ab,kw OR “fractured”:ti,ab,kw OR “bone turnover”:ti,ab,kw OR “bone resorption”:ti,ab,kw OR (“bone” NEAR preserve*):ti,ab,kw OR “bone formation”:ti,ab,kw))	12,289
#2	[mh ^”mass screening”] OR [mh “Diagnostic Screening Programs”] OR [mh ^”algorithms”] OR [mh “Surveys and Questionnaires”] OR [mh “risk assessment”] OR “screening”:ti OR [mh “Absorptiometry, Photon”] OR “Dual-Energy X-Ray Absorptiometry Scan”:ti,ab,kw OR “DXA”:ti,ab,kw OR “DEXA”:ti,ab,kw OR [mh “Densitometry”] OR “densitometry”:ti,ab,kw OR “Age Bulk One or Never Estrogens”:ti,ab,kw OR “ABONE”:ti,ab,kw OR “body weight criterion”:ti,ab,kw OR “BWC”:ti,ab,kw OR “Brown’’s clinical risk assessment”:ti,ab,kw OR “Canadian Risk for Osteoporosis Calculator”:ti,ab,kw OR “CAROC”:ti,ab,kw OR “fracture absolute risk assessment”:ti,ab,kw OR “FARA”:ti,ab,kw OR “fracture risk assessment”:ti,ab,kw OR “FRAX”:ti,ab,kw OR “fracture risk score”:ti,ab,kw OR “fracture risk calculator”:ti,ab,kw OR “fracture risk tool”:ti,ab,kw OR “risk assessment”:ti,ab,kw OR (predictive NEXT model*):ti,ab,kw OR (prognostic NEXT model*):ti,ab,kw OR “Garvan”:ti,ab,kw OR “Hong Kong Osteoporosis Study”:ti,ab,kw OR “HKOS”:ti,ab,kw OR “Male Osteoporosis Risk Estimation Score”:ti,ab,kw OR “MORES”:ti,ab,kw OR “Osteoporosis Self-assessment Tool”:ti,ab,kw OR “OST”:ti,ab,kw OR “OSTA”:ti,ab,kw OR “OSTAi”:ti,ab,kw OR “risk assessment instrument”:ti,ab,kw OR “ORAI”:ti,ab,kw OR “Osteoporosis Index of Risk”:ti,ab,kw OR “OSIRIS”:ti,ab,kw OR “Q fracture”:ti,ab,kw OR “osteoporosis risk estimate”:ti,ab,kw OR “Study of Osteoporotic Fractures”:ti,ab,kw OR “SOF”:ti,ab,kw OR “SOFSURF”:ti,ab,kw OR “Weight-only-EPIDOS”:ti,ab,kw OR (“American Society for Bone and Mineral Research” OR “ASBMR” OR “International Society for Clinical Densitometry” OR “ISCD” OR “National Osteoporosis Foundation” OR “National Osteoporosis Guideline Group” OR “NOGG” OR “World Health Organization” NEAR “guideline” OR “guidelines”):ti,ab,kw	130,942
#3	[mh ^”Diphosphonates”] OR “Bisphosphonates”:ti,ab,kw OR “Bisphosphonate”:ti,ab,kw OR [mh “Alendronate”] OR “Alendronate”:ti,ab,kw OR “alendronic acid”:ti,ab,kw OR “Fosamax”:ti,ab,kw OR “Binosto”:ti,ab,kw OR [mh “Ibandronic Acid”] OR “Ibandronic Acid”:ti,ab,kw OR “Ibandronate”:ti,ab,kw OR “Boniva”:ti,ab,kw OR “Bonviva”:ti,ab,kw OR “Bondronat”:ti,ab,kw OR [mh “Risedronic Acid”] OR “Risedronic Acid”:ti,ab,kw OR “Atelvia”:ti,ab,kw OR “Actonel”:ti,ab,kw OR “Risedronate”:ti,ab,kw OR [mh “Zoledronic Acid”] OR “Zoledronic Acid”:ti,ab,kw OR “Zometa”:ti,ab,kw OR “Zoledronate”:ti,ab,kw OR “Reclast”:ti,ab,kw OR [mh “abaloparatide”] OR “abaloparatide”:ti,ab,kw OR “Tymlos”:ti,ab,kw OR [mh “Teriparatide”] OR “Teriparatide”:ti,ab,kw OR “Forteo”:ti,ab,kw OR “Parathar”:ti,ab,kw OR [mh “romosozumab”] OR “romosozumab”:ti,ab,kw OR “evenity”:ti,ab,kw OR “sclerostin inhibitor”:ti,ab,kw OR [mh “Denosumab”] OR “Denosumab”:ti,ab,kw OR “Prolia”:ti,ab,kw OR “Xgeva”:ti,ab,kw OR [mh “RANK Ligand”] OR “RANK Ligand”:ti,ab,kw OR “Osteoprotegerin Ligand”:ti,ab,kw OR “TRANCE Protein”:ti,ab,kw OR “RANKL Protein”:ti,ab,kw OR “Osteoclast Differentiation Factor”:ti,ab,kw	6,619
#4	#1 AND (OR #2-#3) with Publication Year from 2016 to 2021, with Cochrane Library publication date from Apr 2016 to Jul 2021, in Trials	1,781

Embase April 1, 2016, through July 14, 2021

Search	Query	Results
#1	''osteoporosis''/de OR ''fragility fracture''/de OR ((''metabolic bone disease''/de OR ''bone demineralization''/de OR ''Osteoporosis'':ti,ab OR ''Osteoporoses'':ti,ab OR ''osteoporotic'':ti,ab OR ''osteopenia'':ti,ab OR ''Age-Related Bone Loss'':ti,ab OR ''Age-Related Bone Losses'':ti,ab OR ''calcaneus''/exp OR ''menopause and climacterium''/exp OR ''menopause'':ti,ab OR ''menopausal'':ti,ab OR ''postmenopause'':ti,ab OR ''postmenopausal'':ti,ab OR ''perimenopause'':ti,ab OR ''perimenopausal'':ti,ab OR ''risk factor''/exp) AND (''bone density''/exp OR ''bone mineral density'':ti,ab OR ''bone density'':ti,ab OR ''density of bone'':ti,ab OR ''density of bones'':ti,ab OR ''bone loss'':ti,ab OR ''bone mass'':ti,ab OR ''brittle bone'':ti,ab OR ''brittle bones'':ti,ab OR ''fragile bone'':ti,ab OR ''fragile bones'':ti,ab OR ''broken bone'':ti,ab OR ''broken bones'':ti,ab OR ''bone health'':ti,ab OR ''health of bones'':ti,ab OR ''fracture''/exp OR ''fracture'':ti,ab OR ''fractures'':ti,ab OR ''fractured'':ti,ab OR ''bone turnover'':ti,ab OR ''bone resorption'':ti,ab OR (''bone'' NEAR/6 preserve*):ti,ab OR ''bone formation'':ti,ab)) AND [humans]/lim AND [english]/lim AND [2016-2021]/py	46,522
#2	''mass screening''/exp OR ''mass screening'':ti,ab OR ''diagnostic screening programs''/exp OR ''diagnostic screening'':ti,ab OR ''diagnostic imaging''/exp OR ''diagnostic imaging'':ti,ab OR ''algorithms''/de OR ''algorithms'':ti,ab OR ''surveys and questionnaires''/de OR ''surveys and questionnaires'':ti,ab OR ''screening''/exp OR screening OR ''photon absorptiometry''/exp OR ''photon absorptiometry'':ti,ab OR ''dual-energy x-ray absorptiometry scan'':ti,ab OR ''dxa''/exp OR ''dxa'':ti,ab OR ''dexa''/exp OR ''dexa'':ti,ab OR ''densitometry''/exp OR ''densitometry'':ti,ab OR ''age bulk one or never estrogens'':ti,ab OR ''abone'':ti,ab OR ''body weight criterion'':ti,ab OR ''bwc'':ti,ab OR ''clinical risk assessment'':ti,ab OR ''canadian risk for osteoporosis calculator'':ti,ab OR ''caroc'':ti,ab OR ''fracture absolute risk assessment'':ti,ab OR ''fara'':ti,ab OR ''fracture risk assessment''/exp OR ''fracture risk assessment'':ti,ab OR ''frax''/exp OR ''frax'':ti,ab OR ''fracture risk score'':ti,ab OR ''fracture risk calculator''/exp OR ''fracture risk calculator'':ti,ab OR ''fracture risk tool'':ti,ab OR ''risk assessment''/exp OR ''risk assessment'':ti,ab OR ''predictive model*'':ti,ab OR ''prognostic model*'':ti,ab OR ''garvan'':ti,ab OR ''hong kong osteoporosis study'':ti,ab OR ''hkos'':ti,ab OR ''male osteoporosis risk estimation score'':ti,ab OR ''mores'':ti,ab OR ''osteoporosis self-assessment tool'':ti,ab OR ''ost'':ti,ab OR ''osta'':ti,ab OR ''ostai'':ti,ab OR ''risk assessment instrument'':ti,ab OR ''orai'':ti,ab OR ''osteoporosis index of risk'':ti,ab OR ''osiris'':ti,ab OR ''q fracture'':ti,ab OR ''osteoporosis risk estimate'':ti,ab OR ''study of osteoporotic fractures'':ti,ab OR ''sof'':ti,ab OR ''sofsurf'':ti,ab OR ''weight-only-epidos'':ti,ab OR ((''American Society for Bone and Mineral Research'':ti,ab OR ''ASBMR'':ti,ab OR ''International Society for Clinical Densitometry'':ti,ab OR ''ISCD'':ti,ab OR ''National Osteoporosis Foundation'':ti,ab OR ''National Osteoporosis Guideline Group'':ti,ab OR ''NOGG'':ti,ab OR ''World Health Organization'') AND (''guideline'':ti,ab OR ''guidelines'':ti,ab))	9,552,835
#3	''diphosphonates''/exp OR ''diphosphonates'':ti,ab OR ''bisphosphonates''/exp OR ''bisphosphonates'':ti,ab OR ''bisphosphonate''/exp OR ''bisphosphonate'':ti,ab OR ''alendronate''/exp OR ''alendronate'':ti,ab OR ''alendronic acid''/exp OR ''alendronic acid'':ti,ab OR ''fosamax''/exp OR ''fosamax'':ti,ab OR ''binosto''/exp OR ''binosto'':ti,ab OR ''ibandronic acid''/exp OR ''ibandronic acid'':ti,ab OR ''ibandronate''/exp OR ''ibandronate'':ti,ab OR ''boniva''/exp OR ''boniva'':ti,ab OR ''bonviva''/exp OR ''bonviva'':ti,ab OR ''bondronat''/exp OR ''bondronat'':ti,ab OR ''risedronic acid''/exp OR ''risedronic acid'':ti,ab OR ''atelvia''/exp OR ''atelvia'':ti,ab OR ''actonel''/exp OR ''actonel'':ti,ab OR ''risedronate''/exp OR ''risedronate'':ti,ab OR ''zoledronic acid''/exp OR ''zoledronic acid'':ti,ab OR ''zometa''/exp OR ''zometa'':ti,ab OR ''zoledronate''/exp OR ''zoledronate'':ti,ab OR ''reclast''/exp OR ''reclast'':ti,ab OR ''abaloparatide''/exp OR ''abaloparatide'':ti,ab OR ''tymlos''/exp OR ''tymlos'':ti,ab OR ''teriparatide''/exp OR ''teriparatide'':ti,ab OR ''forteo''/exp OR ''forteo'':ti,ab OR ''parathar''/exp OR ''parathar'':ti,ab OR ''romosozumab''/exp OR ''romosozumab'':ti,ab OR ''evenity''/exp OR ''evenity'':ti,ab OR ''sclerostin inhibitor''/exp OR ''sclerostin inhibitor'':ti,ab OR ''denosumab''/exp OR ''denosumab'':ti,ab OR ''prolia''/exp OR ''prolia'':ti,ab OR ''xgeva''/exp OR ''xgeva'':ti,ab OR ''rank ligand''/exp OR ''rank ligand'':ti,ab OR ''osteoprotegerin ligand''/exp OR ''osteoprotegerin ligand'':ti,ab OR ''trance protein'':ti,ab OR ''rankl protein'':ti,ab OR ''osteoclast differentiation factor''/exp OR ''osteoclast differentiation factor'':ti,ab	97,487
#4	#1 AND (#2 OR #3)	29,710
#5	#4 AND ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim OR	15,163

Appendix B. Additional Methods

Search	Query	Results
	[controlled clinical trial]/lim OR [randomized controlled trial]/lim OR ''observational study''/exp OR ''cohort studies''/exp OR ''follow-up studies''/exp OR ''longitudinal studies''/exp OR ''prospective studies''/exp OR ''case-control studies''/exp OR ''retrospective studies''/exp OR ''adverse effects''/exp)	
#6	#1 AND #2 AND (''predictive value of tests''/exp OR ''predictive value of tests'' OR ''logistic models''/exp OR ''logistic models'' OR ''sensitivity next specificity'' OR ''roc curve''/exp OR ''roc curve'' OR ''proportional hazards models''/exp OR ''proportional hazards models'' OR ''area under curve''/exp OR ''area under curve'' OR ''analysis of variance''/exp OR ''analysis of variance'' OR ''models, statistical''/exp OR ''fracture prediction'' OR ''reproducibility of results''/exp OR ''reproducibility of results'' OR ''accuracy''/exp OR ''accuracy'' OR ''discrimination''/exp OR ''discrimination'' OR ''discriminant validity''/exp OR ''discriminant validity'' OR ''goodness-of-fit'' OR ''hosmer-lemeshow'' OR ''c-statistic*'' OR ''cstatistic*'' OR ''calibrat*'' OR ((''accurac*'' OR ''reliability'' OR ''validity'' OR ''value*'') NEAR/4 ''predict*'') OR ((''accurac*'' OR ''effectiveness'' OR ''efficac*'' OR ''error*'' OR ''perform*'' OR ''reliability'' OR ''validity'' OR ''value'' OR ''yield*'') NEAR/4 ''diagnostic*'') OR ''receiver operat*'')	3,325
	((#5 OR #6) AND ''osteoporosis''/dm AND ''article''/it) NOT [medline]/lim	903

Bridge Search PubMed, July 1, 2021, through November 10, 2022

Search	Query	Results
#1	"Osteoporosis"[mh] OR "Osteoporotic Fractures"[mh] OR "Fractures, Bone/prevention and control"[mh:noexp] OR "Decalcification, Pathologic"[mh] OR (("Bone Diseases, Metabolic"[mh:noexp] OR "Osteoporosis"[tiab] OR "Osteoporoses"[tiab] OR "osteoporotic"[tiab] OR "osteopenia"[tiab] OR "Age-Related Bone Loss"[tiab] OR "Age-Related Bone Losses"[tiab] OR "Calcaneus"[mh] OR "Menopause"[mh] OR "menopause"[tiab] OR "menopausal"[tiab] OR "postmenopause"[tiab] OR "postmenopausal"[tiab] OR "perimenopause"[tiab] OR "perimenopausal"[tiab] OR "Risk Factors"[Mesh]) AND ("Bone Density"[mh] OR "bone mineral density"[tiab] OR "bone density"[tiab] OR "density of bone"[tiab] OR "density of bones"[tiab] OR "bone loss"[tiab] OR "bone mass"[tiab] OR "brittle bone"[tiab] OR "brittle bones"[tiab] OR "fragile bone"[tiab] OR "fragile bones"[tiab] OR "broken bone"[tiab] OR "broken bones"[tiab] OR "bone health"[tiab] OR "health of bones"[tiab] OR "fractures, bone"[mh] OR "hip fractures"[mh] OR "spinal fractures"[mh] OR "fractures, spontaneous"[mh] OR "femoral fractures"[mh] OR "humeral fractures"[mh] OR "radius fractures"[mh] OR "ulna fractures"[mh] OR "fracture"[tiab] OR "fractures"[tiab] OR "fractured"[tiab] OR "bone turnover"[tiab] OR "bone resorption"[tiab] OR ("bone"[tiab] AND preserve*[tiab]) OR "bone formation"[tiab]))	118,277
#2	#1 AND (English[lang] AND ("2021/07/01"[Date - MeSH] : "3000"[Date - MeSH])) NOT (animals[mh] NOT humans[mh]) NOT (rat[ti] OR rats[ti] OR mouse[ti] OR mice[ti])	8,340
#3	#2 AND ("mass screening"[mh:noexp] OR "Diagnostic Screening Programs"[mh] OR "diagnostic imaging"[Subheading] OR "algorithms"[mh:noexp] OR "Surveys and Questionnaires"[mh] OR "risk assessment"[mh] OR screening[ti] OR screening[ot] OR "Absorptiometry, Photon"[mh] OR "Dual-Energy X-Ray Absorptiometry Scan"[tiab] OR "DXA"[tiab] OR "DEXA"[tiab] OR "Densitometry"[mh] OR "densitometry"[tiab] OR "Age Bulk One or Never Estrogens"[tiab] OR "ABONE"[tiab] OR "body weight criterion"[tiab] OR "BWC"[tiab] OR "Brown’s clinical risk assessment"[tiab] OR "Canadian Risk for Osteoporosis Calculator"[tiab] OR "CAROC"[tiab] OR "fracture absolute risk assessment"[tiab] OR "FARA"[tiab] OR "fracture risk assessment"[tiab] OR "FRAX"[tiab] OR "fracture risk score"[tiab] OR "fracture risk calculator"[tiab] OR "fracture risk tool"[tiab] OR "risk assessment"[ti] OR "risk assessment"[ot] OR "predictive model*"[tiab] OR "prognostic model*"[tiab] OR "Garvan"[tiab] OR "Hong Kong Osteoporosis Study"[tiab] OR "HKOS"[tiab] OR "Male Osteoporosis Risk Estimation Score"[tiab] OR "MORES"[tiab] OR "Osteoporosis Self-assessment Tool"[tiab] OR "OST"[tiab] OR "OSTA"[tiab] OR "OSTAi"[tiab] OR "risk assessment instrument"[tiab] OR "ORAI"[tiab] OR "Osteoporosis Index of Risk"[tiab] OR "OSIRIS"[tiab] OR "Q fracture"[tiab] OR "osteoporosis risk estimate"[tiab] OR "Study of Osteoporotic Fractures"[tiab] OR "SOF"[tiab] OR "SOFSURF"[tiab] OR "Weight-only-EPIDOS"[tiab] OR (("American Society for Bone and	2,805

Appendix B. Additional Methods

Search	Query	Results
#3 (contin-ued)	Mineral Research"[tiab] OR "ASBMR"[tiab] OR "International Society for Clinical Densitometry"[tiab] OR "ISCD"[tiab] OR "National Osteoporosis Foundation"[tiab] OR "National Osteoporosis Guideline Group"[tiab] OR "NOGG"[tiab] OR "World Health Organization"[tiab]) AND ("guideline"[tiab] OR "guidelines"[tiab])))	
#4	#2 AND ("Diphosphonates"[mh:noexp] OR "Bisphosphonates"[tiab] OR "Bisphosphonate"[tiab] OR "Alendronate"[mh] OR "Alendronate"[tiab] or "alendronic acid"[tiab] OR "Fosamax"[tiab] OR "Binosto"[tiab] OR "Ibandronic Acid"[mh] OR "Ibandronic Acid"[tiab] OR "Ibandronate"[tiab] OR "Boniva"[tiab] OR "Bonviva"[tiab] OR "Bondronat"[tiab] OR "Risedronic Acid"[mh] OR "Risedronic Acid"[tiab] OR "Atelvia"[tiab] OR "Actonel"[tiab] OR "Risedronate"[tiab] OR "Zoledronic Acid"[mh] OR "Zoledronic Acid"[tiab] OR "Zometa"[tiab] OR "Zoledronate"[tiab] OR "Reclast"[tiab] OR "abaloparatide"[Supplementary Concept] OR "abaloparatide"[tiab] OR "Tymlos"[tiab] OR "Teriparatide"[mh] OR "Teriparatide"[tiab] OR "Forteo"[tiab] OR "Parathar"[tiab] OR "romosozumab"[Supplementary Concept] OR "romosozumab"[tiab] OR "evenity"[tiab] OR "sclerostin inhibitor"[tiab] OR "Denosumab"[mh] OR "Denosumab"[tiab] OR "Prolia"[tiab] OR "Xgeva"[tiab] OR "RANK Ligand"[mh] OR "RANK Ligand"[tiab] OR "Osteoprotegerin Ligand"[tiab] OR "TRANCE Protein"[tiab] OR "RANKL Protein"[tiab] OR "Osteoclast Differentiation Factor"[tiab])	1,066
#5	(#3 OR #4) AND ("randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR "random allocation"[mh] OR "randomized"[tiab] OR "randomized"[tiab] OR "randomization"[tiab] OR "randomization"[tiab] OR "randomly"[tiab] OR "placebos"[mh] OR placebo[tiab] OR placebos[tiab] OR "multicenter study"[pt] OR "comparative study"[pt] OR "comparative study"[tiab] OR "comparative"[ti] OR "clinical study"[pt:noexp] OR "clinical trial"[pt] OR "clinical trials as topic"[mh] OR "clinical protocols"[mh] OR "clinical trial"[tiab] OR "clinical trials"[tiab] OR ((trial[tiab] OR trials[tiab] OR study[tiab] OR studies[tiab]) AND ("control"[tiab] OR "controlled"[tiab] OR "controls"[tiab] OR group[tiab] OR groups[tiab] OR volunteer*[tiab] OR cohort[tiab] OR cohorts[tiab])) OR "single-blind method"[mh] OR single blind*[tiab] OR "double-blind method"[mh] OR double blind*[tiab] OR triple blind*[tiab] OR ((singl*[tiab] OR doubl*[tiab] OR trebl*[tiab] OR tripl*[tiab]) AND (mask*[tiab] OR blind*[tiab])) OR "treatment outcome"[mh] OR "evaluation studies"[pt] OR "evaluation studies as topic"[mh] OR "evaluation study"[tiab] OR "evaluation studies"[tiab] OR "intervention study"[tiab] OR "intervention studies"[tiab] OR systematic[sb] OR "systematic review"[tiab] OR "meta-analysis"[pt] OR meta-analysis[mh] OR "meta-analysis as topic"[mh] OR "meta-analysis"[tiab] OR "meta-analyses"[tiab] OR "meta-synthesis"[tiab] OR "meta-syntheses"[tiab] OR "survival analysis"[mh] OR "systematic overview"[tiab] OR "quantitative review"[tiab] OR "quantitative synthesis"[tiab] OR "quantitative syntheses"[tiab] OR "pooled analysis"[tiab] OR "pooled analyses"[tiab] OR "meta-regression"[tiab] OR "data synthesis"[tiab] OR "data syntheses"[tiab] OR "data extraction"[tiab] OR "data abstraction"[tiab] OR "fixed effect"[tiab] OR "fixed effects"[tiab] OR "indirect comparison"[tiab] OR (("indirect treatment"[tiab] OR "mixed-treatment"[tiab]) AND ("comparison"[tiab] OR "comparisons"[tiab])) OR "comparative efficacy"[tiab] OR "comparative effectiveness"[tiab])	2,044
#6	(#3 OR #4) AND ("observational study"[pt] OR "observational studies as topic"[mh] OR "observation"[mh] OR "observational"[tiab] OR "cohort studies"[mh] OR "cohort"[tiab] OR "cohorts"[tiab] OR "concurrent study"[tiab] OR "concurrent studies"[tiab] OR "incidence study"[tiab] OR "incidence studies"[tiab] OR "follow-up studies"[mh] OR "follow-up"[tiab] OR "followup"[tiab] OR "longitudinal studies"[mh] OR "longitudinal"[tiab] OR "longitudinally"[tiab] OR "prospective studies"[mh] OR "prospective"[tiab] OR "prospectively"[tiab] OR "case-control studies"[mh] OR "case-control"[tiab] OR "case-crossover"[tiab] OR "retrospective studies"[mh] OR "retrospective"[tiab] OR "nonexperimental"[tiab] OR "non-experimental"[tiab] OR "nonrandomized"[tiab] OR "nonrandomized"[tiab] OR "non-randomised"[tiab] OR "nonrandomized"[tiab] OR "adverse effects"[subheading])	2,002

Search	Query	Results
#7	#3 AND ("predictive value of tests"[mh] OR "models, statistical"[mh] OR "logistic models"[mh] OR "logistic models"[mh] OR "sensitivity and specificity"[mh] OR "roc curve"[mh] OR "proportional hazards models"[mh] OR "area under curve"[mh] OR "analysis of variance"[mh] OR "models, statistical"[mh] OR "fracture prediction"[ot] OR "reproducibility of results"[mh] OR "accuracy"[tiab] OR "discrimination"[tiab] OR "discriminant validity"[tiab] OR "goodness-of-fit"[tiab] OR "Hosmer-Lemeshow"[tiab] OR "c-statistic*"[tiab] OR "cstatistic*"[tiab] OR "calibrat*"[tiab] OR (("accurac*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value*"[tiab]) AND "predict*"[tiab]) OR (("accurac*"[tiab] OR "effectiveness"[tiab] OR "efficac*"[tiab] OR "error*"[tiab] OR "perform*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value"[tiab] OR "yield*"[tiab]) AND "diagnostic*"[tiab]) OR "receiver operat*"[tiab])	620
#9	#5 OR #6 OR #7	2,773

Bridge Search Cochrane Central 2020 through November 10, 2022

Search	Query	Results
#1	[mh "Osteoporosis"] OR [mh "Osteoporotic Fractures"] OR [mh ^"Fractures, Bone"/PC] OR [mh "Decalcification, Pathologic"] OR (([mh ^"Bone Diseases, Metabolic"] OR "Osteoporosis":ti,ab,kw OR "Osteoporoses":ti,ab,kw OR "osteoporotic":ti,ab,kw OR "osteopenia":ti,ab,kw OR "Age-Related Bone Loss":ti,ab,kw OR "Age-Related Bone Losses":ti,ab,kw OR [mh "Calcaneus"] OR [mh "Menopause"] OR "menopause":ti,ab,kw OR "menopausal":ti,ab,kw OR "postmenopause":ti,ab,kw OR "postmenopausal":ti,ab,kw OR "perimenopause":ti,ab,kw OR "perimenopausal":ti,ab,kw OR [mh "Risk Factors"]) AND ([mh "Bone Density"] OR "bone mineral density":ti,ab,kw OR "bone density":ti,ab,kw OR "density of bone":ti,ab,kw OR "density of bones":ti,ab,kw OR "bone loss":ti,ab,kw OR "bone mass":ti,ab,kw OR "brittle bone":ti,ab,kw OR "brittle bones":ti,ab,kw OR "fragile bone":ti,ab,kw OR "fragile bones":ti,ab,kw OR "broken bone":ti,ab,kw OR "broken bones":ti,ab,kw OR "bone health":ti,ab,kw OR "health of bones":ti,ab,kw OR [mh "fractures, bone"] OR [mh "hip fractures"] OR [mh "spinal fractures"] OR [mh "fractures, spontaneous"] OR [mh "femoral fractures"] OR [mh "humeral fractures"] OR [mh "radius fractures"] OR [mh "ulna fractures"] OR "fracture":ti,ab,kw OR "fractures":ti,ab,kw OR "fractured":ti,ab,kw OR "bone turnover":ti,ab,kw OR "bone resorption":ti,ab,kw OR ("bone" NEAR preserve*):ti,ab,kw OR "bone formation":ti,ab,kw))	12,906
#2	[mh ^"mass screening"] OR [mh "Diagnostic Screening Programs"] OR [mh ^"algorithms"] OR [mh "Surveys and Questionnaires"] OR [mh "risk assessment"] OR "screening":ti OR [mh "Absorptiometry, Photon"] OR "Dual-Energy X-Ray Absorptiometry Scan":ti,ab,kw OR "DXA":ti,ab,kw OR "DEXA":ti,ab,kw OR [mh "Densitometry"] OR "densitometry":ti,ab,kw OR "Age Bulk One or Never Estrogens":ti,ab,kw OR "ABONE":ti,ab,kw OR "body weight criterion":ti,ab,kw OR "BWC":ti,ab,kw OR "Brown’s clinical risk assessment":ti,ab,kw OR "Canadian Risk for Osteoporosis Calculator":ti,ab,kw OR "CAROC":ti,ab,kw OR "fracture absolute risk assessment":ti,ab,kw OR "FARA":ti,ab,kw OR "fracture risk assessment":ti,ab,kw OR "FRAX":ti,ab,kw OR "fracture risk score":ti,ab,kw OR "fracture risk calculator":ti,ab,kw OR "fracture risk tool":ti,ab,kw OR "risk assessment":ti,ab,kw OR (predictive NEXT model*):ti,ab,kw OR (prognostic NEXT model*):ti,ab,kw OR "Garvan":ti,ab,kw OR "Hong Kong Osteoporosis Study":ti,ab,kw OR "HKOS":ti,ab,kw OR "Male Osteoporosis Risk Estimation Score":ti,ab,kw OR "MORES":ti,ab,kw OR "Osteoporosis Self-assessment Tool":ti,ab,kw OR "OST":ti,ab,kw OR "OSTA":ti,ab,kw OR "OSTAi":ti,ab,kw OR "risk assessment instrument":ti,ab,kw OR "ORAI":ti,ab,kw OR "Osteoporosis Index of Risk":ti,ab,kw OR "OSIRIS":ti,ab,kw OR "Q fracture":ti,ab,kw OR "osteoporosis risk estimate":ti,ab,kw OR "Study of Osteoporotic Fractures":ti,ab,kw OR "SOF":ti,ab,kw OR "SOFSURF":ti,ab,kw OR "Weight-only-EPIDOS":ti,ab,kw OR ("American Society for Bone and Mineral Research" OR "ASBMR" OR "International Society for Clinical Densitometry" OR "ISCD" OR "National Osteoporosis Foundation" OR "National Osteoporosis Guideline Group" OR "NOGG" OR "World Health Organization" NEAR "guideline" OR "guidelines"):ti,ab,kw	143,262

Appendix B. Additional Methods

Search	Query	Results
#3	[mh ^"Diphosphonates"] OR "Bisphosphonates":ti,ab,kw OR "Bisphosphonate":ti,ab,kw OR [mh "Alendronate"] OR "Alendronate":ti,ab,kw OR "alendronic acid":ti,ab,kw OR "Fosamax":ti,ab,kw OR "Binosto":ti,ab,kw OR [mh "Ibandronic Acid"] OR "Ibandronic Acid":ti,ab,kw OR "Ibandronate":ti,ab,kw OR "Boniva":ti,ab,kw OR "Bonviva":ti,ab,kw OR "Bondronat":ti,ab,kw OR [mh "Risedronic Acid"] OR "Risedronic Acid":ti,ab,kw OR "Atelvia":ti,ab,kw OR "Actonel":ti,ab,kw OR "Risedronate":ti,ab,kw OR [mh "Zoledronic Acid"] OR "Zoledronic Acid":ti,ab,kw OR "Zometa":ti,ab,kw OR "Zoledronate":ti,ab,kw OR "Reclast":ti,ab,kw OR [mh "abaloparatide"] OR "abaloparatide":ti,ab,kw OR "Tymlos":ti,ab,kw OR [mh "Teriparatide"] OR "Teriparatide":ti,ab,kw OR "Forteo":ti,ab,kw OR "Parathar":ti,ab,kw OR [mh "romosozumab"] OR "romosozumab":ti,ab,kw OR "evenity":ti,ab,kw OR "sclerostin inhibitor":ti,ab,kw OR [mh "Denosumab"] OR "Denosumab":ti,ab,kw OR "Prolia":ti,ab,kw OR "Xgeva":ti,ab,kw OR [mh "RANK Ligand"] OR "RANK Ligand":ti,ab,kw OR "Osteoprotegerin Ligand":ti,ab,kw OR "TRANCE Protein":ti,ab,kw OR "RANKL Protein":ti,ab,kw OR "Osteoclast Differentiation Factor":ti,ab,kw	6,891
#4	#1 AND (OR #2-#3) with Publication Year from 2020 to 2022, in Trials	692

Bridge Search Embase 2021 through November 10, 2022

Search	Query	Results
#1	'osteoporosis'/de OR 'fragility fracture'/de OR (('metabolic bone disease'/de OR 'bone demineralization'/de OR 'Osteoporosis':ti,ab OR 'Osteoporoses':ti,ab OR 'osteoporotic':ti,ab OR 'osteopenia':ti,ab OR 'Age-Related Bone Loss':ti,ab OR 'Age-Related Bone Losses':ti,ab OR 'calcaneus'/exp OR 'menopause and climacterium'/exp OR 'menopause':ti,ab OR 'menopausal':ti,ab OR 'postmenopause':ti,ab OR 'postmenopausal':ti,ab OR 'perimenopause':ti,ab OR 'perimenopausal':ti,ab OR 'risk factor'/exp) AND ('bone density'/exp OR 'bone mineral density':ti,ab OR 'bone density':ti,ab OR 'density of bone':ti,ab OR 'density of bones':ti,ab OR 'bone loss':ti,ab OR 'bone mass':ti,ab OR 'brittle bone':ti,ab OR 'brittle bones':ti,ab OR 'fragile bone':ti,ab OR 'fragile bones':ti,ab OR 'broken bone':ti,ab OR 'broken bones':ti,ab OR 'bone health':ti,ab OR 'health of bones':ti,ab OR 'fracture'/exp OR 'fracture':ti,ab OR 'fractures':ti,ab OR 'fractured':ti,ab OR 'bone turnover':ti,ab OR 'bone resorption':ti,ab OR ('bone' NEAR/6 preserve*):ti,ab OR 'bone formation':ti,ab)) AND [humans]/lim AND [english]/lim AND [2021-2022]/py	17,571
#2	'mass screening'/exp OR 'mass screening':ti,ab OR 'diagnostic screening programs'/exp OR 'diagnostic screening':ti,ab OR 'diagnostic imaging'/exp OR 'diagnostic imaging':ti,ab OR 'algorithms'/de OR 'algorithms':ti,ab OR 'surveys and questionnaires'/de OR 'surveys and questionnaires':ti,ab OR 'screening'/exp OR screening OR 'photon absorptiometry'/exp OR 'photon absorptiometry':ti,ab OR 'dual-energy x-ray absorptiometry scan':ti,ab OR 'dxa'/exp OR 'dxa':ti,ab OR 'dexa'/exp OR 'dexa':ti,ab OR 'densitometry'/exp OR 'densitometry':ti,ab OR 'age bulk one or never estrogens':ti,ab OR 'abone':ti,ab OR 'body weight criterion':ti,ab OR 'bwc':ti,ab OR 'clinical risk assessment':ti,ab OR 'canadian risk for osteoporosis calculator':ti,ab OR 'caroc':ti,ab OR 'fracture absolute risk assessment':ti,ab OR 'fara':ti,ab OR 'fracture risk assessment'/exp OR 'fracture risk assessment':ti,ab OR 'frax'/exp OR 'frax':ti,ab OR 'fracture risk score':ti,ab OR 'fracture risk calculator'/exp OR 'fracture risk calculator':ti,ab OR 'fracture risk tool':ti,ab OR 'risk assessment'/exp OR 'risk assessment':ti,ab OR 'predictive model*':ti,ab OR 'prognostic model*':ti,ab OR 'garvan':ti,ab OR 'hong kong osteoporosis study':ti,ab OR 'hkos':ti,ab OR 'male osteoporosis risk estimation score':ti,ab OR 'mores':ti,ab OR 'osteoporosis self-assessment tool':ti,ab OR 'ost':ti,ab OR 'osta':ti,ab OR 'ostai':ti,ab OR 'risk assessment instrument':ti,ab OR 'orai':ti,ab OR 'osteoporosis index of risk':ti,ab OR 'osiris':ti,ab OR 'q fracture':ti,ab OR 'osteoporosis risk estimate':ti,ab OR 'study of osteoporotic fractures':ti,ab OR 'sof':ti,ab OR 'sofsurf':ti,ab OR 'weight-only-epidos':ti,ab OR (('American Society for Bone and Mineral Research':ti,ab OR 'ASBMR':ti,ab OR 'International Society for Clinical Densitometry':ti,ab OR 'ISCD':ti,ab OR 'National Osteoporosis Foundation':ti,ab OR 'National Osteoporosis Guideline Group':ti,ab OR 'NOGG':ti,ab OR 'World Health Organization') AND ('guideline':ti,ab OR 'guidelines':ti,ab))	10,393,945

Appendix B. Additional Methods

Search	Query	Results
#3	'diphosphonates'/exp OR 'diphosphonates':ti,ab OR 'bisphosphonates'/exp OR 'bisphosphonates':ti,ab OR 'bisphosphonate'/exp OR 'bisphosphonate':ti,ab OR 'alendronate'/exp OR 'alendronate':ti,ab OR 'alendronic acid'/exp OR 'alendronic acid':ti,ab OR 'fosamax'/exp OR 'fosamax':ti,ab OR 'binosto'/exp OR 'binosto':ti,ab OR 'ibandronic acid'/exp OR 'ibandronic acid':ti,ab OR 'ibandronate'/exp OR 'ibandronate':ti,ab OR 'boniva'/exp OR 'boniva':ti,ab OR 'bonviva'/exp OR 'bonviva':ti,ab OR 'bondronat'/exp OR 'bondronat':ti,ab OR 'risedronic acid'/exp OR 'risedronic acid':ti,ab OR 'atelvia'/exp OR 'atelvia':ti,ab OR 'actonel'/exp OR 'actonel':ti,ab OR 'risedronate'/exp OR 'risedronate':ti,ab OR 'zoledronic acid'/exp OR 'zoledronic acid':ti,ab OR 'zometa'/exp OR 'zometa':ti,ab OR 'zoledronate'/exp OR 'zoledronate':ti,ab OR 'reclast'/exp OR 'reclast':ti,ab OR 'abaloparatide'/exp OR 'abaloparatide':ti,ab OR 'tymlos'/exp OR 'tymlos':ti,ab OR 'teriparatide'/exp OR 'teriparatide':ti,ab OR 'forteo'/exp OR 'forteo':ti,ab OR 'parathar'/exp OR 'parathar':ti,ab OR 'romosozumab'/exp OR 'romosozumab':ti,ab OR 'evenity'/exp OR 'evenity':ti,ab OR 'sclerostin inhibitor'/exp OR 'sclerostin inhibitor':ti,ab OR 'denosumab'/exp OR 'denosumab':ti,ab OR 'prolia'/exp OR 'prolia':ti,ab OR 'xgeva'/exp OR 'xgeva':ti,ab OR 'rank ligand'/exp OR 'rank ligand':ti,ab OR 'osteoprotegerin ligand'/exp OR 'osteoprotegerin ligand':ti,ab OR 'trance protein':ti,ab OR 'rankl protein':ti,ab OR 'osteoclast differentiation factor'/exp OR 'osteoclast differentiation factor':ti,ab	105,192
#4	#1 AND (#2 OR #3)	11,133
#5	#4 AND ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim OR [controlled clinical trial]/lim OR [randomized controlled trial]/lim OR 'observational study'/exp OR 'cohort studies'/exp OR 'follow-up studies'/exp OR 'longitudinal studies'/exp OR 'prospective studies'/exp OR 'case-control studies'/exp OR 'retrospective studies'/exp OR 'adverse effects'/exp)	6,478
#6	#1 AND #2 AND ('predictive value of tests'/exp OR 'predictive value of tests' OR 'logistic models'/exp OR 'logistic models' OR 'sensitivity next specificity' OR 'roc curve'/exp OR 'roc curve' OR 'proportional hazards models'/exp OR 'proportional hazards models' OR 'area under curve'/exp OR 'area under curve' OR 'analysis of variance'/exp OR 'analysis of variance' OR 'models, statistical'/exp OR 'fracture prediction' OR 'reproducibility of results'/exp OR 'reproducibility of results' OR 'accuracy'/exp OR 'accuracy' OR 'discrimination'/exp OR 'discrimination' OR 'discriminant validity'/exp OR 'discriminant validity' OR 'goodness-of-fit' OR 'hosmer-lemeshow' OR 'c-statistic*' OR 'cstatistic*' OR 'calibrat*' OR (('accurac*' OR 'reliability' OR 'validity' OR 'value*') NEAR/4 'predict*') OR (('accurac*' OR 'effectiveness' OR 'efficac*' OR 'error*' OR 'perform*' OR 'reliability' OR 'validity' OR 'value' OR 'yield*') NEAR/4 'diagnostic*') OR 'receiver operat*')	1,676
#7	((#5 OR #6) AND 'osteoporosis'/dm AND 'article'/it) NOT [medline]/lim	849

Bridge Search PubMed August 1, 2022, through January 9, 2024

Search	Query	Results
#1	"Osteoporosis"[mh] OR "Osteoporotic Fractures"[mh] OR "Fractures, Bone/prevention and control"[mh:noexp] OR "Decalcification, Pathologic"[mh] OR (("Bone Diseases, Metabolic"[mh:noexp] OR "Osteoporosis"[tiab] OR "Osteoporoses"[tiab] OR "osteoporotic"[tiab] OR "osteopenia"[tiab] OR "Age-Related Bone Loss"[tiab] OR "Age-Related Bone Losses"[tiab] OR "Calcaneus"[mh] OR "Menopause"[mh] OR "menopause"[tiab] OR "menopausal"[tiab] OR "postmenopause"[tiab] OR "postmenopausal"[tiab] OR "perimenopause"[tiab] OR "perimenopausal"[tiab] OR "Risk Factors"[Mesh]) AND ("Bone Density"[mh] OR "bone mineral density"[tiab] OR "bone density"[tiab] OR "density of bone"[tiab] OR "density of bones"[tiab] OR "bone loss"[tiab] OR "bone mass"[tiab] OR "brittle bone"[tiab] OR "brittle bones"[tiab] OR "fragile bone"[tiab] OR "fragile bones"[tiab] OR "broken bone"[tiab] OR "broken bones"[tiab] OR "bone health"[tiab] OR "health of bones"[tiab] OR "fractures, bone"[mh] OR "hip fractures"[mh] OR "spinal fractures"[mh] OR "fractures, spontaneous"[mh] OR "femoral fractures"[mh] OR "humeral fractures"[mh] OR "radius fractures"[mh] OR "ulna fractures"[mh] OR "fracture"[tiab] OR "fractures"[tiab] OR "fractured"[tiab] OR "bone turnover"[tiab] OR "bone resorption"[tiab] OR ("bone"[tiab] AND preserve*[tiab]) OR "bone formation"[tiab]))	124,382

Appendix B. Additional Methods

#2	#1 AND (English[lang] AND ("2022/08/01"[Date - MeSH] : "3000"[Date - MeSH])) NOT (animals[mh] NOT humans[mh]) NOT (rat[ti] OR rats[ti] OR mouse[ti] OR mice[ti])	7,171
#3	#2 AND ("mass screening"[mh:noexp] OR "Diagnostic Screening Programs"[mh] OR "diagnostic imaging"[Subheading] OR "algorithms"[mh:noexp] OR "Surveys and Questionnaires"[mh] OR "risk assessment"[mh] OR screening[ti] OR screening[ot] OR "Absorptiometry, Photon"[mh] OR "Dual-Energy X-Ray Absorptiometry Scan"[tiab] OR "DXA"[tiab] OR "DEXA"[tiab] OR "Densitometry"[mh] OR "densitometry"[tiab] OR "Age Bulk One or Never Estrogens"[tiab] OR "ABONE"[tiab] OR "body weight criterion"[tiab] OR "BWC"[tiab] OR "Brown's clinical risk assessment"[tiab] OR "Canadian Risk for Osteoporosis Calculator"[tiab] OR "CAROC"[tiab] OR "fracture absolute risk assessment"[tiab] OR "FARA"[tiab] OR "fracture risk assessment"[tiab] OR "FRAX"[tiab] OR "fracture risk score"[tiab] OR "fracture risk calculator"[tiab] OR "fracture risk tool"[tiab] OR "risk assessment"[ti] OR "risk assessment"[ot] OR "predictive model*"[tiab] OR "prognostic model*"[tiab] OR "Garvan"[tiab] OR "Hong Kong Osteoporosis Study"[tiab] OR "HKOS"[tiab] OR "Male Osteoporosis Risk Estimation Score"[tiab] OR "MORES"[tiab] OR "Osteoporosis Self-assessment Tool"[tiab] OR "OST"[tiab] OR "OSTA"[tiab] OR "OSTAi"[tiab] OR "risk assessment instrument"[tiab] OR "ORAI"[tiab] OR "Osteoporosis Index of Risk"[tiab] OR "OSIRIS"[tiab] OR "Q fracture"[tiab] OR "osteoporosis risk estimate"[tiab] OR "Study of Osteoporotic Fractures"[tiab] OR "SOF"[tiab] OR "SOFSURF"[tiab] OR "Weight-only-EPIDOS"[tiab] OR (("American Society for Bone and Mineral Research"[tiab] OR "ASBMR"[tiab] OR "International Society for Clinical Densitometry"[tiab] OR "ISCD"[tiab] OR "National Osteoporosis Foundation"[tiab] OR "National Osteoporosis Guideline Group"[tiab] OR "NOGG"[tiab] OR "World Health Organization"[tiab]) AND ("guideline"[tiab] OR "guidelines"[tiab])))	2,111
#4	#2 AND ("Diphosphonates"[mh:noexp] OR "Bisphosphonates"[tiab] OR "Bisphosphonate"[tiab] OR "Alendronate"[mh] OR "Alendronate"[tiab] or "alendronic acid"[tiab] OR "Fosamax"[tiab] OR "Binosto"[tiab] OR "Ibandronic Acid"[mh] OR "Ibandronic Acid"[tiab] OR "Ibandronate"[tiab] OR "Boniva"[tiab] OR "Bonviva"[tiab] OR "Bondronat"[tiab] OR "Risedronic Acid"[mh] OR "Risedronic Acid"[tiab] OR "Atelvia"[tiab] OR "Actonel"[tiab] OR "Risedronate"[tiab] OR "Zoledronic Acid"[mh] OR "Zoledronic Acid"[tiab] OR "Zometa"[tiab] OR "Zoledronate"[tiab] OR "Reclast"[tiab] OR "abaloparatide"[Supplementary Concept] OR "abaloparatide"[tiab] OR "Tymlos"[tiab] OR "Teriparatide"[mh] OR "Teriparatide"[tiab] OR "Forteo"[tiab] OR "Parathar"[tiab] OR "romosozumab"[Supplementary Concept] OR "romosozumab"[tiab] OR "evenity"[tiab] OR "sclerostin inhibitor"[tiab] OR "Denosumab"[mh] OR "Denosumab"[tiab] OR "Prolia"[tiab] OR "Xgeva"[tiab] OR "RANK Ligand"[mh] OR "RANK Ligand"[tiab] OR "Osteoprotegerin Ligand"[tiab] OR "TRANCE Protein"[tiab] OR "RANKL Protein"[tiab] OR "Osteoclast Differentiation Factor"[tiab])	841
#5	(#3 OR #4) AND ("randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR "random allocation"[mh] OR "randomized"[tiab] OR "randomized"[tiab] OR "randomization"[tiab] OR "randomization"[tiab] OR "randomly"[tiab] OR "placebos"[mh] OR placebo[tiab] OR placebos[tiab] OR "multicenter study"[pt] OR "comparative study"[pt] OR "comparative study"[tiab] OR "comparative"[ti] OR "clinical study"[pt:noexp] OR "clinical trial"[pt] OR "clinical trials as topic"[mh] OR "clinical protocols"[mh] OR "clinical trial"[tiab] OR "clinical trials"[tiab] OR ((trial[tiab] OR trials[tiab] OR study[tiab] OR studies[tiab]) AND ("control"[tiab] OR "controlled"[tiab] OR "controls"[tiab] OR group[tiab] OR groups[tiab] OR volunteer*[tiab] OR cohort[tiab] OR cohorts[tiab])) OR "single-blind method"[mh] OR single blind*[tiab] OR "double-blind method"[mh] OR double blind*[tiab] OR triple blind*[tiab] OR ((singl*[tiab] OR doubl*[tiab] OR trebl*[tiab] OR tripl*[tiab]) AND (mask*[tiab] OR blind* [tiab])) OR "treatment outcome"[mh] OR "evaluation studies"[pt] OR "evaluation studies as topic"[mh] OR "evaluation study"[tiab] OR "evaluation studies"[tiab] OR "intervention study"[tiab] OR "intervention studies"[tiab] OR systematic[sb] OR "systematic review"[tiab] OR "meta-analysis"[pt] OR meta-analysis[mh] OR "meta-analysis as topic"[mh] OR "meta-analysis"[tiab] OR "meta-analyses"[tiab] OR "meta-synthesis"[tiab] OR "meta-syntheses"[tiab] OR "survival analysis"[mh] OR "systematic overview"[tiab] OR "quantitative review"[tiab] OR "quantitative synthesis"[tiab] OR "quantitative syntheses"[tiab] OR "pooled analysis"[tiab] OR "pooled analyses"[tiab] OR "meta-regression"[tiab] OR "data synthesis"[tiab] OR "data syntheses"[tiab] OR "data extraction"[tiab] OR "data abstraction"[tiab] OR "fixed effect"[tiab] OR "fixed effects"[tiab] OR "indirect comparison"[tiab] OR (("indirect treatment"[tiab] OR "mixed-treatment"[tiab]) AND ("comparison"[tiab] OR "comparisons"[tiab])) OR "comparative efficacy"[tiab] OR "comparative effectiveness"[tiab])	1,502
#6	(#3 OR #4) AND ("observational study"[pt] OR "observational studies as topic"[mh] OR "observation"[mh] OR "observational"[tiab] OR "cohort studies"[mh] OR "cohort"[tiab] OR	1,522

	"cohorts"[tiab] OR "concurrent study"[tiab] OR "concurrent studies"[tiab] OR "incidence study"[tiab] OR "incidence studies"[tiab] OR "follow-up studies"[mh] OR "follow-up"[tiab] OR "followup"[tiab] OR "longitudinal studies"[mh] OR "longitudinal"[tiab] OR "longitudinally"[tiab] OR "prospective studies"[mh] OR "prospective"[tiab] OR "prospectively"[tiab] OR "case-control studies"[mh] OR "case-control"[tiab] OR "case-crossover"[tiab] OR "retrospective studies"[mh] OR "retrospective"[tiab] OR "nonexperimental"[tiab] OR "non-experimental"[tiab] OR "nonrandomized"[tiab] OR "nonrandomized"[tiab] OR "non-randomised"[tiab] OR "nonrandomized"[tiab] OR "adverse effects"[subheading])	
#7	#3 AND ("predictive value of tests"[mh] OR "models, statistical"[mh] OR "logistic models"[mh] OR "logistic models"[mh] OR "sensitivity and specificity"[mh] OR "roc curve"[mh] OR "proportional hazards models"[mh] OR "area under curve"[mh] OR "analysis of variance"[mh] OR "models, statistical"[mh] OR "fracture prediction"[ot] OR "reproducibility of results"[mh] OR "accuracy"[tiab] OR "discrimination"[tiab] OR "discriminant validity"[tiab] OR "goodness-of-fit"[tiab] OR "Hosmer-Lemeshow"[tiab] OR "c-statistic*"[tiab] OR "cstatistic*"[tiab] OR "calibrat*"[tiab] OR (("accurac*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value*"[tiab]) AND "predict*"[tiab]) OR (("accurac*"[tiab] OR "effectiveness"[tiab] OR "efficac*"[tiab] OR "error*"[tiab] OR "perform*"[tiab] OR "reliability"[tiab] OR "validity"[tiab] OR "value"[tiab] OR "yield*"[tiab]) AND "diagnostic*"[tiab]) OR "receiver operat*"[tiab])	530
#8	#5 OR #6 OR #7	2,108

Bridge Search Cochrane Central 2022 through January 9, 2024

Search	Query	Results
#1	[mh "Osteoporosis"] OR [mh "Osteoporotic Fractures"] OR [mh ^"Fractures, Bone"/PC] OR [mh "Decalcification, Pathologic"] OR (([mh ^"Bone Diseases, Metabolic"] OR "Osteoporosis":ti,ab,kw OR "Osteoporoses":ti,ab,kw OR "osteoporotic":ti,ab,kw OR "osteopenia":ti,ab,kw OR "Age-Related Bone Loss":ti,ab,kw OR "Age-Related Bone Losses":ti,ab,kw OR [mh "Calcaneus"] OR [mh "Menopause"] OR "menopause":ti,ab,kw OR "menopausal":ti,ab,kw OR "postmenopause":ti,ab,kw OR "postmenopausal":ti,ab,kw OR "perimenopause":ti,ab,kw OR "perimenopausal":ti,ab,kw OR [mh "Risk Factors"]) AND ([mh "Bone Density"] OR "bone mineral density":ti,ab,kw OR "bone density":ti,ab,kw OR "density of bone":ti,ab,kw OR "density of bones":ti,ab,kw OR "bone loss":ti,ab,kw OR "bone mass":ti,ab,kw OR "brittle bone":ti,ab,kw OR "brittle bones":ti,ab,kw OR "fragile bone":ti,ab,kw OR "fragile bones":ti,ab,kw OR "broken bone":ti,ab,kw OR "broken bones":ti,ab,kw OR "bone health":ti,ab,kw OR "health of bones":ti,ab,kw OR [mh "fractures, bone"] OR [mh "hip fractures"] OR [mh "spinal fractures"] OR [mh "fractures, spontaneous"] OR [mh "femoral fractures"] OR [mh "humeral fractures"] OR [mh "radius fractures"] OR [mh "ulna fractures"] OR "fracture":ti,ab,kw OR "fractures":ti,ab,kw OR "fractured":ti,ab,kw OR "bone turnover":ti,ab,kw OR "bone resorption":ti,ab,kw OR ("bone" NEAR preserve*):ti,ab,kw OR "bone formation":ti,ab,kw))	13,958
#2	[mh ^"mass screening"] OR [mh "Diagnostic Screening Programs"] OR [mh ^"algorithms"] OR [mh "Surveys and Questionnaires"] OR [mh "risk assessment"] OR "screening":ti OR [mh "Absorptiometry, Photon"] OR "Dual-Energy X-Ray Absorptiometry Scan":ti,ab,kw OR "DXA":ti,ab,kw OR "DEXA":ti,ab,kw OR [mh "Densitometry"] OR "densitometry":ti,ab,kw OR "Age Bulk One or Never Estrogens":ti,ab,kw OR "ABONE":ti,ab,kw OR "body weight criterion":ti,ab,kw OR "BWC":ti,ab,kw OR "Brown’s clinical risk assessment":ti,ab,kw OR "Canadian Risk for Osteoporosis Calculator":ti,ab,kw OR "CAROC":ti,ab,kw OR "fracture absolute risk assessment":ti,ab,kw OR "FARA":ti,ab,kw OR "fracture risk assessment":ti,ab,kw OR "FRAX":ti,ab,kw OR "fracture risk score":ti,ab,kw OR "fracture risk calculator":ti,ab,kw OR "fracture risk tool":ti,ab,kw OR "risk assessment":ti,ab,kw OR (predictive NEXT model*):ti,ab,kw OR (prognostic NEXT model*):ti,ab,kw OR "Garvan":ti,ab,kw OR "Hong Kong Osteoporosis Study":ti,ab,kw OR "HKOS":ti,ab,kw OR "Male Osteoporosis Risk Estimation Score":ti,ab,kw OR "MORES":ti,ab,kw OR "Osteoporosis Self-assessment Tool":ti,ab,kw OR "OST":ti,ab,kw OR "OSTA":ti,ab,kw OR "OSTAi":ti,ab,kw OR "risk assessment instrument":ti,ab,kw OR "ORAI":ti,ab,kw OR "Osteoporosis Index of Risk":ti,ab,kw OR "OSIRIS":ti,ab,kw OR "Q fracture":ti,ab,kw OR "osteoporosis risk estimate":ti,ab,kw OR "Study of Osteoporotic Fractures":ti,ab,kw OR "SOF":ti,ab,kw OR "SOFSURF":ti,ab,kw OR "Weight-only-EPIDOS":ti,ab,kw OR ("American Society for Bone and Mineral Research" OR "ASBMR" OR "International Society for Clinical Densitometry" OR	164,755

	"ISCD" OR "National Osteoporosis Foundation" OR "National Osteoporosis Guideline Group" OR "NOGG" OR "World Health Organization" NEAR "guideline" OR "guidelines"):ti,ab,kw	
#3	[mh ^"Diphosphonates"] OR "Bisphosphonates":ti,ab,kw OR "Bisphosphonate":ti,ab,kw OR [mh "Alendronate"] OR "Alendronate":ti,ab,kw OR "alendronic acid":ti,ab,kw OR "Fosamax":ti,ab,kw OR "Binosto":ti,ab,kw OR [mh "Ibandronic Acid"] OR "Ibandronic Acid":ti,ab,kw OR "Ibandronate":ti,ab,kw OR "Boniva":ti,ab,kw OR "Bonviva":ti,ab,kw OR "Bondronat":ti,ab,kw OR [mh "Risedronic Acid"] OR "Risedronic Acid":ti,ab,kw OR "Atelvia":ti,ab,kw OR "Actonel":ti,ab,kw OR "Risedronate":ti,ab,kw OR [mh "Zoledronic Acid"] OR "Zoledronic Acid":ti,ab,kw OR "Zometa":ti,ab,kw OR "Zoledronate":ti,ab,kw OR "Reclast":ti,ab,kw OR [mh "abaloparatide"] OR "abaloparatide":ti,ab,kw OR "Tymlos":ti,ab,kw OR [mh "Teriparatide"] OR "Teriparatide":ti,ab,kw OR "Forteo":ti,ab,kw OR "Parathar":ti,ab,kw OR [mh "romosozumab"] OR "romosozumab":ti,ab,kw OR "evenity":ti,ab,kw OR "sclerostin inhibitor":ti,ab,kw OR [mh "Denosumab"] OR "Denosumab":ti,ab,kw OR "Prolia":ti,ab,kw OR "Xgeva":ti,ab,kw OR [mh "RANK Ligand"] OR "RANK Ligand":ti,ab,kw OR "Osteoprotegerin Ligand":ti,ab,kw OR "TRANCE Protein":ti,ab,kw OR "RANKL Protein":ti,ab,kw OR "Osteoclast Differentiation Factor":ti,ab,kw	7,271
#4	#1 AND (OR #2-#3) with Publication Year from 2022 to 2024, in Trials	413

Bridge Search Embase 2022 through January 9, 2024

Search	Query	Results
#1	'osteoporosis'/de OR 'fragility fracture'/de OR (('metabolic bone disease'/de OR 'bone demineralization'/de OR 'Osteoporosis':ti,ab OR 'Osteoporoses':ti,ab OR 'osteoporotic':ti,ab OR 'osteopenia':ti,ab OR 'Age-Related Bone Loss':ti,ab OR 'Age-Related Bone Losses':ti,ab OR 'calcaneus'/exp OR 'menopause and climacterium'/exp OR 'menopause':ti,ab OR 'menopausal':ti,ab OR 'postmenopause':ti,ab OR 'postmenopausal':ti,ab OR 'perimenopause':ti,ab OR 'perimenopausal':ti,ab OR 'risk factor'/exp) AND ('bone density'/exp OR 'bone mineral density':ti,ab OR 'bone density':ti,ab OR 'density of bone':ti,ab OR 'density of bones':ti,ab OR 'bone loss':ti,ab OR 'bone mass':ti,ab OR 'brittle bone':ti,ab OR 'brittle bones':ti,ab OR 'fragile bone':ti,ab OR 'fragile bones':ti,ab OR 'broken bone':ti,ab OR 'broken bones':ti,ab OR 'bone health':ti,ab OR 'health of bones':ti,ab OR 'fracture'/exp OR 'fracture':ti,ab OR 'fractures':ti,ab OR 'fractured':ti,ab OR 'bone turnover':ti,ab OR 'bone resorption':ti,ab OR ('bone' NEAR/6 preserve*):ti,ab OR 'bone formation':ti,ab)) AND [humans]/lim AND [english]/lim AND [2022-2024]/py	20,539
#2	'mass screening'/exp OR 'mass screening':ti,ab OR 'diagnostic screening programs'/exp OR 'diagnostic screening':ti,ab OR 'diagnostic imaging'/exp OR 'diagnostic imaging':ti,ab OR 'algorithms'/de OR 'algorithms':ti,ab OR 'surveys and questionnaires'/de OR 'surveys and questionnaires':ti,ab OR 'screening'/exp OR screening OR 'photon absorptiometry'/exp OR 'photon absorptiometry':ti,ab OR 'dual-energy x-ray absorptiometry scan':ti,ab OR 'dxa'/exp OR 'dxa':ti,ab OR 'dexa'/exp OR 'dexa':ti,ab OR 'densitometry'/exp OR 'densitometry':ti,ab OR 'age bulk one or never estrogens':ti,ab OR 'abone':ti,ab OR 'body weight criterion':ti,ab OR 'bwc':ti,ab OR 'clinical risk assessment':ti,ab OR 'canadian risk for osteoporosis calculator':ti,ab OR 'caroc':ti,ab OR 'fracture absolute risk assessment':ti,ab OR 'fara':ti,ab OR 'fracture risk assessment'/exp OR 'fracture risk assessment':ti,ab OR 'frax'/exp OR 'frax':ti,ab OR 'fracture risk score':ti,ab OR 'fracture risk calculator'/exp OR 'fracture risk calculator':ti,ab OR 'fracture risk tool':ti,ab OR 'risk assessment'/exp OR 'risk assessment':ti,ab OR 'predictive model*':ti,ab OR 'prognostic model*':ti,ab OR 'garvan':ti,ab OR 'hong kong osteoporosis study':ti,ab OR 'hkos':ti,ab OR 'male osteoporosis risk estimation score':ti,ab OR 'mores':ti,ab OR 'osteoporosis self-assessment tool':ti,ab OR 'ost':ti,ab OR 'osta':ti,ab OR 'ostai':ti,ab OR 'risk assessment instrument':ti,ab OR 'orai':ti,ab OR 'osteoporosis index of risk':ti,ab OR 'osiris':ti,ab OR 'q fracture':ti,ab OR 'osteoporosis risk estimate':ti,ab OR 'study of osteoporotic fractures':ti,ab OR 'sof':ti,ab OR 'sofsurf':ti,ab OR 'weight-only-epidos':ti,ab OR (('American Society for Bone and Mineral Research':ti,ab OR 'ASBMR':ti,ab OR 'International Society for Clinical Densitometry':ti,ab OR 'ISCD':ti,ab OR 'National Osteoporosis Foundation':ti,ab OR 'National Osteoporosis Guideline Group':ti,ab OR 'NOGG':ti,ab OR 'World Health Organization') AND ('guideline':ti,ab OR 'guidelines':ti,ab))	11,145,144
#3	'diphosphonates'/exp OR 'diphosphonates':ti,ab OR 'bisphosphonates'/exp OR 'bisphosphonates':ti,ab OR 'bisphosphonate'/exp OR 'bisphosphonate':ti,ab OR	111,931

	'alendronate'/exp OR 'alendronate':ti,ab OR 'alendronic acid'/exp OR 'alendronic acid':ti,ab OR 'fosamax'/exp OR 'fosamax':ti,ab OR 'binosto'/exp OR 'binosto':ti,ab OR 'ibandronic acid'/exp OR 'ibandronic acid':ti,ab OR 'ibandronate'/exp OR 'ibandronate':ti,ab OR 'boniva'/exp OR 'boniva':ti,ab OR 'bonviva'/exp OR 'bonviva':ti,ab OR 'bondronat'/exp OR 'bondronat':ti,ab OR 'risedronic acid'/exp OR 'risedronic acid':ti,ab OR 'atelvia'/exp OR 'atelvia':ti,ab OR 'actonel'/exp OR 'actonel':ti,ab OR 'risedronate'/exp OR 'risedronate':ti,ab OR 'zoledronic acid'/exp OR 'zoledronic acid':ti,ab OR 'zometa'/exp OR 'zometa':ti,ab OR 'zoledronate'/exp OR 'zoledronate':ti,ab OR 'reclast'/exp OR 'reclast':ti,ab OR 'abaloparatide'/exp OR 'abaloparatide':ti,ab OR 'tymlos'/exp OR 'tymlos':ti,ab OR 'teriparatide'/exp OR 'teriparatide':ti,ab OR 'forteo'/exp OR 'forteo':ti,ab OR 'parathar'/exp OR 'parathar':ti,ab OR 'romosozumab'/exp OR 'romosozumab':ti,ab OR 'evenity'/exp OR 'evenity':ti,ab OR 'sclerostin inhibitor'/exp OR 'sclerostin inhibitor':ti,ab OR 'denosumab'/exp OR 'denosumab':ti,ab OR 'prolia'/exp OR 'prolia':ti,ab OR 'xgeva'/exp OR 'xgeva':ti,ab OR 'rank ligand'/exp OR 'rank ligand':ti,ab OR 'osteoprotegerin ligand'/exp OR 'osteoprotegerin ligand':ti,ab OR 'trance protein':ti,ab OR 'rankl protein':ti,ab OR 'osteoclast differentiation factor'/exp OR 'osteoclast differentiation factor':ti,ab	
#4	#1 AND (#2 OR #3)	13,017
#5	#4 AND ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim OR [controlled clinical trial]/lim OR [randomized controlled trial]/lim OR 'observational study'/exp OR 'cohort studies'/exp OR 'follow-up studies'/exp OR 'longitudinal studies'/exp OR 'prospective studies'/exp OR 'case-control studies'/exp OR 'retrospective studies'/exp OR 'adverse effects'/exp)	7,503
#6	#1 AND #2 AND ('predictive value of tests'/exp OR 'predictive value of tests' OR 'logistic models'/exp OR 'logistic models' OR 'sensitivity next specificity' OR 'roc curve'/exp OR 'roc curve' OR 'proportional hazards models'/exp OR 'proportional hazards models' OR 'area under curve'/exp OR 'area under curve' OR 'analysis of variance'/exp OR 'analysis of variance' OR 'models, statistical'/exp OR 'fracture prediction' OR 'reproducibility of results'/exp OR 'reproducibility of results' OR 'accuracy'/exp OR 'accuracy' OR 'discrimination'/exp OR 'discrimination' OR 'discriminant validity'/exp OR 'discriminant validity' OR 'goodness-of-fit' OR 'hosmer-lemeshow' OR 'c-statistic*' OR 'cstatistic*' OR 'calibrat*' OR (('accurac*' OR 'reliability' OR 'validity' OR 'value*') NEAR/4 'predict*') OR (('accurac*' OR 'effectiveness' OR 'efficac*' OR 'error*' OR 'perform*' OR 'reliability' OR 'validity' OR 'value' OR 'yield*') NEAR/4 'diagnostic*') OR 'receiver operat*')	2,251
#7	((#5 OR #6) AND 'osteoporosis'/dm AND 'article'/it) NOT [medline]/lim	1,016

B2 Detailed Eligibility Criteria

Category	Included	Excluded
Population	**KQs 1–3 (Screening benefits, accuracy, harms):** Adults age 40 years or older without known osteoporosis or history of fragility fractures **KQs 4, 5 (Treatment benefits and harms):** Adults age 40 years or older with osteoporosis, low bone mass, or increased fracture risk (as defined by study authors) *Studies in which less than 50% of the enrolled population includes persons with conditions or medications listed as excluded will be included, and results will be stratified if possible.* *Subpopulations of interest include men, women age 65 years or older, and postmenopausal women younger than 65 years.*[*]	**All KQs:** Studies that exclusively enroll adults younger than age 40 years **KQs 1–3:** Studies that exclusively enroll: • Adults with known osteoporosis or prior history of fragility fracture • Adults with cancer, metabolic bone diseases, or medical conditions associated with bone loss, including but not limited to hyperparathyroidism, premature ovarian failure, hypogonadism, untreated hyperthyroidism, acromegaly, adrenal insufficiency, Cushing's syndrome, celiac disease, inflammatory bowel disease, history of gastric bypass surgery, anorexia, chronic liver disease, multiple myeloma, chronic kidney disease, rheumatoid arthritis, lupus, multiple sclerosis, spinal cord injury • Adults taking chronic medications associated with bone loss or strengthening, including glucocorticosteroids, select antiepileptic medications, hypogonadism-inducing agents (e.g., aromatase inhibitors, medroxyprogesterone acetate, gonadotropin-releasing hormone agonists), thiazolidinediones, calcineurin inhibitors, antiretroviral therapy, and testosterone **KQs 4, 5:** Studies that exclusively enroll or in which the majority of the population has: • Secondary osteoporosis because of an underlying medical condition or chronic use of a medication associated with bone loss or • Prior fragility fracture In addition, studies that exclusively enroll participants who have failed prior medication use for osteoporosis are not eligible.
Screening Interventions	KQs 1–3 (Screening benefits, accuracy, and harms): • FRA or ORA that has been evaluated in at least two independent cohorts external to the development cohort (unless males are included, then only one independent cohort external to the development cohort is required) • DXA measurement of BMD at the femoral neck (T-scores based on NHANES III reference range) or lumbar spine (local reference range) • A combination of FRA or ORA and DXA together or in sequence (e.g., two-step approach)	• FRAs or ORAs that are not publicly available • Studies of FRAs or ORAs using split sample validation • Fall risk assessments (i.e., instruments validated to predict falls, not fractures) • FRAs or ORAs using risk factors not readily available or feasible within primary care settings • Quantitative ultrasound • Quantitative CT • Magnetic resonance imaging • Trabecular bone score • Vertebral fracture assessment • DXA measured at peripheral skeletal sites (e.g., radius, wrist, heel) • DXA measured at central skeletal sites, but hip T-scores based on local reference ranges • Bone turnover biomarkers • Finite element analysis • Hip structural analysis • Opportunistic screening for osteoporosis on images taken for other indications (e.g., dental X-rays, abdominal CT)

Appendix B. Additional Methods

Category	Included	Excluded
Screening Comparators	KQs 1, 3 (Screening benefits and harms): • No screening • FRA/ORA or BMD or both, but no results shared with patient or their primary care provider KQ 2 (Accuracy): • For predictive accuracy: Observed fracture incidence from nationally representative and verified sources • For diagnostic accuracy: DXA-measured BMD at the femoral neck (T-scores based on NHANES III reference range) or lumbar spine	KQs 1, 3: • No control group • Another screening strategy (active comparator) **KQ 2:** Any comparator not specifically identified as included
Treatment Interventions	**KQs 4, 5 (Treatment benefits and harms):** Bisphosphonates with FDA-approved indications for the treatment of osteoporosis (i.e., alendronate, ibandronate, risedronate, zoledronic acid), denosumab *Males only:* Teriparatide, abaloparatide, and romosozumab are also eligible†	**KQs 4, 5:** • Bisphosphonates that do not have FDA-approved indications for the treatment of osteoporosis (e.g., etidronate, pamidronate) • Estrogen (with or without progesterone), raloxifene, or bazedoxifene† • *Females only:* Teriparatide, abaloparatide, or romosozumab† • Medications that are sometimes used off-label to treat osteoporosis (e.g., testosterone, tamoxifen) • Treatments that are no longer used in practice or that have been recalled, specifically calcitonin and parathyroid hormone 1-84 • Vitamin D or calcium supplements alone (these are considered adjuncts to treatment) • Dietary supplements Nonpharmacologic treatments (e.g., exercise, fall prevention interventions)
Treatment Comparators	**KQs 4, 5 (Treatment benefits and harms):** Placebo, vitamin D or calcium or both, no treatment	• Active drug comparators (e.g., head-to-head comparisons of active drugs or comparisons of multiple drugs in combination or in sequence with monotherapy) • Nonpharmacologic interventions (e.g., exercise)

Appendix B. Additional Methods

Category	Included	Excluded
Outcomes	KQs 1, 4 (Screening and treatment benefits): • All-cause mortality • Fracture-related mortality • Fractures (all-cause, hip, major osteoporotic fractures‡, clinical vertebral fractures, any clinical fragility fractures) • Fracture-related morbidity (e.g., disability) KQ 2 (Accuracy): 1. Calibration outcomes (e.g., observed vs. Expected ratio, calibration slope, calibration plot, Hosmer-Lemeshow goodness-of-fit), overall prediction model performance (e.g., Brier score, explained variation [R^2]) 2. Discrimination outcomes (e.g., c-statistic, discrimination slope, sensitivity, specificity, area under the receiver operating characteristic curve) **KQ 3 (Screening harms):** 1. Overdiagnosis 1. Unnecessary treatment from inaccurate risk prediction 2. Radiation exposure • Anxiety from labeling **KQ 5 (Treatment harms):** • Total adverse events • Total serious adverse events • Specific serious adverse events: Major cardiovascular events (i.e., myocardial infarction, stroke, cardiovascular death), atrial fibrillation, osteonecrosis of the jaw, atypical femur fractures, incident gastrointestinal cancer, serious gastrointestinal events, rebound fractures after discontinuing denosumab treatment • Discontinuations because of adverse events	KQs 1, 4: • Radiographic (i.e., morphometric) vertebral fractures • Fractures based on patient self-report without verification/confirmation • BMD • Other outcomes not specifically identified as included **KQs 2, 3, 5:** Outcomes not specifically identified as included

Appendix B. Additional Methods

Category	Included	Excluded
Timing	KQs 1, 4 (Screening and treatment benefits): Followup for at least 1 year KQ 2 (Accuracy): • For predictive accuracy of FRAs or ORAs, observed fracture incidence over at least a median or mean of 80% of the time specified by the FRA (e.g., at least 8 years for a tool designed to predict 10-year risk). For FRAs or ORAs that do not specify a prediction interval, a minimum of 3 years of observed incidence is required • For predictive accuracy of DXA, observed fracture incidence over at least 1 year • For diagnostic accuracy of risk assessments, no longer than 8 weeks between FRA or ORA and BMD measurement KQs 3, 5 (Screening and treatment harms): Any length of followup	• Timing that does not meet inclusion criteria
Study Design	**KQs 1, 3 (Screening benefits and harms):** RCTs, clinical controlled trials, or systematic reviews of RCTs or controlled trials. Cohort studies and systematic reviews of cohort studies are also eligible for KQ 3 only. **KQ 2 (Accuracy):** Recent (published in the last 5 years) systematic reviews of cohort or test accuracy studies, cohort studies designed for evaluating predictive accuracy (i.e., prognosis for fracture risk) or diagnostic accuracy (for identification of osteoporosis), comparative studies in which a single group is treated as a cohort for purposes of evaluating predictive or diagnostic accuracy are also eligible **KQs 4, 5 (Treatment benefits and harms):** RCTs and controlled trials (including those in which participants serve as their own controls); controlled cohort studies are also eligible for KQ 5 only	**All KQs:** Case series; case reports; case-control studies; conference abstracts, posters, or proceedings without data or information available to assess risk of bias; unpublished data; editorials; commentaries; narrative reviews **KQs 4, and 5:** Systematic reviews are not eligible but will be hand searched to identify studies potentially missed by our search
Settings	**KQs 1, 3, 4, 5 (Screening and treatment benefits and harms):** Primary care settings in countries designated as "very high" on the 2020 Human Development Index (as defined by the United Nations Development Programme)[104] **KQ 2 (Accuracy):** Predictive accuracy: United States or countries with similar hip fracture incidence as the United States[§] for synthesis of any primary research studies	**KQs 1, 3, 4, 5:** Long-term care settings such as nursing homes, inpatient settings **KQs 1, 3:** Specialty medical settings (e.g., endocrinology, rheumatology) **KQ 2:** Predictive accuracy: studies in single countries with high or low fracture incidence

Appendix B. Additional Methods

Category	Included	Excluded
Study Quality	**KQ 1, 2c, 3, 4, 5:** Good or fair quality as determined by standard risk of bias instruments and existing USPSTF criteria tailored to study design **KQ 2a and KQ 2b:** Poor quality studies were also included	KQ 1, 2c, 3, 4, 5: Poor quality **KQ 2c:** Any study quality were allowed

* For the purposes of this review, we use the terms *men* and *women* consistent with how they are typically used in the underlying evidence base for this topic. *Men* refers to persons assigned male sex at birth. *Women* refers to persons assigned female sex at birth. Studies that include gender-diverse individuals, including those who have undergone gender-affirming therapy (e.g., transmen, transwomen), were not excluded from the scope of this review. However, studies that *exclusively* enrolled populations who take hormone therapy that affects bone density were excluded from this review, consistent with our criteria that exclude studies that focused on populations with secondary osteoporosis or who took chronic medications that have known effects on bone metabolism. For such populations, individualized clinical decisions about bone density testing in the context of condition and medication management are required.

† This review is not intended as a comprehensive review of all available pharmacologic therapies. Second-line therapies (abaloparatide, teriparatide, romosozumab) were excluded for women because the USPSTF is likely to have sufficient evidence to determine the net benefit of treatment based on the evidence for FDA-approved bisphosphonates and denosumab, as determined by the most recent review before this update. We only consider these drugs for men given the paucity of treatment studies generally available for men. Although romosozumab is not currently FDA approved for men, it is currently in Phase 3 studies for men, so it was included in this update. Hormone therapy and selective estrogen receptor modulators were reviewed in a separate USPSTF review on hormone therapy, so they were not included in this update.

‡ Major osteoporotic fracture is typically defined as fractures of the hip, wrist, and humerus and clinical vertebral fractures.

§ Countries with “moderate” hip fracture incidence in addition to the United States include Australia, Canada, Chile, Estonia, Finland, France, Israel, Japan, Kuwait, Lithuania, Malaysia, Mexico, the Netherlands, New Zealand, Poland, Portugal, Russia, South Korea, Spain, and Thailand.[105]

Abbreviations: BMD=bone mineral density; CT=computerized tomography; DXA=dual-energy X-ray absorptiometry; FDA=Food and Drug Administration; FRA=fracture risk assessment; KQ=key question; NHANES=National Health and Nutrition Examination Survey; ORA=osteoporosis risk assessment; RCT=randomized, controlled trial; USPSTF=U.S. Preventive Services Task Force.

B.3 U.S. Preventive Services Task Force Quality Rating Criteria

Criteria for Randomized, Controlled Trials and Cohort Studies

- Initial assembly of comparable groups
- Randomized, controlled trials (RCTs)—adequate randomization, including concealment and whether potential confounders were distributed equally among groups; cohort studies—consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts
- Maintenance of comparable groups (includes attrition, crossovers, adherence, and contamination)
- Important differential loss to followup or overall high loss to followup
- Measurements that are equal, reliable, and valid (includes masking of outcome assessment)
- Clear definition of interventions
- Important outcomes considered
- Analysis: Adjustment for potential confounders for cohort studies or intention-to-treat analysis for RCTs; for cluster RCTs, correction for correlation coefficient

Definition of Ratings Based on Above Criteria Randomized, Controlled Trials and Cohort Studies

Good: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (followup ≥80%); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; important outcomes are considered; and appropriate attention is given to confounders in analysis. In addition, intention-to-treat analysis is used for RCTs.

Fair: Studies will be graded “fair” if any or all of the following problems occur, without the important limitations noted in the “poor” category below: Generally comparable groups are assembled initially, but some question remains on whether some (although not major) differences occurred in followup; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is lacking for RCTs.

Poor: Studies will be graded “poor” if any of the following major limitations exist: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. Intention-to-treat analysis is lacking for RCTs.

Criteria for Systematic Reviews

- Comprehensiveness of sources considered/search strategy used
- Standard appraisal of included studies
- Validity of conclusions
- Recency and relevance (especially important for systematic reviews)

Definition of Ratings Based on Above Criteria for Systematic Reviews

Good: Recent, relevant review with comprehensive sources and search strategies; explicit and relevant selection criteria; standard appraisal of included studies; and valid conclusions

Fair: Recent, relevant review that is not clearly biased but lacks comprehensive sources and search strategies

Poor: Outdated, irrelevant, or biased review without systematic search for studies, explicit selection criteria, or standard appraisal of studies

Sources: U.S. Preventive Services Task Force, Procedure Manual, Appendix VI. 2017[420]; Harris et al, 2001.[421]

Appendix C. Excluded Studies

List of Exclusion Codes:

X1 Not published in English or ineligible publication type
X2 Ineligible population
X3 Ineligible study design or timing
X4 Ineligible geographic setting (except non very high HDI)
X5 Ineligible or no intervention
X6 Ineligible or no comparator
X7 Ineligible or no outcome
X8 Not in very high HDI country
X9 Study superseded by new evidence or duplicate or covered by included SR
X10 Poor quality

1. What evidence for bisphosphonate drug holidays? *Drug Ther Bull*. 2020 Jun;58(6):88. doi: 10.1136/dtb.2020.000017. PMID: 32188686. Exclusion Code: X5.
2. Abrahamsen B, Pazianas M, Eiken P, et al. Esophageal and gastric cancer incidence and mortality in alendronate users. *J Bone Miner Res*. 2012 Mar;27(3):679-86. doi: 10.1002/jbmr.1481. PMID: 22113985. Exclusion Code: X10.
3. Adachi JD, Bone HG, Daizadeh NS, et al. Influence of subject discontinuation on long-term nonvertebral fracture rate in the denosumab FREEDOM Extension study. *BMC Musculoskelet Disord*. 2017 Apr 27;18(1):174. doi: 10.1186/s12891-017-1520-6. PMID: 28449657. Exclusion Code: X6.
4. Adami G, Arioli G, Bianchi G, et al. Radiofrequency echographic multi spectrometry for the prediction of incident fragility fractures: a 5-year follow-up study. *Bone*. 2020 May;134:115297. doi: 10.1016/j.bone.2020.115297. PMID: 32092480. Exclusion Code: X5.
5. Agarwal A, Baleanu F, Moreau M, et al. External validation of FRISBEE 5-year fracture prediction models: a registry-based cohort study. *Arch Osteoporos*. 2022 Dec 23;18(1):13. doi: 10.1007/s11657-022-01205-7. PMID: 36564674. Exclusion Code: X3.
6. Agarwal A, Leslie WD, Nguyen TV, et al. Performance of the Garvan Fracture Risk Calculator in individuals with diabetes: a registry-based cohort study. *Calcif Tissue Int*. 2022 Jun;110(6):658-65. doi: 10.1007/s00223-021-00941-1. PMID: 34994831. Exclusion Code: X9.
7. Ahadzadeh Ardebili A, Fu T, Dunnewold N, et al. Bisphosphonates preserve bone mineral density and suppress bone turnover markers in early menopausal women: a systematic review and meta-analysis of randomized trials. *JBMR Plus*. 2023 Jun;7(6):e10748. doi: 10.1002/jbm4.10748. PMID: 37283657. Exclusion Code: X2.
8. Ahmed LA, Nguyen ND, Bjornerem A, et al. External validation of the Garvan nomograms for predicting absolute fracture risk: the Tromso study. *PLoS One*. 2014;9(9):e107695. doi: 10.1371/journal.pone.0107695. PMID: 25255221. Exclusion Code: X4.
9. Ahmed LA, Schirmer H, Fonnebo V, et al. Validation of the Cummings' risk score; how well does it identify women with high risk of hip fracture: the Tromso Study. *Eur J Epidemiol*. 2006;21(11):815-22. doi: 10.1007/s10654-006-9072-3. PMID: 17119878. Exclusion Code: X7.
10. Aktas I, Nazikoglu C, Kepez A, et al. Effect of intravenous zoledronic acid infusion on electrocardiographic parameters in patients with osteoporosis. *Osteoporos Int*. 2016 Dec;27(12):3543-7. doi: 10.1007/s00198-016-3684-6. PMID: 27344642. Exclusion Code: X7.

11. Akyea RK, McKeever TM, Gibson J, et al. Predicting fracture risk in patients with chronic obstructive pulmonary disease: a UK-based population-based cohort study. *BMJ Open*. 2019 Apr 3;9(4):e024951. doi: 10.1136/bmjopen-2018-024951. PMID: 30948576. Exclusion Code: X2.

12. Albert SG, Wood E. Meta-analysis of clinical fracture risk reduction of anti-osteoporosis drugs: direct and indirect comparisons and meta-regressions. *Endocr Pract*. 2021 Jul 9doi: 10.1016/j.eprac.2021.06.015. PMID: 34252583. Exclusion Code: X3.

13. Albertsson D, Mellstrom D, Petersson C, et al. Hip and fragility fracture prediction by 4-item clinical risk score and mobile heel BMD: a women cohort study. *BMC Musculoskelet Disord*. 2010 Mar 24;11:55. doi: 10.1186/1471-2474-11-55. PMID: 20334634. Exclusion Code: X7.

14. Albertsson DM, Mellstrom D, Petersson C, Eggertsen R. Validation of a 4-item score predicting hip fracture and mortality risk among elderly women. *Ann Fam Med*. 2007 Jan-Feb;5(1):48-56. doi: 10.1370/afm.602. PMID: 17261864. Exclusion Code: X7.

15. Alibasic E, Ljuca F, Brkic S, et al. Secondary prevention of osteoporosis through assessment of individual and multiple risk factors. *Mater Sociomed*. 2020 Mar;32(1):10-4. doi: 10.5455/msm.2020.32.10-14. PMID: 32410886. Exclusion Code: X7.

16. Alshahrani F, Alsulaiman YA, Almashari YM, et al. Use of the Osteoporosis Self-Assessment Tool as a screening tool for osteoporosis in Saudi postmenopausal women. *Cureus*. 2023 Apr;15(4):e37755. doi: 10.7759/cureus.37755. PMID: 37213999. Exclusion Code: X7.

17. Amith N, Augustine S, Raj SPS, et al. Evaluation of effectiveness of MORES to identify men at risk of osteoporosis. *European Journal of Molecular and Clinical Medicine*. 2022;9(7):5772-9. Exclusion Code: X8.

18. Anagnostis P, Paschou SA, Gkekas NN, et al. Efficacy of anti-osteoporotic medications in patients with type 1 and 2 diabetes mellitus: a systematic review. *Endocrine*. 2018 Jun;60(3):373-83. doi: 10.1007/s12020-018-1548-x. PMID: 29411304. Exclusion Code: X2.

19. Anastasilakis AD, Papapoulos SE, Polyzos SA, et al. Zoledronate for the prevention of bone loss in women discontinuing denosumab treatment. a prospective 2-year clinical trial. *J Bone Miner Res*. 2019 Dec;34(12):2220-8. doi: 10.1002/jbmr.3853. PMID: 31433518. Exclusion Code: X6.

20. Ang SB, Xia JY, Cheng SJ, et al. A pilot screening study for low bone mass in Singaporean women using years since menopause and BMI. *Climacteric*. 2021 Apr 30:1-7. doi: 10.1080/13697137.2021.1908989. PMID: 33928868. Exclusion Code: X6.

21. Asghar ZB, Godoy Caballero A, Pathirannehelage S, et al. Saving bones without risking brain-bisphosphonates and risk of stroke: matched case-control study. *Osteoporos Int*. 2019 Sep;30(9):1845-54. doi: 10.1007/s00198-019-05045-z. PMID: 31214750. Exclusion Code: X3.

22. Ashok Kumar D, Anburajan M, Snekhalatha U. Evaluation of low bone mass and prediction of fracture risk using metacarpal radiogrammetry method: a comparative study with DXA and X-ray phantom. *Int J Rheum Dis*. 2018 Jul;21(7):1350-71. doi: 10.1111/1756-185x.13326. PMID: 29968333. Exclusion Code: X8.

23. Asirvatham AR, Balachandran K, Kannan S, et al. FRAX first - Pragmatic approach in resource poor settings. *Indian J Endocrinol Metab*. 2018 Nov-Dec;22(6):757-9. doi: 10.4103/ijem.IJEM_412_18. PMID: 30766813. Exclusion Code: X7.

24. Azagra R, Roca G, Encabo G, et al. FRAX(R) tool, the WHO algorithm to predict osteoporotic fractures: the first analysis of its discriminative and predictive ability in the Spanish FRIDEX cohort. *BMC Musculoskelet Disord*. 2012;13:204. doi: 10.1186/1471-2474-13-204. PMID: 23088223. Exclusion Code: X9.

25. Baek KH, Chung YS, Koh JM, et al. Romosozumab in postmenopausal Korean women with osteoporosis: a randomized, double-blind, placebo-controlled efficacy and safety study. *Endocrinol Metab (Seoul)*. 2021 Feb;36(1):60-9. doi: 10.3803/EnM.2020.848. PMID: 33677928. Exclusion Code: X5.

26. Bager CL, Bay F, Christiansen C, Karsdal M. Low bone turnover levels predict increased risk of cancer. *Bone*. 2019 Oct;127:75-81. doi: 10.1016/j.bone.2019.05.032. PMID: 31150870. Exclusion Code: X7.

27. Bala Y, Chapurlat R, Cheung AM, et al. Risedronate slows or partly reverses cortical and trabecular microarchitectural deterioration in postmenopausal women. *J Bone Miner Res*. 2014 Feb;29(2):380-8. doi: 10.1002/jbmr.2101 [doi]. PMID: 24115129. Exclusion Code: X10.

28. Bansal B, Mithal A, Chopra SR, et al. Judicious use of DXA-BMD in assessing fracture risk by using clinical risk factors in the Indian population. *Arch Osteoporos*. 2018 Oct 29;13(1):115. doi: 10.1007/s11657-018-0536-3. PMID: 30374781. Exclusion Code: X3.

29. Barrett-Connor E, Cauley JA, Kulkarni PM, et al. Risk-benefit profile for raloxifene: 4-year data From the Multiple Outcomes of Raloxifene Evaluation (MORE) randomized trial. *J Bone Miner Res*. 2004 Aug;19(8):1270-5. doi: 10.1359/JBMR.040406 [doi]. PMID: 15231013. Exclusion Code: X5.

30. Barrett-Connor E, Grady D, Sashegyi A, et al. Raloxifene and cardiovascular events in osteoporotic postmenopausal women: four-year results from the MORE (Multiple Outcomes of Raloxifene Evaluation) randomized trial. *JAMA*. 2002 Feb 20;287(7):847-57. doi: joc11015 [pii]. PMID: 11851576. Exclusion Code: X5.

31. Barrionuevo P, Kapoor E, Asi N, et al. Efficacy of pharmacological therapies for the prevention of fractures in postmenopausal women: a network meta-analysis. *J Clin Endocrinol Metab*. 2019 May 1;104(5):1623-30. doi: 10.1210/jc.2019-00192. PMID: 30907957. Exclusion Code: X3.

32. Bartosch P, Malmgren L. Can frailty in conjunction with FRAX identify additional women at risk of fracture - a longitudinal cohort study of community dwelling older women. *BMC Geriatr*. 2022 Dec 9;22(1):951. doi: 10.1186/s12877-022-03639-7. PMID: 36494774. Exclusion Code: X4.

33. Bastounis A, Langley T, Davis S, et al. Assessing the Effectiveness of Bisphosphonates for the Prevention of Fragility Fractures: An Updated Systematic Review and Network Meta-Analyses. *JBMR Plus*. 2022 May;6(5):e10620. doi: 10.1002/jbm4.10620. PMID: 35509636. Exclusion Code: X3.

34. Bauer DC, Black DM, Bouxsein ML, et al. Treatment-related changes in bone turnover and fracture risk reduction in clinical trials of anti-resorptive drugs: a meta-regression. *J Bone Miner Res*. 2018 Apr;33(4):634-42. doi: 10.1002/jbmr.3355. PMID: 29318649. Exclusion Code: X3.

35. Bauer DC, Ewing SK, Cauley JA, et al. Quantitative ultrasound predicts hip and non-spine fracture in men: the MrOS study. *Osteoporos Int*. 2007 Jun;18(6):771-7. doi: 10.1007/s00198-006-0317-5 [doi]. PMID: 17273893. Exclusion Code: X5.

36. Beattie JR, Cummins NM, Caraher C, et al. Raman spectroscopic analysis of fingernail clippings can help differentiate between postmenopausal women who have and have not suffered a fracture. *Clin Med Insights: Arthritis Musculoskelet Disord*. 2016;9:109-16. doi: 10.4137/cmamd.S38493. PMID: 27429561. Exclusion Code: X5.

37. Beaudart C, Demonceau C, Sabico S, et al. Efficacy of osteoporosis pharmacological treatments in men: a systematic review and meta-analysis. *Aging Clin Exp Res*. 2023 Sep;35(9):1789-806. doi: 10.1007/s40520-023-02478-9. PMID: 37400668. Exclusion Code: X3.

38. Beaudoin C, Jean S, Moore L, et al. Number, location, and time since prior fracture as predictors of future fracture in the elderly from the general population. *J Bone Miner Res*. 2018 Nov;33(11):1956-66. doi: 10.1002/jbmr.3526. PMID: 29924429. Exclusion Code: X3.

39. Behanova M, Reichardt B, Stamm TA, et al. Treatment effects of bisphosphonates and denosumab on survival and refracture from real-world data of hip-fractured patients. *Calcif Tissue Int*. 2019 Dec;105(6):630-41. doi: 10.1007/s00223-019-00611-3. PMID: 31531720. Exclusion Code: X2.

40. Bell KJ, Hayen A, Glasziou P, et al. Potential usefulness of BMD and bone turnover monitoring of zoledronic acid therapy among women with osteoporosis: secondary analysis of randomized controlled trial data. *J Bone Miner Res*. 2016 Sep;31(9):1767-73. doi: 10.1002/jbmr.2847. PMID: 27027655. Exclusion Code: X7.

41. Berry ME. Preventing and treating osteoporosis. *Radiol Technol*. 2019 Jan;90(3):286-93. PMID: 30635461. Exclusion Code: X1.

42. Berry SD, Kiel DP, Colón-Emeric C. Hip fractures in older adults in 2019. *JAMA*. 2019 Jun 11;321(22):2231-2. doi: 10.1001/jama.2019.5453. PMID: 31074763. Exclusion Code: X1.

43. Bhandari M, Jin L, See K, et al. Does teriparatide improve femoral neck fracture healing: results from a randomized placebo-controlled trial. *Clin Orthop Relat Res*. 2016 May;474(5):1234-44. doi: 10.1007/s11999-015-4669-z. PMID: 26932738. Exclusion Code: X2.

44. Bhat KA, Kakaji M, Awasthi A, et al. Utility of osteoporosis self-assessment tool as a screening tool for predicting osteoporosis in Indian men. *J Clin Densitom*. 2017 Apr-Jun;20(2):160-3. doi: 10.1016/j.jocd.2016.04.005. PMID: 27210803. Exclusion Code: X8.

45. Bilek LD, Waltman NL, Lappe JM, et al. Protocol for a randomized controlled trial to compare bone-loading exercises with risedronate for preventing bone loss in osteopenic postmenopausal women. *BMC Womens Health*. 2016 Aug 30;16(1):59. doi: 10.1186/s12905-016-0339-x. PMID: 27576310. Exclusion Code: X7.

46. Bilezikian JP, Lin CJF, Brown JP, et al. Long-term denosumab treatment restores cortical bone loss and reduces fracture risk at the forearm and humerus: analyses from the FREEDOM Extension cross-over group. *Osteoporos Int*. 2019 Sep;30(9):1855-64. doi: 10.1007/s00198-019-05020-8. PMID: 31201481. Exclusion Code: X6.

47. Binkley N, Silverman SL, Simonelli C, et al. Monthly ibandronate suppresses serum CTX-I within 3 days and maintains a monthly fluctuating pattern of suppression. *Osteoporos Int*. 2009 Sep;20(9):1595-601. doi: 10.1007/s00198-008-0827-4. PMID: 19145396. Exclusion Code: X10.

48. Biver E, Durosier-Izart C, Chevalley T, et al. Evaluation of radius microstructure and areal bone mineral density improves fracture prediction in postmenopausal women. *J Bone Miner Res*. 2018 Feb;33(2):328-37. doi: 10.1002/jbmr.3299. PMID: 28960489. Exclusion Code: X3.

49. Black DM, Bauer DC, Vittinghoff E, et al. Treatment-related changes in bone mineral density as a surrogate biomarker for fracture risk reduction: meta-regression analyses of individual patient data from multiple randomised controlled trials. *Lancet Diabetes Endocrinol*. 2020 Aug;8(8):672-82. doi: 10.1016/s2213-8587(20)30159-5. PMID: 32707115. Exclusion Code: X3.

50. Black DM, Condra K, Adams AL, Eastell R. Bisphosphonates and the risk of atypical femur fractures. *Bone*. 2022 Mar;156:116297. doi: 10.1016/j.bone.2021.116297. PMID: 34920168. Exclusion Code: X1.

51. Black DM, Geiger EJ, Eastell R, et al. Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. *N Engl J Med*. 2020 Aug 20;383(8):743-53. doi: 10.1056/NEJMoa1916525. PMID: 32813950. Exclusion Code: X10.

52. Black DM, Schwartz AV, Ensrud KE, et al. Effects of continuing or stopping alendronate after 5 years of treatment: the Fracture Intervention Trial Long-term Extension (FLEX): a randomized trial. *JAMA*. 2006 Dec 27;296(24):2927-38. doi: 10.1001/jama.296.24.2927. PMID: 17190893. Exclusion Code: X2.

53. Bliuc D, Tran T, van Geel T, et al. Reduced bone loss is associated with reduced mortality risk in subjects exposed to nitrogen bisphosphonates: a mediation analysis. *J Bone Miner Res*. 2019 Nov;34(11):2001-11. doi: 10.1002/jbmr.3816. PMID: 31402509. Exclusion Code: X7.

54. Bliuc D, Tran T, van Geel T, et al. Mortality risk reduction differs according to bisphosphonate class: a 15-year observational study. *Osteoporos Int*. 2019 Apr;30(4):817-28. doi: 10.1007/s00198-018-4806-0. PMID: 30607457. Exclusion Code: X3.

55. Bolton JM, Morin SN, Majumdar SR, et al. Association of mental disorders and related medication use with risk for major osteoporotic fractures. *JAMA Psychiatry*. 2017 Jun 1;74(6):641-8. doi: 10.1001/jamapsychiatry.2017.0449. PMID: 28423154. Exclusion Code: X7.

56. Bonaccorsi G, Fila E, Cervellati C, et al. Assessment of fracture risk in a population of postmenopausal Italian women: a comparison of two different tools. *Calcif Tissue Int*. 2015 Jul;97(1):50-7. doi: 10.1007/s00223-015-0009-2. PMID: 25939647. Exclusion Code: X4.

57. Bone HG, Greenspan SL, McKeever C, et al. Alendronate and estrogen effects in postmenopausal women with low bone mineral density. Alendronate/Estrogen Study Group. *J Clin Endocrinol Metab*. 2000 Feb;85(2):720-6. doi: 10.1210/jcem.85.2.6393. PMID: 10690882. Exclusion Code: X5.

58. Bone HG, Wagman RB, Brandi ML, et al. 10 years of denosumab treatment in postmenopausal women with osteoporosis: results from the phase 3 randomised FREEDOM trial and open-label extension. *Lancet Diabetes Endocrinol*. 2017 Jul;5(7):513-23. doi: 10.1016/s2213-8587(17)30138-9. PMID: 28546097. Exclusion Code: X6.

59. Bovijn J, Krebs K, Chen CY, et al. Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. *Sci Transl Med*. 2020 Jun 24;12(549)doi: 10.1126/scitranslmed.aay6570. PMID: 32581134. Exclusion Code: X3.

60. Bregigeon S, Galinier A, Zaegel-Faucher O, et al. Frailty in HIV infected people: a new risk factor for bone mineral density loss. *AIDS*. 2017 Jul 17;31(11):1573-7. doi: 10.1097/qad.0000000000001507. PMID: 28426535. Exclusion Code: X2.

61. Briot K, Paternotte S, Kolta S, et al. FRAX®: prediction of major osteoporotic fractures in women from the general population: the OPUS study. *PLoS One*. 2013;8(12):e83436. doi: 10.1371/journal.pone.0083436. PMID: 24386199. Exclusion Code: X9.

62. Brozek W, Reichardt B, Zwerina J, et al. Antiresorptive therapy and risk of mortality and refracture in osteoporosis-related hip fracture: a nationwide study. *Osteoporos Int*. 2016 Jan;27(1):387-96. doi: 10.1007/s00198-015-3415-4. PMID: 26576544. Exclusion Code: X2.

63. Buehring B, Hansen KE, Lewis BL, et al. Dysmobility syndrome independently increases fracture risk in the osteoporotic fractures in men (MrOS) prospective cohort study. *J Bone Miner Res*. 2018 Sep;33(9):1622-9. doi: 10.1002/jbmr.3455. PMID: 29701911. Exclusion Code: X9.

64. Byun JH, Jang S, Lee S, et al. The efficacy of bisphosphonates for prevention of osteoporotic fracture: an update meta-analysis. *J Bone Metab*. 2017 Feb;24(1):37-49. doi: 10.11005/jbm.2017.24.1.37. PMID: 28326300. Exclusion Code: X3.

65. Cai G, Keen HI, Host LV, et al. Once-yearly zoledronic acid and change in abdominal aortic calcification over 3 years in postmenopausal women with osteoporosis: results from the HORIZON Pivotal Fracture Trial. *Osteoporos Int*. 2020 Sep;31(9):1741-7. doi: 10.1007/s00198-020-05430-z. PMID: 32361951. Exclusion Code: X7.

66. Caillet P, Klemm S, Ducher M, et al. Hip fracture in the elderly: a re-analysis of the EPIDOS study with causal Bayesian networks. *PLoS One*. 2015;10(3):e0120125. doi: 10.1371/journal.pone.0120125. PMID: 25822373. Exclusion Code: X7.

67. Cappelle SI, Ramon I, Dekelver C, et al. Distribution of clinical risk factors for fracture in a Brussels cohort of postmenopausal women: The FRISBEE study and comparison with other major cohort studies. *Maturitas*. 2017 Dec;106:1-7. doi: 10.1016/j.maturitas.2017.08.010. PMID: 29150162. Exclusion Code: X7.

68. Cardwell CR, Abnet CC, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of esophageal cancer. *JAMA*. 2010 Aug 11;304(6):657-63. doi: 10.1001/jama.2010.1098. PMID: 20699457. Exclusion Code: X10.

69. Carvajal A, Martín Arias LH, Sáinz M, et al. Carpal tunnel syndrome associated with oral bisphosphonates. a population-based cohort study. *PLoS One*. 2016;11(1):e0146772. doi: 10.1371/journal.pone.0146772. PMID: 26765346. Exclusion Code: X7.

70. Cass AR, Shepherd AJ, Asirot R, et al. Comparison of the Male Osteoporosis Risk Estimation Score (MORES) with FRAX in identifying men at risk for osteoporosis. *Ann Fam Med*. 2016 Jul;14(4):365-9. doi: 10.1370/afm.1945. PMID: 27401426. Exclusion Code: X9.

71. Cauley JA, Norton L, Lippman ME, et al. Continued breast cancer risk reduction in postmenopausal women treated with raloxifene: 4-year results from the MORE trial. Multiple outcomes of raloxifene evaluation. *Breast Cancer Res Treat*. 2001 Jan;65(2):125-34. doi: 10.1023/a:1006478317173. PMID: 11261828. Exclusion Code: X5.

72. Cha Y, Kim JT, Kim JW, et al. Effect of artificial intelligence or machine learning on prediction of hip fracture risk: systematic review. *J Bone Metab*. 2023 Aug;30(3):245-52. doi: 10.11005/jbm.2023.30.3.245. PMID: 37718902. Exclusion Code: X5.

73. Chalitsios CV, Shaw DE, McKeever TM. Risk of subtrochanteric and femoral shaft fractures due to bisphosphonate therapy in asthma: a population-based nested case-control study. *Osteoporos Int*. 2022 Apr;33(4):931-5. doi: 10.1007/s00198-021-06197-7. PMID: 34635953. Exclusion Code: X3.

74. Chambers JD, Winn A, Zhong Y, et al. Potential role of network meta-analysis in value-based insurance design. *Am J Manag Care*. 2014 Aug;20(8):641-8. PMID: 25295678. Exclusion Code: X3.

75. Champakanath A, Keshawarz A, Pyle L, et al. Fracture risk assessment (FRAX) without BMD and risk of major osteoporotic fractures in adults with type 1 diabetes. *Bone*. 2021 Feb;143:115614. doi: 10.1016/j.bone.2020.115614. PMID: 32853853. Exclusion Code: X7.

76. Chan MY, Nguyen ND, Center JR, et al. Absolute fracture-risk prediction by a combination of calcaneal quantitative ultrasound and bone mineral density. *Calcif Tissue Int*. 2012 Feb;90(2):128-36. doi: 10.1007/s00223-011-9556-3 [doi]. PMID: 22179560. Exclusion Code: X5.

77. Chandran M. Fracture risk assessment in clinical practice: why do it? what to do it with? . *J Clin Densitom*. 2017 Jul-Sep;20(3):274-9. doi: 10.1016/j.jocd.2017.06.010. PMID: 28711299. Exclusion Code: X1.

78. Chandran M, Chin YA, Choo KS, et al. Comparison of the osteoporosis self-assessment tool for Asians and the fracture risk assessment tool - FRAX to identify densitometric defined osteoporosis: a discriminatory value analysis in a multi-ethnic female population in Southeast Asia. *Osteoporos Sarcopenia*. 2020 Jun;6(2):53-8. doi: 10.1016/j.afos.2020.04.001. PMID: 32715094. Exclusion Code: X6.

79. Chanplakorn P, Lertudomphonwanit T, Daraphongsataporn N, et al. Development of prediction model for osteoporotic vertebral compression fracture screening without using clinical risk factors, compared with FRAX and other previous models. *Arch Osteoporos*. 2021 Jun 3;16(1):84. doi: 10.1007/s11657-021-00957-y. PMID: 34085107. Exclusion Code: X8.

80. Chao AS, Chen FP, Lin YC, et al. Application of the World Health Organization Fracture Risk Assessment Tool to predict need for dual-energy X-ray absorptiometry scanning in postmenopausal women. *Taiwan J Obstet Gynecol*. 2015 Dec;54(6):722-5. doi: 10.1016/j.tjog.2015.10.005. PMID: 26700992. Exclusion Code: X10.

81. Chavassieux P, Portero-Muzy N, Roux JP, et al. Reduction of cortical bone turnover and erosion depth after 2 and 3 years of denosumab: iliac bone histomorphometry in the FREEDOM Trial. *J Bone Miner Res*. 2019 Apr;34(4):626-31. doi: 10.1002/jbmr.3631. PMID: 30601581. Exclusion Code: X7.

82. Chen L, Liu XP, Zhou B, et al. Relationship between change in bone mineral density of lumbar spine and risk of new vertebral and nonvertebral fractures: a meta-analysis. *Orthop Surg*. 2022 Feb;14(2):199-206. doi: 10.1111/os.13184. PMID: 34985187. Exclusion Code: X3.

83. Chen L, Wang G, Zheng F, et al. Efficacy of bisphosphonates against osteoporosis in adult men: a meta-analysis of randomized controlled trials. *Osteoporos Int*. 2015 Sep;26(9):2355-63. doi: 10.1007/s00198-015-3148-4. PMID: 25944731. Exclusion Code: X3.

84. Chen W, Simpson JM, March LM, et al. Comorbidities only account for a small proportion of excess mortality after fracture: a record linkage study of individual fracture types. *J Bone Miner Res*. 2018 May;33(5):795-802. doi: 10.1002/jbmr.3374. PMID: 29314242. Exclusion Code: X3.

85. Chen Y, Liu R, Hettinghouse A, et al. Clinical application of teriparatide in fracture prevention: a systematic review. *JBJS Rev*. 2019 Jan;7(1):e10. doi: 10.2106/jbjs.Rvw.18.00052. PMID: 30694878. Exclusion Code: X3.

86. Cherian KE, Kapoor N, Shetty S, et al. Evaluation of different screening tools for predicting femoral neck osteoporosis in rural south Indian postmenopausal women. *J Clin Densitom*. 2018 Jan-Mar;21(1):119-24. doi: 10.1016/j.jocd.2017.08.002. PMID: 28958825. Exclusion Code: X8.

87. Chiang CH, Huang CC, Chan WL, et al. Oral alendronate use and risk of cancer in postmenopausal women with osteoporosis: a nationwide study. *J Bone Miner Res*. 2012 Sep;27(9):1951-8. doi: 10.1002/jbmr.1645. PMID: 22532232. Exclusion Code: X10.

88. Chiu WY, Yang WS, Chien JY, et al. The influence of alendronate and tooth extraction on the incidence of osteonecrosis of the jaw among osteoporotic subjects. *PLoS One*. 2018;13(4):e0196419. doi: 10.1371/journal.pone.0196419. PMID: 29694412. Exclusion Code: X6.

89. Choi D, Choi S, Chang J, Park SM. Exposure to oral bisphosphonates and risk of gastrointestinal cancer. *Osteoporos Int*. 2020 Apr;31(4):775-82. doi: 10.1007/s00198-020-05327-x. PMID: 32034451. Exclusion Code: X10.

90. Chotiyarnwong P, McCloskey E, Eastell R, et al. A pooled analysis of fall incidence from placebo-controlled trials of denosumab. *J Bone Miner Res*. 2020 Jun;35(6):1014-21. doi: 10.1002/jbmr.3972. PMID: 31999376. Exclusion Code: X3.

91. Cipriani C, Castro C, Curione M, et al. Acute effect of zoledronic acid on the risk of cardiac dysrhythmias. *Intern Emerg Med*. 2015 Mar;10(2):151-6. doi: 10.1007/s11739-014-1125-0. PMID: 25190623. Exclusion Code: X7.

92. Cipriani C, Pepe J, Bertoldo F, et al. The epidemiology of osteoporosis in Italian postmenopausal women according to the National Bone Health Alliance (NBHA) diagnostic criteria: a multicenter cohort study. *J Endocrinol Invest*. 2018 Apr;41(4):431-8. doi: 10.1007/s40618-017-0761-4. PMID: 28956296. Exclusion Code: X7.

93. Cohen A, Shiau S, Nair N, et al. Effect of teriparatide on bone remodeling and density in premenopausal idiopathic osteoporosis: a phase II trial. *J Clin Endocrinol Metab*. 2020 Oct 1;105(10)doi: 10.1210/clinem/dgaa489. PMID: 32876328. Exclusion Code: X2.

94. Condurache CI, Chiu S, Chotiyarnwong P, et al. Screening for high hip fracture risk does not impact on falls risk: a post hoc analysis from the SCOOP study. *Osteoporos Int*. 2020 Mar;31(3):457-64. doi: 10.1007/s00198-019-05270-6. PMID: 31960099. Exclusion Code: X7.

95. Cortet B, Modi A, Tang J, et al. Association between gastrointestinal events and osteoporosis treatment initiation in women diagnosed with osteoporosis in France: a retrospective analysis. *BMC Musculoskelet Disord*. 2016 Apr 30;17:195. doi: 10.1186/s12891-016-1041-8. PMID: 27139225. Exclusion Code: X6.

96. Cosman F, Crittenden DB, Adachi JD, et al. Romosozumab treatment in postmenopausal women with osteoporosis. *N Engl J Med*. 2016 Oct 20;375(16):1532-43. doi: 10.1056/NEJMoa1607948. PMID: 27641143. Exclusion Code: X5.

97. Cosman F, Crittenden DB, Ferrari S, et al. FRAME study: the foundation effect of building bone with 1 year of romosozumab leads to continued lower fracture risk after transition to denosumab. *J Bone Miner Res*. 2018 Jul;33(7):1219-26. doi: 10.1002/jbmr.3427. PMID: 29573473. Exclusion Code: X3.

98. Cosman F, Hattersley G, Hu MY, et al. Effects of abaloparatide-SC on fractures and bone mineral density in subgroups of postmenopausal women with osteoporosis and varying baseline risk factors. *J Bone Miner Res*. 2017 Jan;32(1):17-23. doi: 10.1002/jbmr.2991. PMID: 27612281. Exclusion Code: X2.

99. Cosman F, Miller PD, Williams GC, et al. Eighteen months of treatment with subcutaneous abaloparatide followed by 6 months of treatment with alendronate in postmenopausal women with osteoporosis: results of the ACTIVExtend Trial. *Mayo Clin Proc*. 2017 Feb;92(2):200-10. doi: 10.1016/j.mayocp.2016.10.009. PMID: 28160873. Exclusion Code: X6.

100. Cosman F, Peterson LR, Towler DA, et al. Cardiovascular safety of abaloparatide in postmenopausal women with osteoporosis: analysis from the ACTIVE phase 3 trial. *J Clin Endocrinol Metab*. 2020 Nov 1;105(11):3384-95. doi: 10.1210/clinem/dgaa450. PMID: 32658264. Exclusion Code: X5.

101. Crandall CJ. Risk assessment tools for osteoporosis screening in postmenopausal women: a systematic review. *Curr Osteoporos Rep*. 2015 Oct;13(5):287-301. doi: 10.1007/s11914-015-0282-z. PMID: 26233285. Exclusion Code: X10.

102. Cronholm F, Rosengren BE, Nilsson J, et al. The fracture predictive ability of a musculoskeletal composite score in old men - data from the MrOs Sweden study. *BMC Geriatr*. 2019 Mar 22;19(1):90. doi: 10.1186/s12877-019-1106-2. PMID: 30902044. Exclusion Code: X3.

103. Cummings SR, Ferrari S, Eastell R, et al. Vertebral fractures after discontinuation of denosumab: a post hoc analysis of the randomized placebo-controlled FREEDOM trial and its extension. *J Bone Miner Res*. 2018 Feb;33(2):190-8. doi: 10.1002/jbmr.3337. PMID: 29105841. Exclusion Code: X3.

104. Cummins NM, Poku EK, Towler MR, et al. Clinical risk factors for osteoporosis in Ireland and the UK: a comparison of FRAX and QFractureScores. *Calcif Tissue Int*. 2011 Aug;89(2):172-7. doi: 10.1007/s00223-011-9504-2 [doi]. PMID: 21647704. Exclusion Code: X9.

105. Curhan SG, Stankovic K, Halpin C, et al. Osteoporosis, bisphosphonate use, and risk of moderate or worse hearing loss in women. *J Am Geriatr Soc*. 2021 May 24doi: 10.1111/jgs.17275. PMID: 34028002. Exclusion Code: X7.

106. Curtis EM, Harvey NC, D'Angelo S, et al. Bone mineral content and areal density, but not bone area, predict an incident fracture risk: a comparative study in a UK prospective cohort. *Arch Osteoporos*. 2016 Dec;11(1):39. doi: 10.1007/s11657-016-0293-0. PMID: 27915448. Exclusion Code: X4.

107. Czerwiński E, Borowy P, Kumorek A, et al. Fracture risk prediction in outpatients from Krakow region using FRAX tool versus fracture risk in 11-year follow-up. *Ortop Traumatol Rehabil*. 2013 Nov-Dec;15(6):617-28. doi: 10.5604/15093492.1091517. PMID: 24662908. Exclusion Code: X4.

108. Czerwinski E, Cardona J, Plebanski R, et al. The efficacy and safety of abaloparatide-SC in men with osteoporosis: a randomized clinical trial. *J Bone Miner Res*. 2022;37(12):2435-42. doi: 10.1002/jbmr.4719. PMID: CN-02470664. Exclusion Code: X9.

109. Czerwinski E, Cardona J, Plebanski R, et al. The Efficacy and Safety of Abaloparatide-SC in Men With Osteoporosis: A Randomized Clinical Trial. *J Bone Miner Res*. 2022 Oct 3doi: 10.1002/jbmr.4719. PMID: 36190391. Exclusion Code: X2.

110. Dagan N, Elnekave E, Barda N, et al. Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization. *Nat Med*. 2020 Jan;26(1):77-82. doi: 10.1038/s41591-019-0720-z. PMID: 31932801. Exclusion Code: X9.

111. Das P, Goenka S, Sharma RK. Effect of subcutaneous administration of denosumab in postmenopausal osteoporosis. *European journal of molecular and clinical medicine*. 2022;9(3):2529-36. PMID: CN-02396006. Exclusion Code: X8.

112. Davis S, Martyn-St James M, Sanderson J, et al. A systematic review and economic evaluation of bisphosphonates for the prevention of fragility fractures. *Health Technol Assess*. 2016 Oct;20(78):1-406. doi: 10.3310/hta20780. PMID: 27801641. Exclusion Code: X3.

113. Davis SR, Tan A, Bell RJ. Targeted assessment of fracture risk in women at midlife. *Osteoporos Int*. 2015 Jun;26(6):1705-12. doi: 10.1007/s00198-015-3046-9. PMID: 25630574. Exclusion Code: X3.

114. de Abreu LLF, Holloway-Kew KL, Sajjad MA, et al. FRAX (Australia) scores in women with impaired fasting glucose and diabetes. *Bone Rep*. 2019 Dec;11:100223. doi: 10.1016/j.bonr.2019.100223. PMID: 31463338. Exclusion Code: X7.

115. de Vries BCS, Hegeman JH, Nijmeijer W, et al. Comparing three machine learning approaches to design a risk assessment tool for future fractures: predicting a subsequent major osteoporotic fracture in fracture patients with osteopenia and osteoporosis. *Osteoporos Int*. 2021 Mar;32(3):437-49. doi: 10.1007/s00198-020-05735-z. PMID: 33415373. Exclusion Code: X2.

116. Deal CL, Mitlak BH, Wang Y, et al. Response rates for hip, femoral neck, and lumbar spine bone mineral density in patients treated with abaloparatide followed by alendronate: results from phase 3 ACTIVExtend. *Bone Rep*. 2019 Dec;11:100230. doi: 10.1016/j.bonr.2019.100230. PMID: 31799340. Exclusion Code: X6.

117. Delmas PD, Ensrud KE, Adachi JD, et al. Efficacy of raloxifene on vertebral fracture risk reduction in postmenopausal women with osteoporosis: four-year results from a randomized clinical trial. *J Clin Endocrinol Metab*. 2002 Aug;87(8):3609-17. doi: 10.1210/jcem.87.8.8750. PMID: 12161484. Exclusion Code: X5.

118. Dempster DW, Brown JP, Fahrleitner-Pammer A, et al. Effects of long-term denosumab on bone histomorphometry and mineralization in women with postmenopausal osteoporosis. *J Clin Endocrinol Metab*. 2018 Jul 1;103(7):2498-509. doi: 10.1210/jc.2017-02669. PMID: 29672714. Exclusion Code: X7.

119. Dhiman P, Andersen S, Vestergaard P, et al. Does bone mineral density improve the predictive accuracy of fracture risk assessment? A prospective cohort study in Northern Denmark. *BMJ Open*. 2018 Apr 12;8(4):e018898. doi: 10.1136/bmjopen-2017-018898. PMID: 29654006. Exclusion Code: X3.

120. DiGiulio M, Loveless T, Heider G, et al. Bisphosphonate drug holidays: one size does not fit all. *Nurse Pract*. 2020 Mar;45(3):50-5. doi: 10.1097/01.NPR.0000605524.34681.e7. PMID: 32068658. Exclusion Code: X1.

121. Dömötör ZR, Vörhendi N, Hanák L, et al. Oral treatment with bisphosphonates of osteoporosis does not increase the risk of severe gastrointestinal side effects: a meta-analysis of randomized controlled trials. *Front Endocrinol (Lausanne)*. 2020;11:573976. doi: 10.3389/fendo.2020.573976. PMID: 33240217. Exclusion Code: X3.

122. Donaldson MG, Palermo L, Schousboe JT, et al. FRAX and risk of vertebral fractures: the fracture intervention trial. *J Bone Miner Res*. 2009 Nov;24(11):1793-9. doi: 10.1359/jbmr.090511 [doi]. PMID: 19419318. Exclusion Code: X9.

123. Drieling RL, LaCroix AZ, Beresford SA, et al. Long-term oral bisphosphonate use in relation to fracture risk in postmenopausal women with breast cancer: findings from the Women's Health Initiative. *Menopause*. 2016 Nov;23(11):1168-75. doi: 10.1097/gme.0000000000000696. PMID: 27433859. Exclusion Code: X6.

124. Duerksen DR, Lix LM, Johansson H, et al. Fracture risk assessment in celiac disease: a registry-based cohort study. *Osteoporos Int*. 2021 Jan;32(1):93-9. doi: 10.1007/s00198-020-05579-7. PMID: 32748311. Exclusion Code: X3.

125. Dundar U, Kavuncu V, Ciftci IH, et al. The effect of risedronate treatment on serum cytokines in postmenopausal osteoporosis: a 6-month randomized and controlled study. *J Bone Miner Metab*. 2009;27(4):464-70. doi: 10.1007/s00774-009-0055-9. PMID: 19301089. Exclusion Code: X10.

126. Eastell R, Mitlak BH, Wang Y, et al. Bone turnover markers to explain changes in lumbar spine BMD with abaloparatide and teriparatide: results from ACTIVE. *Osteoporos Int*. 2019 Mar;30(3):667-73. doi: 10.1007/s00198-018-04819-1. PMID: 30635696. Exclusion Code: X7.

127. Eiken PA, Prieto-Alhambra D, Eastell R, Abrahamsen B. Surgically treated osteonecrosis and osteomyelitis of the jaw and oral cavity in patients highly adherent to alendronate treatment: a nationwide user-only cohort study including over 60,000 alendronate users. *Osteoporos Int*. 2017 Oct;28(10):2921-8. doi: 10.1007/s00198-017-4132-y. PMID: 28664276. Exclusion Code: X3.

128. Ensrud KE, Taylor BC, Peters KW, et al. Implications of expanding indications for drug treatment to prevent fracture in older men in United States: cross sectional and longitudinal analysis of prospective cohort study. *BMJ*. 2014 Jul 3;349:g4120. doi: 10.1136/bmj.g4120. PMID: 24994809. Exclusion Code: X7.

129. Erjiang E, Wang T, Yang L, et al. Machine learning can improve clinical detection of low BMD: the DXA-HIP Study. *J Clin Densitom*. 2021 Oct-Dec;24(4):527-37. doi: 10.1016/j.jocd.2020.10.004. PMID: 33187864. Exclusion Code: X9.

130. Ettinger B, Black DM, Mitlak BH, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. Multiple Outcomes of Raloxifene Evaluation (MORE) Investigators. *JAMA*. 1999 Aug 18;282(7):637-45. doi: 10.1001/jama.282.7.637. PMID: 10517716. Exclusion Code: X5.

131. Ettinger B, Hillier TA, Pressman A, et al. Simple computer model for calculating and reporting 5-year osteoporotic fracture risk in postmenopausal women. *J Womens Health (Larchmt)*. 2005 Mar;14(2):159-71. doi: 10.1089/jwh.2005.14.159 [doi]. PMID: 15775734. Exclusion Code: X9.

132. Everts-Graber J, Reichenbach S, Gahl B, et al. Risk factors for vertebral fractures and bone loss after denosumab discontinuation: a real-world observational study. *Bone*. 2021 Mar;144:115830. doi: 10.1016/j.bone.2020.115830. PMID: 33359006. Exclusion Code: X6.

133. Fadda V, Maratea D, Trippoli S, Messori A. Gastrointestinal and renal side effects of bisphosphonates: differentiating between no proof of difference and proof of no difference. *J Endocrinol Invest*. 2015 Feb;38(2):189-92. doi: 10.1007/s40618-014-0211-5. PMID: 25412945. Exclusion Code: X3.

134. Fan Q, Wang J. The efficacy and safety of bisphosphonates for osteoporosis in women older than 65 years: a meta-analysis. *Curr Pharm Des*. 2020;26(32):4022-30. doi: 10.2174/1381612826666200423092602. PMID: 32324507. Exclusion Code: X3.

135. Fan Z, Li X, Zhang X, et al. Comparison of OSTA, FRAX and BMI for predicting postmenopausal osteoporosis in a Han population in Beijing: a cross sectional study. *Clin Interv Aging*. 2020;15:1171-80. doi: 10.2147/cia.S257166. PMID: 32764904. Exclusion Code: X8.

136. Ferrari S, Adachi JD, Lippuner K, et al. Further reductions in nonvertebral fracture rate with long-term denosumab treatment in the FREEDOM open-label extension and influence of hip bone mineral density after 3 years. *Osteoporos Int*. 2015 Dec;26(12):2763-71. doi: 10.1007/s00198-015-3179-x. PMID: 26068295. Exclusion Code: X6.

137. Ferrari S, Butler PW, Kendler DL, et al. Further nonvertebral fracture reduction beyond 3 years for up to 10 years of denosumab treatment. *J Clin Endocrinol Metab*. 2019 Aug 1;104(8):3450-61. doi: 10.1210/jc.2019-00271. PMID: 31125092. Exclusion Code: X6.

138. Ferrari S, Eastell R, Napoli N, et al. Denosumab in postmenopausal women with osteoporosis and diabetes: subgroup analysis of FREEDOM and FREEDOM extension. *Bone*. 2020 May;134:115268. doi: 10.1016/j.bone.2020.115268. PMID: 32058020. Exclusion Code: X2.

139. Ferrari S, Lewiecki EM, Butler PW, et al. Favorable skeletal benefit/risk of long-term denosumab therapy: a virtual-twin analysis of fractures prevented relative to skeletal safety events observed. *Bone*. 2020 May;134:115287. doi: 10.1016/j.bone.2020.115287. PMID: 32092479. Exclusion Code: X6.

140. Ferrari S, Libanati C, Lin CJF, et al. Relationship between bone mineral density T-Score and nonvertebral fracture risk over 10 years of denosumab treatment. *J Bone Miner Res*. 2019 Jun;34(6):1033-40. doi: 10.1002/jbmr.3722. PMID: 30919997. Exclusion Code: X6.

141. Fink HA, Butler ME, Claussen AM, et al. Performance of fracture risk assessment tools by race and ethnicity: a systematic review for the ASBMR Task Force on Clinical Algorithms for Fracture Risk. *J Bone Miner Res*. 2023 Dec;38(12):1731-41. doi: 10.1002/jbmr.4895. PMID: 37597237. Exclusion Code: X10.

142. Fogelman I, Ribot C, Smith R, et al. Risedronate reverses bone loss in postmenopausal women with low bone mass: results from a multinational, double-blind, placebo-controlled trial. BMD-MN Study Group. *J Clin Endocrinol Metab*. 2000 May;85(5):1895-900. doi: 10.1210/jcem.85.5.6603. PMID: 10843171. Exclusion Code: X10.

143. Forti P, Rietti E, Pisacane N, et al. A comparison of frailty indexes for prediction of adverse health outcomes in an elderly cohort. *Arch Gerontol Geriatr*. 2012 Jan-Feb;54(1):16-20. doi: 10.1016/j.archger.2011.01.007. PMID: 21339007. Exclusion Code: X4.

144. Francesco L, Elisa B, Raffaella M, et al. Assessing risk of osteoporotic fractures in primary care: development and validation of the FRA-HS algorithm. *Calcif Tissue Int*. 2017 Jun;100(6):537-49. doi: 10.1007/s00223-016-0230-7. PMID: 28160026. Exclusion Code: X3.

145. Friis-Holmberg T, Rubin KH, Brixen K, et al. Fracture risk prediction using phalangeal bone mineral density or FRAX((R))?-a Danish cohort study on men and women. *J Clin Densitom*. 2014 Jan-Mar;17(1):7-15. doi: 10.1016/j.jocd.2013.03.014. PMID: 23623379. Exclusion Code: X9.

146. Fuggle NR, Cooper C, Harvey NC, et al. Assessment of cardiovascular safety of anti-osteoporosis drugs. *Drugs*. 2020 Oct;80(15):1537-52. doi: 10.1007/s40265-020-01364-2. PMID: 32725307. Exclusion Code: X3.

147. Fujimaki H, Tomioka M, Kanoshima Y, et al. Accuracy of the Fracture Risk Assessment Tool for judging pharmacotherapy initiation for primary osteoporosis. *J Bone Miner Metab*. 2022 Sep;40(5):860-8. doi: 10.1007/s00774-022-01356-0. PMID: 35945295. Exclusion Code: X6.

148. Gartlehner G, Patel S, Viswanathan M, et al. Hormone therapy for the primary prevention of chronic conditions in postmenopausal women: an evidence review for the U.S. Preventive Services Task Force [Internet] (Prepared by the RTI International-University of North Carolina Evidence-based Practice Center under Contract No. HHSA-290-2012-00015-I.) U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews, Evidence Synthesis, No. 155. Report No.: 15-05227-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; December 2017. Exclusion Code: X5.

149. Gazoni FM, Civile VT, Atallah Á N, et al. Intravenous zoledronate for postmenopausal women with osteopenia and osteoporosis: a systematic review and metanalysis. *Sao Paulo Med J*. 2023;141(6):e2022480. doi: 10.1590/1516-3180.2022.0480.R1.27032023. PMID: 37255065. Exclusion Code: X3.

150. Genant HK, Engelke K, Bolognese MA, et al. Effects of romosozumab compared with teriparatide on bone density and mass at the spine and hip in postmenopausal women with low bone mass. *J Bone Miner Res*. 2017 Jan;32(1):181-7. doi: 10.1002/jbmr.2932. PMID: 27487526. Exclusion Code: X7.

151. Geusens P, Oates M, Miyauchi A, et al. The effect of 1 year of romosozumab on the incidence of clinical vertebral fractures in postmenopausal women with osteoporosis: results from the FRAME Study. *JBMR Plus*. 2019 Oct;3(10):e10211. doi: 10.1002/jbm4.10211. PMID: 31687647. Exclusion Code: X5.

152. Ghannam S, Blaney H, Gelfond J, Bruder JM. The use of FRAX in identifying women less than 65 years needing bone mineral density testing. *J Clin Densitom*. 2021 Jan-Mar;24(1):36-43. doi: 10.1016/j.jocd.2020.05.002. PMID: 32792202. Exclusion Code: X2.

153. Girman CJ, Chandler JM, Zimmerman SI, et al. Prediction of fracture in nursing home residents. *J Am Geriatr Soc*. 2002 Aug;50(8):1341-7. doi: 10.1046/j.1532-5415.2002.50354.x. PMID: 12164989. Exclusion Code: X2.

154. Gnudi S, Sitta E. Clinical risk factor evaluation to defer postmenopausal women from bone mineral density measurement: an Italian study. *J Clin Densitom*. 2005 Summer;8(2):199-205. doi: 10.1385/jcd:8:2:199. PMID: 15908708. Exclusion Code: X5.

155. Gough Courtney M, Roberts J, Godde K. Development of a diverse osteoporosis screening tool for older US adults from the health and retirement study. *Heliyon*. 2024 Jan 15;10(1):e23806. doi: 10.1016/j.heliyon.2023.e23806. PMID: 38192805. Exclusion Code: X5.

156. Graeff C, Campbell GM, Peña J, et al. Administration of romosozumab improves vertebral trabecular and cortical bone as assessed with quantitative computed tomography and finite element analysis. *Bone*. 2015 Dec;81:364-9. doi: 10.1016/j.bone.2015.07.036. PMID: 26232375. Exclusion Code: X7.

157. Greenspan SL, Bone HG, Ettinger MP, et al. Effect of recombinant human parathyroid hormone (1-84) on vertebral fracture and bone mineral density in postmenopausal women with osteoporosis: a randomized trial. *Ann Intern Med*. 2007 Mar 6;146(5):326-39. doi: 10.7326/0003-4819-146-5-200703060-00005. PMID: 17339618. Exclusion Code: X5.

158. Greenspan SL, Fitzpatrick LA, Mitlak B, et al. Abaloparatide followed by alendronate in women ≥80 years with osteoporosis: post hoc analysis of ACTIVExtend. *Menopause*. 2020 Oct;27(10):1137-42. doi: 10.1097/gme.0000000000001593. PMID: 32665529. Exclusion Code: X6.

159. Gregson CL, Armstrong DJ, Bowden J, et al. UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos*. 2022 Apr 5;17(1):58. doi: 10.1007/s11657-022-01061-5. PMID: 35378630. Exclusion Code: X1.

160. Grey A, Horne A, Gamble G, et al. Ten years of very infrequent zoledronate therapy in older women: an open-label extension of a randomized trial. *J Clin Endocrinol Metab*. 2020 Apr 1;105(4)doi: 10.1210/clinem/dgaa062. PMID: 32016386. Exclusion Code: X6.

161. Habib ZA. Bisphosphonates in the treatment of osteoporosis: a review of skeletal safety concerns. *Expert Rev Endocrinol Metab*. 2017 Jan;12(1):59-71. doi: 10.1080/17446651.2017.1256199. PMID: 30058876. Exclusion Code: X1.

162. Hagino H, Soen S, Sugimoto T, et al. Changes in quality of life in patients with postmenopausal osteoporosis receiving weekly bisphosphonate treatment: a 2-year multicenter study in Japan. *J Bone Miner Metab*. 2019 Mar;37(2):273-81. doi: 10.1007/s00774-018-0914-3. PMID: 29523963. Exclusion Code: X6.

163. Harvey NC, Kanis JA, Odén A, et al. FRAX and the effect of teriparatide on vertebral and non-vertebral fracture. *Osteoporos Int*. 2015 Nov;26(11):2677-84. doi: 10.1007/s00198-015-3173-3. PMID: 26092063. Exclusion Code: X5.

164. Harvey NC, Kanis JA, Odén A, et al. Efficacy of weekly teriparatide does not vary by baseline fracture probability calculated using FRAX. *Osteoporos Int*. 2015 Sep;26(9):2347-53. doi: 10.1007/s00198-015-3129-7. PMID: 26092062. Exclusion Code: X2.

165. Harvey NC, Odén A, Orwoll E, et al. Falls Falls predict fractures independently of FRAX probability: a meta-analysis of the Osteoporotic Fractures in Men (MrOS) Study. *J Bone Miner Res*. 2018 Mar;33(3):510-6. doi: 10.1002/jbmr.3331. PMID: 29220072. Exclusion Code: X3.

166. Hassanabadi N, Berger C, Papaioannou A, et al. Variation in bone mineral density and fractures over 20 years among Canadians: a comparison of the Canadian Multicenter Osteoporosis Study and the Canadian Longitudinal Study on Aging. *Osteoporos Int*. 2023 Feb;34(2):357-67. doi: 10.1007/s00198-022-06623-4. PMID: 36449036. Exclusion Code: X7.

167. Henry MJ, Pasco JA, Merriman EN, et al. Fracture risk score and absolute risk of fracture. *Radiology*. 2011 May;259(2):495-501. doi: 10.1148/radiol.10101406. PMID: 21292868. Exclusion Code: X9.

168. Henry MJ, Pasco JA, Sanders KM, et al. Fracture risk (FRISK) score: Geelong Osteoporosis Study. *Radiology*. 2006 Oct;241(1):190-6. doi: 10.1148/radiol.2411051290. PMID: 16928979. Exclusion Code: X9.

169. Herd RJ, Balena R, Blake GM, et al. The prevention of early postmenopausal bone loss by cyclical etidronate therapy: a 2-year, double-blind, placebo-controlled study. *Am J Med*. 1997 Aug;103(2):92-9. doi: 10.1016/s0002-9343(97)00019-3. PMID: 9274891. Exclusion Code: X5.

170. Hillier TA, Cauley JA, Rizzo JH, et al. WHO absolute fracture risk models (FRAX): do clinical risk factors improve fracture prediction in older women without osteoporosis? *J Bone Miner Res*. 2011 Aug;26(8):1774-82. doi: 10.1002/jbmr.372. PMID: 21351144. Exclusion Code: X7.

171. Ho-Le TP, Center JR, Eisman JA, et al. Prediction of hip fracture in post-menopausal women using artificial neural network approach. *Annu Int Conf IEEE Eng Med Biol Soc*. 2017 Jul;2017:4207-10. doi: 10.1109/embc.2017.8037784. PMID: 29060825. Exclusion Code: X5.

172. Ho-Pham LT, Doan MC, Van LH, Nguyen TV. Development of a model for identification of individuals with high risk of osteoporosis. *Arch Osteoporos*. 2020 Jul 22;15(1):111. doi: 10.1007/s11657-020-00788-3. PMID: 32699999. Exclusion Code: X3.

173. Hoff M, Meyer HE, Skurtveit S, et al. Validation of FRAX and the impact of self-reported falls among elderly in a general population: the HUNT study, Norway. *Osteoporos Int*. 2017 Oct;28(10):2935-44. doi: 10.1007/s00198-017-4134-9. PMID: 28668994. Exclusion Code: X4.

174. Hoff M, Skovlund E, Meyer HE, et al. Does treatment with bisphosphonates protect against fractures in real life? The HUNT study, Norway. *Osteoporos Int*. 2021 Jul;32(7):1395-404. doi: 10.1007/s00198-021-05845-2. PMID: 33479844. Exclusion Code: X3.

175. Høiberg MP, Rubin KH, Hermann AP, et al. Diagnostic devices for osteoporosis in the general population: a systematic review. *Bone*. 2016 Nov;92:58-69. doi: 10.1016/j.bone.2016.08.011. PMID: 27542659. Exclusion Code: X5.

176. Holloway KL, Mohebbi M, Betson AG, et al. Prediction of major osteoporotic and hip fractures in Australian men using FRAX scores adjusted with trabecular bone score. *Osteoporos Int*. 2018 Jan;29(1):101-8. doi: 10.1007/s00198-017-4226-6. PMID: 28940052. Exclusion Code: X9.

177. Holloway-Kew KL, Zhang Y, Betson AG, et al. How well do the FRAX (Australia) and Garvan calculators predict incident fractures? Data from the Geelong osteoporosis study. *Osteoporos Int*. 2019 Oct;30(10):2129-39. doi: 10.1007/s00198-019-05088-2. PMID: 31317250. Exclusion Code: X9.

178. Hsu CY, Wu CH, Yu SF, et al. Novel algorithm generating strategy to identify high fracture risk population using a hybrid intervention threshold. *J Bone Miner Metab*. 2020 Mar;38(2):213-21. doi: 10.1007/s00774-019-01046-4. PMID: 31583541. Exclusion Code: X7.

179. Huang G, Coviello A, LaValley MP, et al. Surgical menopause and frailty risk in community-dwelling older women: study of osteoporotic fractures. *J Am Geriatr Soc*. 2018 Nov;66(11):2172-7. doi: 10.1111/jgs.15505. PMID: 30251302. Exclusion Code: X5.

180. Huang JY, Song WZ, Huang M. Effectiveness of osteoporosis self-assessment tool for Asians in screening for osteoporosis in healthy males over 40 years old in China. *J Clin Densitom*. 2017 Apr-Jun;20(2):153-9. doi: 10.1016/j.jocd.2017.01.003. PMID: 28153410. Exclusion Code: X8.

181. Huang ML, Hsieh TJ, Lin SS, Huang WC. Associations between trabecular bone score and bone mineral density in Taiwanese older adult men. *Journal of Men's Health*. 2023;19(1):15-22. doi: 10.22514/jomh.2023.005. Exclusion Code: X5.

182. Huang X, Chen B, Thabane L, et al. Fragility of results from randomized controlled trials supporting the guidelines for the treatment of osteoporosis: a retrospective analysis. *Osteoporos Int*. 2021 Sep;32(9):1713-23. doi: 10.1007/s00198-021-05865-y. PMID: 33595680. Exclusion Code: X3.

183. Hundrup YA, Jacobsen RK, Andreasen AH, et al. Validation of a 5-year risk score of hip fracture in postmenopausal women. The Danish nurse cohort study. *Osteoporos Int*. 2010 Dec;21(12):2135-42. doi: 10.1007/s00198-010-1176-7 [doi]. PMID: 20157806. Exclusion Code: X4.

184. Hung WC, Lin YL, Cheng TT, et al. Establish and validate the reliability of predictive models in bone mineral density by deep learning as examination tool for women. *Osteoporos Int*. 2023 Sep 20doi: 10.1007/s00198-023-06913-5. PMID: 37728768. Exclusion Code: X4.

185. Iconaru L, Charles A, Baleanu F, et al. Selection for treatment of patients at high risk of fracture by the short versus long term prediction models - data from the Belgian FRISBEE cohort. *Osteoporos Int*. 2023 Jun;34(6):1119-25. doi: 10.1007/s00198-023-06737-3. PMID: 37022466. Exclusion Code: X2.

186. Iconaru L, Moreau M, Kinnard V, et al. Does the prediction accuracy of osteoporotic fractures by BMD and clinical risk factors vary with fracture site? *JBMR Plus*. 2019 Dec;3(12):e10238. doi: 10.1002/jbm4.10238. PMID: 31844826. Exclusion Code: X3.

Appendix C. Excluded Studies

187. Iki M, Fujita Y, Tamaki J, et al. Trabecular bone score may improve FRAX(R) prediction accuracy for major osteoporotic fractures in elderly Japanese men: the Fujiwara-kyo Osteoporosis Risk in Men (FORMEN) Cohort Study. *Osteoporos Int*. 2015 Jun;26(6):1841-8. doi: 10.1007/s00198-015-3092-3. PMID: 25752623. Exclusion Code: X9.

188. Iki M, Fujita Y, Tamaki J, et al. Trabecular bone score may improve FRAX® prediction accuracy for major osteoporotic fractures in elderly Japanese men: the Fujiwara-kyo Osteoporosis Risk in Men (FORMEN) Cohort Study. *Osteoporos Int*. 2015 Jun;26(6):1841-8. doi: 10.1007/s00198-015-3092-3. PMID: 25752623. Exclusion Code: X3.

189. Iki M, Tamaki J, Kadowaki E, et al. Trabecular bone score (TBS) predicts vertebral fractures in Japanese women over 10 years independently of bone density and prevalent vertebral deformity: the Japanese Population-Based Osteoporosis (JPOS) cohort study. *J Bone Miner Res*. 2014 Feb;29(2):399-407. doi: 10.1002/jbmr.2048 [doi]. PMID: 23873699. Exclusion Code: X5.

190. İlgezdi ZD, Aktaş İ, Doğan Metin F, et al. Acute effect of zoledronic acid infusion on atrial fibrillation development in patients with osteoporosis. *Anatol J Cardiol*. 2015 Apr;15(4):320-4. doi: 10.5152/akd.2014.5333. PMID: 25413229. Exclusion Code: X6.

191. Imai T, Hosoi T, Hagino H, et al. Antiresorptive drugs and the risk of femoral shaft fracture in men and women with osteoporosis: a cohort study using the national database of health insurance claims of Japan. *J Epidemiol*. 2023 Dec 5;33(12):633-9. doi: 10.2188/jea.JE20220099. PMID: 36567127. Exclusion Code: X6.

192. Indhavivadhana S, Rattanachaiyanont M, Angsuwathana S, et al. Validation of osteoporosis risk assessment tools in middle-aged Thai women. *Climacteric*. 2016 Dec;19(6):588-93. doi: 10.1080/13697137.2016.1231176. PMID: 27667093. Exclusion Code: X6.

193. Inoue D, Muraoka R, Okazaki R, et al. Efficacy and safety of risedronate in osteoporosis subjects with comorbid diabetes, hypertension, and/or dyslipidemia: a post hoc analysis of phase III trials conducted in Japan. *Calcif Tissue Int*. 2016 Feb;98(2):114-22. doi: 10.1007/s00223-015-0071-9. PMID: 26466937. Exclusion Code: X6.

194. Iqbal SM, Qamar I, Zhi C, et al. Role of bisphosphonate therapy in patients with osteopenia: a systemic review. *Cureus*. 2019 Feb 27;11(2):e4146. doi: 10.7759/cureus.4146. PMID: 31058029. Exclusion Code: X3.

195. Ishibashi H, Crittenden DB, Miyauchi A, et al. Romosozumab increases bone mineral density in postmenopausal Japanese women with osteoporosis: a phase 2 study. *Bone*. 2017 Oct;103:209-15. doi: 10.1016/j.bone.2017.07.005. PMID: 28687496. Exclusion Code: X5.

196. Izano MA, Lo JC, Adams AL, et al. Bisphosphonate treatment beyond 5 years and hip fracture risk in older women. *JAMA Netw Open*. 2020 Dec 1;3(12):e2025190. doi: 10.1001/jamanetworkopen.2020.25190. PMID: 33284336. Exclusion Code: X6.

197. Jain RK, Weiner MG, Zhao H, Vokes T. Comorbid conditions and GFR predict nonvertebral fractures in patients with diabetes in an ethnic-specific manner. *J Clin Endocrinol Metab*. 2020 Jun 1;105(6)doi: 10.1210/clinem/dgaa141. PMID: 32193529. Exclusion Code: X5.

198. Jang SP, Yeo I, So SY, et al. Atypical femoral shaft fractures in female bisphosphonate users were associated with an increased anterolateral femoral bow and a thicker lateral cortex: a case-control study. *Biomed Res Int*. 2017;2017:5932496. doi: 10.1155/2017/5932496. PMID: 28459066. Exclusion Code: X3.

199. Japan Registry for Clinical Trials.. Phase III clinical study of MN-10-T AI for patients with osteoporosis at higher fracture risk. 2016. Exclusion Code: X1.

200. Japan Registry for Clinical Trials. ITM-058 phase 3 study. 2017. Exclusion Code: X1.

201. Jiang X, Good LE, Spinka R, Schnatz PF. Osteoporosis screening in postmenopausal women aged 50-64 years: BMI alone compared with current screening tools. *Maturitas*. 2016 Jan;83:59-64. doi: 10.1016/j.maturitas.2015.09.009. PMID: 26471931. Exclusion Code: X9.

202. Johnson T, Fox E, Hassanbein S. Implementing an electronic medical record osteoporosis self-assessment tool score which identifies patients at risk for osteoporosis promotes osteoporosis evaluation. *Geriatric Orthopaedic Surgery and Rehabilitation.* 2021;12:21514593211002157. doi: 10.1177/21514593211002157. PMID: 35186418. Exclusion Code: X6.

203. Jonasson GB, Sundh V, Hakeberg M, et al. Evaluation of clinical and radiographic indices as predictors of osteoporotic fractures: a 10-year longitudinal study. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2018 May;125(5):487-94. doi: 10.1016/j.oooo.2017.11.009. PMID: 29273194. Exclusion Code: X4.

204. Japan Registry for Clinical Trials. TRINITY study. 2020. Exclusion Code: X1.

205. Jung SM, Han S, Kwon HY. Dose-intensity of bisphosphonates and the risk of osteonecrosis of the jaw in osteoporosis patients. *Front Pharmacol*. 2018;9(JUL):796. doi: 10.3389/fphar.2018.00796. PMID: 30079024. Exclusion Code: X6.

206. Kalvesten J, Lui LY, Brismar T, Cummings S. Digital X-ray radiogrammetry in the study of osteoporotic fractures: comparison to dual energy x-ray absorptiometry and FRAX. *Bone*. 2016 May;86:30-5. doi: 10.1016/j.bone.2016.02.011. PMID: CN-01137905. Exclusion Code: X5.

207. Kälvesten J, Lui LY, Brismar T, Cummings S. Digital X-ray radiogrammetry in the study of osteoporotic fractures: comparison to dual energy x-ray absorptiometry and FRAX. *Bone*. 2016 May;86:30-5. doi: 10.1016/j.bone.2016.02.011. PMID: 26921822. Exclusion Code: X9.

208. Kanazawa I, Notsu M, Miyake H, et al. Assessment using serum insulin-like growth factor-I and bone mineral density is useful for detecting prevalent vertebral fractures in patients with type 2 diabetes mellitus. *Osteoporos Int*. 2018 Nov;29(11):2527-35. doi: 10.1007/s00198-018-4638-y. PMID: 30030585. Exclusion Code: X5.

209. Kanis JA, Harvey NC, Lorentzon M, et al. Combining fracture outcomes in phase 3 trials of osteoporosis: an analysis of the effects of denosumab in postmenopausal women. *Osteoporos Int*. 2021 Jan;32(1):165-71. doi: 10.1007/s00198-020-05699-0. PMID: 33156354. Exclusion Code: X3.

210. Kanis JA, Johansson H, Harvey NC, et al. Adjusting conventional FRAX estimates of fracture probability according to the recency of sentinel fractures. *Osteoporos Int*. 2020 Oct;31(10):1817-28. doi: 10.1007/s00198-020-05517-7. PMID: 32613411. Exclusion Code: X7.

211. Kanis JA, Johnell O, Oden A, et al. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int*. 2008 Apr;19(4):385-97. doi: 10.1007/s00198-007-0543-5. PMID: 18292978. Exclusion Code: X3.

212. Kanis JA, Oden A, Johnell O, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int*. 2007 Aug;18(8):1033-46. doi: 10.1007/s00198-007-0343-y. PMID: 17323110. Exclusion Code: X3.

213. Karimi Fard M, Aminorroaya A, Kachuei A, et al. Alendronate improves fasting plasma glucose and insulin sensitivity, and decreases insulin resistance in prediabetic osteopenic postmenopausal women: a randomized triple-blind clinical trial. *Journal of Diabetes Investigation*. 2019 May;10(3):731-7. doi: 10.1111/jdi.12944. PMID: CN-01991102. Exclusion Code: X8.

214. Karlsson L, Mesterton J, Tepie MF, et al. Exploring methods for comparing the real-world effectiveness of treatments for osteoporosis: adjusted direct comparisons versus using patients as their own control. *Arch Osteoporos*. 2017 Sep 21;12(1):81. doi: 10.1007/s11657-017-0375-7. PMID: 28936581. Exclusion Code: X7.

215. Kasai H, Mori Y, Ose A, et al. Prediction of fracture risk from early-stage bone markers in patients with osteoporosis treated with once-yearly administered zoledronic acid. *J Clin Pharmacol*. 2021 May;61(5):606-13. doi: 10.1002/jcph.1774. PMID: 33135182. Exclusion Code: X2.

216. Keaveny TM, Crittenden DB, Bolognese MA, et al. Greater gains in spine and hip strength for romosozumab compared with teriparatide in postmenopausal women with low bone mass. *J Bone Miner Res*. 2017 Sep;32(9):1956-62. doi: 10.1002/jbmr.3176. PMID: 28543940. Exclusion Code: X7.

217. Keech CA, Sashegyi A, Barrett-Connor E. Year-by-year analysis of cardiovascular events in the Multiple Outcomes of Raloxifene Evaluation (MORE) trial. *Curr Med Res Opin*. 2005 Jan;21(1):135-40. doi: 10.1185/030079904x18045. PMID: 15881485. Exclusion Code: X5.

218. Kellier-Steele N, Casso D, Anderson A, et al. Assessing the incidence of osteosarcoma among teriparatide-treated patients using linkage of commercial pharmacy and state cancer registry data, contributing to the removal of boxed warning and other labeling changes. *Bone*. 2022 Jul;160:116394. doi: 10.1016/j.bone.2022.116394. PMID: 35318162. Exclusion Code: X5.

219. Kendler DL, Bone HG, Massari F, et al. Bone mineral density gains with a second 12-month course of romosozumab therapy following placebo or denosumab. *Osteoporos Int*. 2019 Dec;30(12):2437-48. doi: 10.1007/s00198-019-05146-9. PMID: 31628490. Exclusion Code: X6.

220. Kendler DL, Chines A, Brandi ML, et al. The risk of subsequent osteoporotic fractures is decreased in subjects experiencing fracture while on denosumab: results from the FREEDOM and FREEDOM Extension studies. *Osteoporos Int*. 2019 Jan;30(1):71-8. doi: 10.1007/s00198-018-4687-2. PMID: 30244369. Exclusion Code: X3.

221. Keshtkar A, Tabatabaie O, Matin N, et al. Clinical performance of seven prescreening tools for osteoporosis in Iranian postmenopausal women. *Rheumatol Int*. 2015 Dec;35(12):1995-2004. doi: 10.1007/s00296-015-3286-1. PMID: 25980683. Exclusion Code: X8.

222. Khanizadeh F, Rahmani A, Asadollahi K, Ahmadi MRH. Combination therapy of curcumin and alendronate modulates bone turnover markers and enhances bone mineral density in postmenopausal women with osteoporosis. *Arch Endocrinol Metab*. 2018 Aug;62(4):438-45. doi: 10.20945/2359-3997000000060. PMID: 30304108. Exclusion Code: X7.

223. Kharroubi A, Saba E, Ghannam I, Darwish H. Evaluation of the validity of osteoporosis and fracture risk assessment tools (IOF one minute test, SCORE, and FRAX) in postmenopausal Palestinian women. *Arch Osteoporos*. 2017 Dec;12(1):6. doi: 10.1007/s11657-016-0298-8. PMID: 28013446. Exclusion Code: X8.

224. Kim H, Kim JH, Kim MJ, et al. Low predictive value of FRAX adjusted by trabecular bone score for osteoporotic fractures in Korean women: a community-based cohort study. *Endocrinol Metab (Seoul)*. 2020 Jun;35(2):359-66. doi: 10.3803/EnM.2020.35.2.359. PMID: 32615720. Exclusion Code: X3.

225. Kim HY, Jang EJ, Park B, et al. Development of a Korean Fracture Risk Score (KFRS) for predicting osteoporotic fracture risk: analysis of data from the Korean National Health Insurance Service. *PLoS One*. 2016;11(7):e0158918. doi: 10.1371/journal.pone.0158918. PMID: 27399597. Exclusion Code: X3.

226. Kim S, Bang HH, Yoo H, et al. Difference in bone mineral density change at the lateral femoral cortices according to administration of different bisphosphonate agents. *J Bone Metab*. 2016 May;23(2):85-93. doi: 10.11005/jbm.2016.23.2.85. PMID: 27294080. Exclusion Code: X7.

227. Kim SH, Lee YK, Kim TY, et al. Incidence of and risk for osteonecrosis of the jaw in Korean osteoporosis patients treated with bisphosphonates: a nationwide cohort-study. *Bone*. 2021 Feb;143:115650. doi: 10.1016/j.bone.2020.115650. PMID: 32956854. Exclusion Code: X10.

228. Kim Y, Tian Y, Yang J, et al. Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. *Sci Rep*. 2020 Jul 6;10(1):11115. doi: 10.1038/s41598-020-68037-8. PMID: 32632237. Exclusion Code: X6.

229. Kirilova E, Kirilov N, Bischoff F. Predictive ability of the osteoporosis self-assessment tool for assessing the risk of osteoporosis. *Revmatologiia (Bulgaria)*. 2019;27(3):3-9. Exclusion Code: X10.

230. Kittithaworn A, Toro-Tobon D, Sfeir JG. Cardiovascular benefits and risks associated with calcium, vitamin D, and antiresorptive therapy in the management of skeletal fragility. *Womens Health (Lond)*. 2023 Jan-Dec;19:17455057231170059. doi: 10.1177/17455057231170059. PMID: 37129172. Exclusion Code: X5.

231. Kline GA, Lix LM, Morin SN, Leslie WD. Fracture risk in Asian-Canadian women is significantly over-estimated by the Canadian Association of Radiologists-Osteoporosis Canada risk prediction tool: retrospective cohort study. *Arch Osteoporos*. 2022 Oct 6;17(1):133. doi: 10.1007/s11657-022-01173-y. PMID: 36201065. Exclusion Code: X5.

232. Klop C, de Vries F, Bijlsma JW, et al. Predicting the 10-year risk of hip and major osteoporotic fracture in rheumatoid arthritis and in the general population: an independent validation and update of UK FRAX without bone mineral density. *Ann Rheum Dis*. 2016 Dec;75(12):2095-100. doi: 10.1136/annrheumdis-2015-208958. PMID: 26984006. Exclusion Code: X4.

233. Koh JH, Myong JP, Yoo J, et al. Predisposing factors associated with atypical femur fracture among postmenopausal Korean women receiving bisphosphonate therapy: 8 years' experience in a single center. *Osteoporos Int*. 2017 Nov;28(11):3251-9. doi: 10.1007/s00198-017-4169-y. PMID: 28748389. Exclusion Code: X3.

234. Kong L, Zuo K, Ma L. Clinical effect of zoledronic acid in the treatment of senile osteoporosis. *Pak J Med Sci*. 2020 Nov-Dec;36(7):1703-7. doi: 10.12669/pjms.36.7.1964. PMID: 33235601. Exclusion Code: X2.

235. Kong SH, Kim S, Kim Y, et al. Development and validation of common data model-based fracture prediction model using machine learning algorithm. *Osteoporos Int*. 2023 Aug;34(8):1437-51. doi: 10.1007/s00198-023-06787-7. PMID: 37195320. Exclusion Code: X5.

236. Kranenburg G, Bartstra JW, Weijmans M, et al. Bisphosphonates for cardiovascular risk reduction: a systematic review and meta-analysis. *Atherosclerosis*. 2016 Sep;252:106-15. doi: 10.1016/j.atherosclerosis.2016.06.039. PMID: 27513349. Exclusion Code: X3.

237. Kruse C, Eiken P, Vestergaard P. Machine learning principles can improve hip fracture prediction. *Calcif Tissue Int*. 2017 Apr;100(4):348-60. doi: 10.1007/s00223-017-0238-7. PMID: 28197643. Exclusion Code: X5.

238. Kucukler FK, Simsek Y, Turk A, et al. Osteoporosis and silent vertebral fractures in nursing home resident elderly men in Turkey. *J Clin Densitom*. 2017 Apr-Jun;20(2):188-95. doi: 10.1016/j.jocd.2015.05.064. PMID: 26071170. Exclusion Code: X2.

239. Kužma M, Hans D, Koller T, et al. Less strict intervention thresholds for the FRAX and TBS-adjusted FRAX predict clinical fractures in osteopenic postmenopausal women with no prior fractures. *J Bone Miner Metab*. 2018 Sep;36(5):580-8. doi: 10.1007/s00774-017-0864-1. PMID: 28884422. Exclusion Code: X3.

240. Kvist AV, Faruque J, Vallejo-Yagüe E, et al. Cardiovascular safety profile of romosozumab: a pharmacovigilance analysis of the US Food and Drug Administration Adverse Event Reporting System (FAERS). *Journal of Clinical Medicine*. 2021 Apr 13;10(8)doi: 10.3390/jcm10081660. PMID: 33924496. Exclusion Code: X6.

241. Lam MT, Sing CW, Li GHY, et al. Development and validation of a risk score to predict the first hip fracture in the oldest old: a retrospective cohort study. *J Gerontol A Biol Sci Med Sci*. 2020 Apr 17;75(5):980-6. doi: 10.1093/gerona/glz178. PMID: 31353417. Exclusion Code: X4.

242. Langdahl BL, Hofbauer LC, Forfar JC. Cardiovascular safety and sclerostin inhibition. *J Clin Endocrinol Metab*. 2021 Jun 16;106(7):1845-53. doi: 10.1210/clinem/dgab193. PMID: 33755157. Exclusion Code: X3.

243. Langdahl BL, Ljunggren Ö, Benhamou CL, et al. Fracture rate, quality of life and back pain in patients with osteoporosis treated with teriparatide: 24-month results from the Extended Forsteo Observational Study (ExFOS). *Calcif Tissue Int*. 2016 Sep;99(3):259-71. doi: 10.1007/s00223-016-0143-5. PMID: 27137783. Exclusion Code: X6.

244. Langdahl BL, Silverman S, Fujiwara S, et al. Real-world effectiveness of teriparatide on fracture reduction in patients with osteoporosis and comorbidities or risk factors for fractures: Integrated analysis of 4 prospective observational studies. *Bone*. 2018 Nov;116:58-66. doi: 10.1016/j.bone.2018.07.013. PMID: 30021126. Exclusion Code: X3.

245. Langdahl BL, Teglbjaerg CS, Ho PR, et al. A 24-month study evaluating the efficacy and safety of denosumab for the treatment of men with low bone mineral density: results from the ADAMO trial. *J Clin Endocrinol Metab*. 2015 Apr;100(4):1335-42. doi: 10.1210/jc.2014-4079 [doi]. PMID: 25607608. Exclusion Code: X3.

246. Lauppe R, Åkesson KE, Ljunggren Ö, et al. Differing impact of clinical factors on the risk of fracture in younger and older women in the general population and an osteoporosis clinic population. *Arch Osteoporos*. 2019 Apr 8;14(1):45. doi: 10.1007/s11657-019-0592-3. PMID: 30963310. Exclusion Code: X7.

247. Laura I, Felicia B, Alexia C, et al. Which treatment to prevent an imminent fracture? *Bone Rep*. 2021 Dec;15:101105. doi: 10.1016/j.bonr.2021.101105. PMID: 34386562. Exclusion Code: X3.

248. Leder BZ, Mitlak B, Hu MY, et al. Effect of abaloparatide vs alendronate on fracture risk reduction in postmenopausal women with osteoporosis. *J Clin Endocrinol Metab*. 2020 Mar 1;105(3):938-43. doi: 10.1210/clinem/dgz162. PMID: 31674644. Exclusion Code: X6.

249. Leder BZ, Zapalowski C, Hu MY, et al. Fracture and bone mineral density response by baseline risk in patients treated with abaloparatide followed by alendronate: results from the phase 3 ACTIVExtend trial. *J Bone Miner Res*. 2019 Dec;34(12):2213-9. doi: 10.1002/jbmr.3848. PMID: 31411768. Exclusion Code: X6.

250. Lee HJ, Hwang SY, Kim SC, et al. Relationship between metabolic syndrome and bone fracture risk in mid-aged Korean women using FRAX scoring system. *Metab Syndr Relat Disord*. 2020 May;18(4):219-24. doi: 10.1089/met.2019.0060. PMID: 32077792. Exclusion Code: X3.

251. Lee WY, Sun LM, Lin MC, et al. A higher dosage of oral alendronate will increase the subsequent cancer risk of osteoporosis patients in Taiwan: a population-based cohort study. *PLoS One*. 2012;7(12):e53032. doi: 10.1371/journal.pone.0053032. PMID: 23300854. Exclusion Code: X10.

252. Leeyaphan J, Rojjananukulpong K, Intarasompun P, Peerakul Y. Development and Validation of a New Clinical Diagnostic Screening Model for Osteoporosis in Postmenopausal Women. *J Bone Metab*. 2023 May;30(2):179-88. doi: 10.11005/jbm.2023.30.2.179. PMID: 37449350. Exclusion Code: X8.

253. Lello S, Sorge R, Surico N. Osteoporosis's Menopausal Epidemiological Risk Observation (O.M.E.R.O.) study. *Gynecol Endocrinol*. 2015;31(12):992-8. doi: 10.3109/09513590.2015.1063605. PMID: 26172928. Exclusion Code: X1.

254. Leslie WD, Brennan-Olsen SL, Morin SN, Lix LM. Fracture prediction from repeat BMD measurements in clinical practice. *Osteoporos Int*. 2016 Jan;27(1):203-10. doi: 10.1007/s00198-015-3259-y. PMID: 26243362. Exclusion Code: X7.

255. Leslie WD, Lix LM. Absolute fracture risk assessment using lumbar spine and femoral neck bone density measurements: derivation and validation of a hybrid system. *J Bone Miner Res*. 2011 Mar;26(3):460-7. doi: 10.1002/jbmr.248 [doi]. PMID: 20839285. Exclusion Code: X7.

256. Leslie WD, Lix LM, Johansson H, et al. Spine-hip discordance and fracture risk assessment: a physician-friendly FRAX enhancement. *Osteoporos Int*. 2011 Mar;22(3):839-47. doi: 10.1007/s00198-010-1461-5 [doi]. PMID: 20959961. Exclusion Code: X7.

257. Leslie WD, Lix LM, Johansson H, et al. A comparative study of using non-hip bone density inputs with FRAX(R). *Osteoporos Int*. 2012 Mar;23(3):853-60. doi: 10.1007/s00198-011-1814-8. PMID: 22008881. Exclusion Code: X9.

258. Leslie WD, Lix LM, Majumdar SR, et al. Total hip bone area affects fracture prediction with FRAX® in Canadian white women. *J Clin Endocrinol Metab*. 2017 Nov 1;102(11):4242-9. doi: 10.1210/jc.2017-01327. PMID: 29092086. Exclusion Code: X7.

259. Leslie WD, Luo Y, Yang S, et al. Fracture risk indices from DXA-based finite element analysis predict incident fractures independently from FRAX: the Manitoba BMD registry. *J Clin Densitom*. 2019 Jul-Sep;22(3):338-45. doi: 10.1016/j.jocd.2019.02.001. PMID: 30852033. Exclusion Code: X9.

260. Leslie WD, Morin SN, Lix LM, et al. Fracture prediction from self-reported falls in routine clinical practice: a registry-based cohort study. *Osteoporos Int*. 2019 Nov;30(11):2195-203. doi: 10.1007/s00198-019-05106-3. PMID: 31372711. Exclusion Code: X3.

261. Leslie WD, Morin SN, Lix LM, et al. The Effect of Fracture Recency on Observed 10-Year Fracture Probability: A Registry-Based Cohort Study. *J Bone Miner Res*. 2022 May;37(5):848-55. doi: 10.1002/jbmr.4526. PMID: 35147245. Exclusion Code: X3.

262. Leslie WD, Morin SN, Lix LM, et al. Performance of FRAX in women with breast cancer initiating aromatase inhibitor therapy: a registry-based cohort study. *J Bone Miner Res*. 2019 Aug;34(8):1428-35. doi: 10.1002/jbmr.3726. PMID: 31069862. Exclusion Code: X2.

263. Leslie WD, Shevroja E, Johansson H, et al. Risk-equivalent T-score adjustment for using lumbar spine trabecular bone score (TBS): the Manitoba BMD registry. *Osteoporos Int*. 2018 Mar;29(3):751-8. doi: 10.1007/s00198-018-4405-0. PMID: 29392355. Exclusion Code: X5.

264. Leslie WD, Tsang JF, Lix LM. Simplified system for absolute fracture risk assessment: clinical validation in Canadian women. *J Bone Miner Res*. 2009 Feb;24(2):353-60. doi: 10.1359/jbmr.081012 [doi]. PMID: 19514851. Exclusion Code: X7.

265. Lewiecki EM, Blicharski T, Goemaere S, et al. A phase III randomized placebo-controlled trial to evaluate efficacy and safety of romosozumab in men with osteoporosis. *J Clin Endocrinol Metab*. 2018 Sep 1;103(9):3183-93. doi: 10.1210/jc.2017-02163. PMID: 29931216. Exclusion Code: X2.

266. Lewiecki EM, Dinavahi RV, Lazaretti-Castro M, et al. One year of romosozumab followed by two years of denosumab maintains fracture risk reductions: results of the FRAME extension study. *J Bone Miner Res*. 2019 Mar;34(3):419-28. doi: 10.1002/jbmr.3622. PMID: 30508316. Exclusion Code: X6.

Appendix C. Excluded Studies

267. Li G, Papaioannou A, Thabane L, et al. Modifying the phenotypic frailty model in predicting risk of major osteoporotic fracture in the elderly. *J Am Med Dir Assoc*. 2017 May 1;18(5):414-9. doi: 10.1016/j.jamda.2016.11.015. PMID: 28108205. Exclusion Code: X3.

268. Li GH, Cheung CL, Tan KC, et al. Development and validation of sex-specific hip fracture prediction models using electronic health records: a retrospective, population-based cohort study. *EClinicalMedicine*. 2023 Apr;58:101876. doi: 10.1016/j.eclinm.2023.101876. PMID: 36896245. Exclusion Code: X5.

269. Li J, Sun Y, Chen Z, et al. Effects of Bisphosphonates Treatments in Osteopenic Older Women: A Systematic Review and Meta-Analysis. *Front Pharmacol*. 2022;13:892091. doi: 10.3389/fphar.2022.892091. PMID: 35662708. Exclusion Code: X3.

270. Li P, Wu X, Li Y, Huang J. Denosumab Versus Bisphosphonates for the Prevention of the Vertebral Fractures in Men with Osteoporosis: An Updated Network Meta-Analysis. *Clin Invest Med*. 2022 Sep 21;45(3):E14-22. doi: 10.25011/cim.v45i3.38875. PMID: 36149052. Exclusion Code: X3.

271. Li Z, Zhao W, Lin X, Li F. AI algorithms for accurate prediction of osteoporotic fractures in patients with diabetes: an up-to-date review. *J Orthop Surg Res*. 2023 Dec 12;18(1):956. doi: 10.1186/s13018-023-04446-5. PMID: 38087332. Exclusion Code: X2.

272. Liang BC, Shi ZY, Wang B, et al. Intravenous zoledronic acid 5 mg on bone turnover markers and bone mineral density in East China subjects with newly diagnosed osteoporosis: a 24-month clinical study. *Orthop Surg*. 2017 Feb;9(1):103-9. doi: 10.1111/os.12307. PMID: 28276638. Exclusion Code: X8.

273. Lin J, Yang Y, Zhang X, et al. BFH-OSTM, a new predictive screening tool for identifying osteoporosis in elderly Han Chinese males. *Clin Interv Aging*. 2017;12:1167-74. doi: 10.2147/cia.S140553. PMID: 28814842. Exclusion Code: X8.

274. Lin SY, Hung MC, Chang SF, et al. Efficacy and Safety of Postmenopausal Osteoporosis Treatments: A Systematic Review and Network Meta-Analysis of Randomized Controlled Trials. *J Clin Med*. 2021 Jul 8;10(14)doi: 10.3390/jcm10143043. PMID: 34300210. Exclusion Code: X3.

275. Liu GF, Wang ZQ, Liu L, et al. A network meta-analysis on the short-term efficacy and adverse events of different anti-osteoporosis drugs for the treatment of postmenopausal osteoporosis. *J Cell Biochem*. 2018 Jun;119(6):4469-81. doi: 10.1002/jcb.26550. PMID: 29227547. Exclusion Code: X3.

276. Liu Q, Chen D, Ye Z, et al. Minodronate in the treatment of osteoporosis: a systematic review and meta-analysis. *Medicine (Baltimore)*. 2020 Oct 2;99(40):e22542. doi: 10.1097/md.0000000000022542. PMID: 33019463. Exclusion Code: X5.

277. Livingstone SJ, Guthrie B, McMinn M, et al. Derivation and validation of the CFracture competing risk fracture prediction tool compared with QFracture in older people and people with comorbidity: a population cohort study. *Lancet Healthy Longev*. 2023 Jan;4(1):e43-e53. doi: 10.1016/s2666-7568(22)00290-2. PMID: 36610448. Exclusion Code: X5.

278. Lix LM, Leslie WD, Majumdar SR. Measuring improvement in fracture risk prediction for a new risk factor: a simulation. *BMC Res Notes*. 2018 Jan 22;11(1):62. doi: 10.1186/s13104-018-3178-z. PMID: 29357907. Exclusion Code: X9.

279. Lo JC, Grimsrud CD, Ott SM, et al. Atypical femur fracture incidence in women increases with duration of bisphosphonate exposure. *Osteoporos Int*. 2019 Dec;30(12):2515-20. doi: 10.1007/s00198-019-05112-5. PMID: 31555883. Exclusion Code: X6.

280. Lo JC, Nath S, Patel M, et al. Variation in fracture risk estimation among East Asian women. *Am J Manag Care*. 2021 Mar 1;27(3):e97-e100. doi: 10.37765/ajmc.2021.88604. PMID: 33720675. Exclusion Code: X3.

281. Lo JC, Neugebauer RS, Ettinger B, et al. Risk of complete atypical femur fracture with oral bisphosphonate exposure beyond three years. *BMC Musculoskelet Disord*. 2020 Dec 3;21(1):801. doi: 10.1186/s12891-020-03672-w. PMID: 33272248. Exclusion Code: X6.

282. Logan S, Thu WPP, Lay WK, et al. Chronic joint pain and handgrip strength correlates with osteoporosis in mid-life women: a Singaporean cohort. *Osteoporos Int*. 2017 Sep;28(9):2633-43. doi: 10.1007/s00198-017-4095-z. PMID: 28647803. Exclusion Code: X3.

283. Looker AC, Sarafrazi Isfahani N, Fan B, Shepherd JA. FRAX-based estimates of 10-year probability of hip and major osteoporotic fracture among adults aged 40 and over: United States, 2013 and 2014. *Natl Health Stat Report*. 2017 Mar(103):1-16. PMID: 28459415. Exclusion Code: X6.

284. Lundin H, Sääf M, Strender LE, et al. Gait speed and one-leg standing time each add to the predictive ability of FRAX. *Osteoporos Int*. 2017 Jan;28(1):179-87. doi: 10.1007/s00198-016-3818-x. PMID: 27844133. Exclusion Code: X4.

285. Lundin H, Torabi F, Saaf M, et al. Laser-supported dual energy x-ray absorptiometry (DXL) compared to conventional absorptiometry (DXA) and to FRAX as tools for fracture risk assessments. *PLoS One*. 2015;10(9):e0137535. doi: 10.1371/journal.pone.0137535. PMID: 26413715. Exclusion Code: X4.

286. Lundin H, Torabi F, Sääf M, et al. Laser-Supported dual energy x-ray absorptiometry (DXL) compared to conventional absorptiometry (DXA) and to FRAX as tools for fracture risk assessments. *PLoS One*. 2015;10(9):e0137535. doi: 10.1371/journal.pone.0137535. PMID: 26413715. Exclusion Code: X4.

287. Lyu H, Yoshida K, Zhao SS, et al. Delayed denosumab injections and fracture risk among patients with osteoporosis: a population-based cohort study. *Ann Intern Med*. 2020 Oct 6;173(7):516-26. doi: 10.7326/m20-0882. PMID: 32716706. Exclusion Code: X7.

288. Ma Z, Yang Y, Lin J, et al. BFH-OST, a new predictive screening tool for identifying osteoporosis in postmenopausal Han Chinese women. *Clin Interv Aging*. 2016;11:1051-9. doi: 10.2147/cia.S107675. PMID: 27536085. Exclusion Code: X8.

289. Makras P, Papapoulos SE, Polyzos SA, et al. The three-year effect of a single zoledronate infusion on bone mineral density and bone turnover markers following denosumab discontinuation in women with postmenopausal osteoporosis. *Bone*. 2020 Sep;138:115478. doi: 10.1016/j.bone.2020.115478. PMID: 32534221. Exclusion Code: X6.

290. Martineau P, Leslie WD, Johansson H, et al. Clinical utility of using lumbar spine trabecular bone score to adjust fracture probability: the Manitoba BMD cohort. *J Bone Miner Res*. 2017 Jul;32(7):1568-74. doi: 10.1002/jbmr.3124. PMID: 28276598. Exclusion Code: X5.

291. Martins LHI, Ferreira DC, Silva MT, et al. Frequency of osteonecrosis in bisphosphonate users submitted to dental procedures: A systematic review. *Oral Dis*. 2023 Jan;29(1):75-99. doi: 10.1111/odi.14003. PMID: 34402147. Exclusion Code: X3.

292. Matsumoto T, Sone T, Soen S, et al. Abaloparatide Increases Lumbar Spine and Hip BMD in Japanese Patients With Osteoporosis: The Phase 3 ACTIVE-J Study. *J Clin Endocrinol Metab*. 2022 Sep 28;107(10):e4222-e31. doi: 10.1210/clinem/dgac486. PMID: 35977548. Exclusion Code: X5.

293. Mawatari T, Ikemura S, Matsui G, et al. Assessment of baseline bone turnover marker levels and response to risedronate treatment: data from a Japanese phase III trial. *Bone Rep*. 2020 Jun;12:100275. doi: 10.1016/j.bonr.2020.100275. PMID: 32462056. Exclusion Code: X6.

294. McCloskey EV, Fitzpatrick LA, Hu MY, et al. Effect of abaloparatide on vertebral, nonvertebral, major osteoporotic, and clinical fractures in a subset of postmenopausal women at increased risk of fracture by FRAX probability. *Arch Osteoporos*. 2019 Feb 5;14(1):15. doi: 10.1007/s11657-019-0564-7. PMID: 30719589. Exclusion Code: X5.

295. McCloskey EV, Johansson H, Harvey NC, et al. Romosozumab efficacy on fracture outcomes is greater in patients at high baseline fracture risk: a post hoc analysis of the first year of the frame study. *Osteoporos Int*. 2021 Feb 3;32(8):1601-8. doi: 10.1007/s00198-020-05815-0. PMID: 33537844. Exclusion Code: X5.

296. McCloskey EV, Johansson H, Oden A, et al. The effect of abaloparatide-SC on fracture risk is independent of baseline FRAX fracture probability: a post hoc analysis of the ACTIVE study. *J Bone Miner Res*. 2017 Aug;32(8):1625-31. doi: 10.1002/jbmr.3163. PMID: 28474780. Exclusion Code: X2.

297. McCloskey EV, Odén A, Harvey NC, et al. A meta-analysis of trabecular bone score in fracture risk prediction and its relationship to FRAX. *J Bone Miner Res*. 2016 May;31(5):940-8. doi: 10.1002/jbmr.2734. PMID: 26498132. Exclusion Code: X5.

298. McClung MR, Bolognese MA, Brown JP, et al. Skeletal responses to romosozumab after 12 months of denosumab. *JBMR Plus*. 2021 Jul;5(7):e10512. doi: 10.1002/jbm4.10512. PMID: 34258507. Exclusion Code: X5.

299. McClung MR, Brown JP, Diez-Perez A, et al. Effects of 24 months of treatment with romosozumab followed by 12 months of denosumab or placebo in postmenopausal women with low bone mineral density: a randomized, double-blind, phase 2, parallel group study. *J Bone Miner Res*. 2018 Aug;33(8):1397-406. doi: 10.1002/jbmr.3452. PMID: 29694685. Exclusion Code: X7.

300. McClung MR, Harvey NC, Fitzpatrick LA, et al. Effects of abaloparatide on bone mineral density and risk of fracture in postmenopausal women aged 80 years or older with osteoporosis. *Menopause*. 2018 Jul;25(7):767-71. doi: 10.1097/gme.0000000000001080. PMID: 29462094. Exclusion Code: X5.

301. McClung MR, Lippuner K, Brandi ML, et al. Effect of denosumab on trabecular bone score in postmenopausal women with osteoporosis. *Osteoporos Int*. 2017 Oct;28(10):2967-73. doi: 10.1007/s00198-017-4140-y. PMID: 28748386. Exclusion Code: X7.

302. McClung MR, Siris E, Cummings S, et al. Prevention of bone loss in postmenopausal women treated with lasofoxifene compared with raloxifene. *Menopause*. 2006 May-Jun;13(3):377-86. doi: 10.1097/01.gme.0000188736.69617.4f. PMID: 16735934. Exclusion Code: X5.

303. McClung MR, Wagman RB, Miller PD, et al. Observations following discontinuation of long-term denosumab therapy. *Osteoporos Int*. 2017 May;28(5):1723-32. doi: 10.1007/s00198-017-3919-1. PMID: 28144701. Exclusion Code: X6.

304. McClung MR, Williams GC, Hattersley G, et al. Geography of fracture incidence in postmenopausal women with osteoporosis treated with abaloparatide. *Calcif Tissue Int*. 2018 Jun;102(6):627-33. doi: 10.1007/s00223-017-0375-z. PMID: 29285549. Exclusion Code: X5.

305. McGowan K, Acton C, Ivanovski S, et al. Systemic comorbidities are associated with medication-related osteonecrosis of the jaws: case-control study. *Oral Dis*. 2019 May;25(4):1107-15. doi: 10.1111/odi.13046. PMID: 30674072. Exclusion Code: X5.

306. Mendes D, Penedones A, Alves C, et al. Ibandronate in the Prevention of Vertebral and Nonvertebral Osteoporotic Fractures: A Systematic Review of Experimental and Observational Studies. *J Clin Rheumatol*. 2023 Mar 1;29(2):78-83. doi: 10.1097/rhu.0000000000001902. PMID: 36731043. Exclusion Code: X3.

307. Meng J, Sun N, Chen Y, et al. Artificial neural network optimizes self-examination of osteoporosis risk in women. *J Int Med Res*. 2019 Jul;47(7):3088-98. doi: 10.1177/0300060519850648. PMID: 31179797. Exclusion Code: X8.

308. Meunier PJ, Confavreux E, Tupinon I, et al. Prevention of early postmenopausal bone loss with cyclical etidronate therapy (a double-blind, placebo-controlled study and 1-year follow-up). *J Clin Endocrinol Metab*. 1997 Sep;82(9):2784-91. doi: 10.1210/jcem.82.9.4073. PMID: 9284696. Exclusion Code: X5.

309. Meunier PJ, Vignot E, Garnero P, et al. Treatment of postmenopausal women with osteoporosis or low bone density with raloxifene. Raloxifene study group. *Osteoporos Int*. 1999;10(4):330-6. doi: 10.1007/s001980050236. PMID: 10692984. Exclusion Code: X5.

310. Migliorini F, Colarossi G, Baroncini A, et al. Pharmacological management of postmenopausal osteoporosis: a level I evidence based - expert opinion. *Expert Rev Clin Pharmacol*. 2021 Jan;14(1):105-19. doi: 10.1080/17512433.2021.1851192. PMID: 33183112. Exclusion Code: X1.

311. Miller PD, Chines AA, Christiansen C, et al. Effects of bazedoxifene on BMD and bone turnover in postmenopausal women: 2-yr results of a randomized, double-blind, placebo-, and active-controlled study. *J Bone Miner Res*. 2008 Apr;23(4):525-35. doi: 10.1359/jbmr.071206. PMID: 18072873. Exclusion Code: X5.

312. Miller PD, Hattersley G, Lau E, et al. Bone mineral density response rates are greater in patients treated with abaloparatide compared with those treated with placebo or teriparatide: results from the ACTIVE phase 3 trial. *Bone*. 2019 Mar;120:137-40. doi: 10.1016/j.bone.2018.10.015. PMID: 30359763. Exclusion Code: X7.

313. Miller PD, Hattersley G, Riis BJ, et al. Effect of abaloparatide vs placebo on new vertebral fractures in postmenopausal women with osteoporosis: a randomized clinical trial. *JAMA*. 2016 Aug 16;316(7):722-33. doi: 10.1001/jama.2016.11136. PMID: 27533157. Exclusion Code: X5.

314. Miller PD, Schnitzer T, Emkey R, et al. Weekly oral alendronic Acid in male osteoporosis. *Clin Drug Investig*. 2004;24(6):333-41. doi: 10.2165/00044011-200424060-00003. PMID: 17516720. Exclusion Code: X2.

315. Miller SA, St Onge EL, Whalen KL. Romosozumab: A novel agent in the treatment for postmenopausal osteoporosis. *J Pharm Technol*. 2021 Feb;37(1):45-52. doi: 10.1177/8755122520967632. PMID: 34752536. Exclusion Code: X3.

316. Mirzaei A, Jahed SA, Nojomi M, et al. A study of the value of trabecular bone score in fracture risk assessment of postmenopausal women. *Taiwan J Obstet Gynecol*. 2018 Jun;57(3):389-93. doi: 10.1016/j.tjog.2018.04.011. PMID: 29880171. Exclusion Code: X8.

317. Mitsuboshi S, Kotake K. Risks of serious adverse events and kidney injury in patients treated with ibandronate: A systematic review and meta-analysis. *Pharmacotherapy*. 2022 Aug;42(8):677-86. doi: 10.1002/phar.2713. PMID: 35716279. Exclusion Code: X3.

318. Miyauchi A, Dinavahi RV, Crittenden DB, et al. Increased bone mineral density for 1 year of romosozumab, vs placebo, followed by 2 years of denosumab in the Japanese subgroup of the pivotal FRAME trial and extension. *Arch Osteoporos*. 2019 Jun 5;14(1):59. doi: 10.1007/s11657-019-0608-z. PMID: 31168657. Exclusion Code: X5.

319. Miyauchi A, Hamaya E, Yang W, et al. Romosozumab followed by denosumab in Japanese women with high fracture risk in the FRAME trial. *J Bone Miner Metab*. 2021 Mar;39(2):278-88. doi: 10.1007/s00774-020-01147-5. PMID: 33057807. Exclusion Code: X5.

320. Modi A, Ebeling PR, Lee MS, et al. Impact of gastrointestinal events on patient-reported outcomes in Asia-Pacific women with osteoporosis: baseline results of the MUSIC OS-AP study. *Arch Osteoporos*. 2017 Dec;12(1):65. doi: 10.1007/s11657-017-0350-3. PMID: 28718004. Exclusion Code: X3.

321. Modi A, Fan CPS, Tang J, et al. Association of gastrointestinal events with osteoporosis treatment initiation and treatment compliance in Germany: an observational study. *Bone Reports*. 2016 Dec;5:208-13. doi: 10.1016/j.bonr.2016.06.001. PMID: 28580388. Exclusion Code: X3.

322. Modi A, Sajjan S, Michael Lewiecki E, et al. Relationship between gastrointestinal events and compliance with osteoporosis therapy: an administrative claims analysis of the US managed care population. *Clin Ther*. 2016 May;38(5):1074-80. doi: 10.1016/j.clinthera.2016.03.027. PMID: 27112533. Exclusion Code: X3.

323. Modi A, Sen S, Adachi JD, et al. Gastrointestinal symptoms and association with medication use patterns, adherence, treatment satisfaction, quality of life, and resource use in osteoporosis: baseline results of the MUSIC-OS study. *Osteoporos Int*. 2016 Mar;27(3):1227-38. doi: 10.1007/s00198-015-3388-3. PMID: 26637321. Exclusion Code: X3.

324. Modi A, Sen S, Adachi JD, et al. The impact of GI events on persistence and adherence to osteoporosis treatment: 3-, 6-, and 12-month findings in the MUSIC-OS study. *Osteoporos Int*. 2018 Feb;29(2):329-37. doi: 10.1007/s00198-017-4271-1. PMID: 29110061. Exclusion Code: X6.

325. Mohammed Abou-Hashem R, Mostafa Abd-El-Gawad W, Abouseif HA. Modified SCORE tool for prediction of bone mineral density in Egyptian elderly women. Comparison of seven osteoporosis screening tools. *Health Care Women Int*. 2019 Oct;40(10):1084-100. doi: 10.1080/07399332.2018.1542433. PMID: 30526422. Exclusion Code: X8.

326. Mok J, Brown C, Moore AEB, et al. Skeletal response to treatment with teriparatide (TPD) after bisphosphonate in post-menopausal women with osteoporosis and a high prevalence of secondary risk factors in real-life setting of a metabolic bone clinic; effect of age and vitamin D status. *Endocr Res*. 2018 Aug;43(3):195-202. doi: 10.1080/07435800.2018.1454461. PMID: 29652557. Exclusion Code: X6.

327. Momeni M, Asadzadeh M, Mowla K, et al. Sensitivity and specificity assessment of DWI and ADC for the diagnosis of osteoporosis in postmenopausal patients. *Radiol Med*. 2020 Jan;125(1):68-74. doi: 10.1007/s11547-019-01080-2. PMID: 31531809. Exclusion Code: X5.

328. Moreira CA, Fitzpatrick LA, Wang Y, Recker RR. Effects of abaloparatide-SC (BA058) on bone histology and histomorphometry: the ACTIVE phase 3 trial. *Bone*. 2017 Apr;97:314-9. doi: 10.1016/j.bone.2016.11.004. PMID: 27826127. Exclusion Code: X7.

329. Morii H, Ohashi Y, Taketani Y, et al. Effect of raloxifene on bone mineral density and biochemical markers of bone turnover in Japanese postmenopausal women with osteoporosis: results from a randomized placebo-controlled trial. *Osteoporos Int*. 2003 Oct;14(10):793-800. doi: 10.1007/s00198-003-1424-1. PMID: 12955333. Exclusion Code: X5.

330. Moshi MR, Nicolopoulos K, Stringer D, et al. The clinical effectiveness of denosumab (Prolia®) for the treatment of osteoporosis in postmenopausal women, compared to bisphosphonates, selective estrogen receptor modulators (SERM), and placebo: a systematic review and network meta-analysis. *Calcif Tissue Int*. 2023 Jun;112(6):631-46. doi: 10.1007/s00223-023-01078-z. PMID: 37016189. Exclusion Code: X3.

331. Mueller YK, Monod S, Locatelli I, et al. Performance of a brief geriatric evaluation compared to a comprehensive geriatric assessment for detection of geriatric syndromes in family medicine: a prospective diagnostic study. *BMC Geriatr*. 2018 Mar 13;18(1):72. doi: 10.1186/s12877-018-0761-z. PMID: 29534680. Exclusion Code: X5.

332. Nagashima H, Hirayama M, Ohishi H, Oikawa I. Optimizing a single monthly dose of risedronate based on its weekly dose in healthy Japanese postmenopausal women. *Japanese J Clin Pharmacol Ther*. 2017;48(4):121-9. doi: 10.3999/jscpt.48.121. PMID: CN-01406016. Exclusion Code: X5.

333. Nahar VK, Nelson KM, Ford MA, et al. Predictors of bone mineral density among Asian Indians in northern Mississippi: a pilot study. *J Res Health Sci*. 2016 fall;16(4):228-32. PMID: 28087857. Exclusion Code: X5.

334. Nakamura T, Matsumoto T, Sugimoto T, et al. Clinical trials express: fracture risk reduction with denosumab in Japanese postmenopausal women and men with osteoporosis: denosumab fracture intervention randomized placebo controlled trial (DIRECT). *J Clin Endocrinol Metab*. 2014 Jul;99(7):2599-607. doi: 10.1210/jc.2013-4175 [doi]. PMID: 24646104. Exclusion Code: X2.

335. Nakano T, Yamamoto M, Hashimoto J, et al. Higher response with bone mineral density increase with monthly injectable ibandronate 1 mg compared with oral risedronate in the MOVER study. *J Bone Miner Metab*. 2016 Nov;34(6):678-84. doi: 10.1007/s00774-015-0717-8. PMID: 26462480. Exclusion Code: X6.

336. Napoli N, Pannacciulli N, Vittinghoff E, et al. Effect of denosumab on fasting glucose in women with diabetes or prediabetes from the FREEDOM trial. *Diabetes Metab Res Rev*. 2018 May;34(4):e2991. doi: 10.1002/dmrr.2991. PMID: 29430796. Exclusion Code: X7.

337. Nayak S, Greenspan SL. Osteoporosis treatment efficacy for men: a systematic review and meta-analysis. *J Am Geriatr Soc*. 2017 Mar;65(3):490-5. doi: 10.1111/jgs.14668. PMID: 28304090. Exclusion Code: X3.

338. NCT02791516. A safety and efficacy study to evaluate AMG 785 in South Korean women with osteoporosis. Exclusion Code: X1.

339. NCT03512262. Safety and efficacy of abaloparatide-SC in men with osteoporosis (ATOM). 2018. Exclusion Code: X1.

340. NCT04249232. Abaloparatide and pelvic fracture healing. 2020. Exclusion Code: X2.

341. Nguyen ND, Frost SA, Center JR, et al. Development of a nomogram for individualizing hip fracture risk in men and women. *Osteoporos Int*. 2007 Aug;18(8):1109-17. doi: 10.1007/s00198-007-0362-8. PMID: 17370100. Exclusion Code: X7.

342. Nguyen ND, Frost SA, Center JR, et al. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. *Osteoporos Int*. 2008 Oct;19(10):1431-44. doi: 10.1007/s00198-008-0588-0. PMID: 18324342. Exclusion Code: X3.

343. Niznik JD, Li X, Gilliam MA, et al. Are nursing home residents with dementia appropriately treated for fracture prevention? *J Am Med Dir Assoc*. 2021 Jan;22(1):28-35.e3. doi: 10.1016/j.jamda.2020.11.019. PMID: 33321079. Exclusion Code: X2.

344. Nordström P, Bergman J, Ballin M, et al. Bone-specific drugs and osteonecrosis of sites other than the jaw: a nationwide cohort study. *J Bone Miner Res*. 2020 Sep;35(9):1703-10. doi: 10.1002/jbmr.4040. PMID: 32379370. Exclusion Code: X10.

345. Oh SM, Song BM, Nam BH, et al. Development and validation of osteoporosis risk-assessment model for Korean men. *Yonsei Med J*. 2016 Jan;57(1):187-96. doi: 10.3349/ymj.2016.57.1.187. PMID: 26632400. Exclusion Code: X6.

346. Oka R, Ohira M, Suzuki S, et al. Fracture risk assessment tool (FRAX) and for the diagnosis of osteoporosis in Japanese middle-aged and elderly women: Chiba bone survey. *Endocr J*. 2018 Feb 26;65(2):193-202. doi: 10.1507/endocrj.EJ17-0331. PMID: 29151451. Exclusion Code: X6.

347. Okazaki R, Muraoka R, Maehara M, Inoue D. Factors associated with inadequate responses to risedronate in Japanese patients with osteoporosis. *J Bone Miner Metab*. 2019 Jan;37(1):185-97. doi: 10.1007/s00774-018-0931-2. PMID: 29737412. Exclusion Code: X6.

348. Okubo N, Matsui S, Matsumoto T, et al. Relationship between bone mineral density and risk of vertebral fractures with denosumab treatment in Japanese postmenopausal women and men with osteoporosis. *Calcif Tissue Int*. 2020 Dec;107(6):559-66. doi: 10.1007/s00223-020-00750-y. PMID: 32839843. Exclusion Code: X2.

349. Orwoll ES, Scheele WH, Paul S, et al. The effect of teriparatide [human parathyroid hormone (1-34)] therapy on bone density in men with osteoporosis. *J Bone Miner Res*. 2003 Jan;18(1):9-17. doi: 10.1359/jbmr.2003.18.1.9. PMID: 12510800. Exclusion Code: X5.

350. Osawa R, Ikegami S, Horiuchi H, et al. Osteoporosis detection by physical function tests in resident health exams: a japanese cohort survey randomly sampled from a basic resident registry. *Journal of Clinical Medicine*. 2021 Apr 27;10(9)doi: 10.3390/jcm10091896. PMID: 33925580. Exclusion Code: X5.

351. Oswald AJ, Berg K, Ralston SH, Riches PL. Long-term effects of teriparatide followed by antiresorptive therapy on clinical outcomes in patients with severe spinal osteoporosis. *Calcif Tissue Int*. 2019 Aug;105(2):148-55. doi: 10.1007/s00223-019-00563-8. PMID: 31115640. Exclusion Code: X3.

352. Papapoulos S, Lippuner K, Roux C, et al. The effect of 8 or 5 years of denosumab treatment in postmenopausal women with osteoporosis: results from the FREEDOM Extension study. *Osteoporos Int*. 2015 Dec;26(12):2773-83. doi: 10.1007/s00198-015-3234-7. PMID: 26202488. Exclusion Code: X6.

353. Park JH, Ko HJ. The Association between Treatment with Bisphosphonates and the Risk of Atrial Fibrillation: A Meta-Analysis of Observational Studies. *Korean J Fam Med*. 2022 Jan;43(1):69-76. doi: 10.4082/kjfm.21.0110. PMID: 35130642. Exclusion Code: X3.

354. Passarelli MN, Newcomb PA, LaCroix AZ, et al. Oral bisphosphonate use and colorectal cancer incidence in the women's health initiative. *J Bone Miner Res*. 2013 Sep;28(9):2043-8. doi: 10.1002/jbmr.1930. PMID: 23519920. Exclusion Code: X10.

355. Passeri E, Mazzaccaro D, Sansoni V, et al. Effects of 12-months treatment with zoledronate or teriparatide on intima-media thickness of carotid artery in women with postmenopausal osteoporosis: a pilot study. *Int J Immunopathol Pharmacol*. 2019 Jan-Dec;33:2058738418822439. doi: 10.1177/2058738418822439. PMID: 30791743. Exclusion Code: X3.

356. Pedrassani de Lira C, Toniazzo de Abreu LL, Veiga Silva AC, et al. Use of data mining to predict the risk factors associated with osteoporosis and osteopenia in women. *Comput Inform Nurs*. 2016 Aug;34(8):369-75. doi: 10.1097/cin.0000000000000253. PMID: 27270629. Exclusion Code: X3.

357. Pineda-Moncusí M, El-Hussein L, Delmestri A, et al. Estimating the Incidence and Key Risk Factors of Cardiovascular Disease in Patients at High Risk of Imminent Fracture Using Routinely Collected Real-World Data From the UK. *J Bone Miner Res*. 2022 Oct;37(10):1986-96. doi: 10.1002/jbmr.4648. PMID: 35818312. Exclusion Code: X5.

358. Pizzato S, Trevisan C, Lucato P, et al. Identification of asymptomatic frailty vertebral fractures in post-menopausal women. *Bone*. 2018 Aug;113:89-94. doi: 10.1016/j.bone.2018.05.007. PMID: 29753150. Exclusion Code: X5.

359. Pluskiewicz W, Adamczyk P, Czekajło A, et al. High fracture probability predicts fractures in a 4-year follow-up in women from the RAC-OST-POL study. *Osteoporos Int*. 2015 Dec;26(12):2811-20. doi: 10.1007/s00198-015-3196-9. PMID: 26168766. Exclusion Code: X3.

360. Pluskiewicz W, Adamczyk P, Drozdzowska B. Glucocorticoids Increase Fracture Risk and Fracture Prevalence Independently from Bone Mineral Density and Clinical Risk Factors: Results from the Gliwice Osteoporosis (GO) Study. *Horm Metab Res*. 2022 Jan;54(1):20-4. doi: 10.1055/a-1700-5007. PMID: 34986496. Exclusion Code: X7.

361. Pluskiewicz W, Adamczyk P, Franek E, et al. The efficacy of pharmacotherapy in postmenopausal osteoporosis: a longitudinal observational study. *Endokrynol Pol*. 2019;70(6):473-7. doi: 10.5603/EP.a2019.0058. PMID: 31909456. Exclusion Code: X7.

362. Pluskiewicz W, Adamczyk P, Werner A, et al. POL-RISK: an algorithm for 10-year fracture risk prediction in postmenopausal women from the RAC-OST-POL study. *Pol Arch Intern Med*. 2023 Mar 29;133(3)doi: 10.20452/pamw.16395. PMID: 36601872. Exclusion Code: X5.

363. Popp AW, Varathan N, Buffat H, et al. Bone mineral density changes after 1 year of denosumab discontinuation in postmenopausal women with long-term denosumab treatment for osteoporosis. *Calcif Tissue Int*. 2018 Jul;103(1):50-4. doi: 10.1007/s00223-018-0394-4. PMID: 29380013. Exclusion Code: X6.

364. Poutoglidou F, Samoladas E, Raikos N, Kouvelas D. Efficacy and safety of anti-sclerostin antibodies in the treatment of osteoporosis: A meta-analysis and systematic review. *J Clin Densitom*. 2022 Jul-Sep;25(3):401-15. doi: 10.1016/j.jocd.2021.11.005. PMID: 34920938. Exclusion Code: X3.

365. Poxleitner P, Engelhardt M, Schmelzeisen R, Voss P. The prevention of medication-related osteonecrosis of the jaw. *Dtsch Arztebl Int*. 2017 Feb 3;114(5):63-9. doi: 10.3238/arztebl.2017.0063. PMID: 28241916. Exclusion Code: X3.

366. Prochaska M, Taylor E, Vaidya A, Curhan G. Low bone density and bisphosphonate use and the risk of kidney stones. *Clin J Am Soc Nephrol*. 2017 Aug 7;12(8):1284-90. doi: 10.2215/cjn.01420217. PMID: 28576907. Exclusion Code: X7.

367. Radeva M, Predel D, Winzler S, et al. Reliability of a risk-factor questionnaire for osteoporosis: a primary care survey study with dual energy x-ray absorptiometry ground truth. *Int J Environ Res Public Health*. 2021 Jan 28;18(3)doi: 10.3390/ijerph18031136. PMID: 33525339. Exclusion Code: X7.

368. Rahim F, Zaki Zadeh A, Javanmardi P, et al. Machine learning algorithms for diagnosis of hip bone osteoporosis: a systematic review and meta-analysis study. *Biomed Eng Online*. 2023 Jul 10;22(1):68. doi: 10.1186/s12938-023-01132-9. PMID: 37430259. Exclusion Code: X5.

369. Real J, Galindo G, Galván L, et al. Use of oral bisphosphonates in primary prevention of fractures in postmenopausal women: a population-based cohort study. *PLoS One*. 2015;10(4):e0118178. doi: 10.1371/journal.pone.0118178. PMID: 25861000. Exclusion Code: X3.

370. Reber KC, König HH, Becker C, et al. Development of a risk assessment tool for osteoporotic fracture prevention: a claims data approach. *Bone*. 2018 May;110:170-6. doi: 10.1016/j.bone.2018.02.002. PMID: 29421456. Exclusion Code: X3.

371. Reginster J, Minne HW, Sorensen OH, et al. Randomized trial of the effects of risedronate on vertebral fractures in women with established postmenopausal osteoporosis. Vertebral Efficacy with Risedronate Therapy (VERT) Study Group. *Osteoporos Int*. 2000;11(1):83-91. doi: 10.1007/s001980050010. PMID: 10663363. Exclusion Code: X1.

372. Reginster JY, Hattersley G, Williams GC, et al. Abaloparatide is an effective treatment option for postmenopausal osteoporosis: review of the number needed to treat compared with teriparatide. *Calcif Tissue Int*. 2018 Nov;103(5):540-5. doi: 10.1007/s00223-018-0450-0. PMID: 29951742. Exclusion Code: X3.

373. Reid IR, Horne AM, Mihov B, et al. Zoledronate slows weight loss and maintains fat mass in osteopenic older women: secondary analysis of a randomized controlled trial. *Calcif Tissue Int*. 2020 Apr;106(4):386-91. doi: 10.1007/s00223-019-00653-7. PMID: 31897528. Exclusion Code: X7.

374. Reid IR, Horne AM, Mihov B, et al. Effect of zoledronate on lower respiratory infections in older women: secondary analysis of a randomized controlled trial. *Calcif Tissue Int*. 2021 Jul;109(1):12-6. doi: 10.1007/s00223-021-00830-7. PMID: 33712919. Exclusion Code: X7.

375. Reyes Domínguez A, Sosa Cabrera N, Saavedra Santana P, et al. Assessment of the predictive capacity of the Garvan calculator of 10 year risk of fracture in a Spanish population. *Revista de Osteoporosis y Metabolismo Mineral*. 2017;9:55-61. Exclusion Code: X9.

376. Riar S, Feasel AL, Aghajafari F, et al. Comparison of 2 fracture risk estimation processes in Alberta: a cross-sectional chart review. *CMAJ Open*. 2021 Apr-Jun;9(2):E711-e7. doi: 10.9778/cmajo.20200207. PMID: 34162663. Exclusion Code: X3.

377. Ripamonti C, Lisi L, Ciaffi J, et al. Spine Fragility Fracture Prediction Using TBS and BMD in Postmenopausal Women: A Bayesian Approach. *Int J Environ Res Public Health*. 2022 Nov 2;19(21)doi: 10.3390/ijerph192114315. PMID: 36361195. Exclusion Code: X5.

378. Rizou S, Aravantinos L, Paganas A, et al. Assessment of the antiosteoporotic treatment initiation: a retrospective observational study in Greece. *J Frailty Sarcopenia Falls*. 2017 Dec;2(4):78-82. doi: 10.22540/jfsf-02-078. PMID: 32300684. Exclusion Code: X7.

379. Rocha VM, Gaspar HA, Oliveira CF. Fracture risk assessment in home care patients using the FRAX® tool. *Einstein (Sao Paulo)*. 2018 Sep 6;16(3):eAO4236. doi: 10.1590/s1679-45082018ao4236. PMID: 30208152. Exclusion Code: X7.

380. Rodríguez AJ, Abrahamsen B. Cardiovascular safety of antifracture medications in patients with osteoporosis: a narrative review of evidence from randomized studies. *JBMR Plus*. 2021 Jul;5(7):e10522. doi: 10.1002/jbm4.10522. PMID: 34258509. Exclusion Code: X3.

381. Rodríguez AJ, Ernst MT, Nybo M, et al. Oral bisphosphonate use reduces cardiovascular events in a cohort of Danish patients referred for bone mineral density. *J Clin Endocrinol Metab*. 2020 Oct 1;105(10)doi: 10.1210/clinem/dgaa481. PMID: 32717068. Exclusion Code: X10.

382. Rossini M, Gatti D, Zamberlan N, et al. Long-term effects of a treatment course with oral alendronate of postmenopausal osteoporosis. *J Bone Miner Res*. 1994 Nov;9(11):1833-7. doi: 10.1002/jbmr.5650091121. PMID: 7863833. Exclusion Code: X10.

383. Rubin KH, Abrahamsen B, Friis-Holmberg T, et al. Comparison of different screening tools (FRAX(R), OST, ORAI, OSIRIS, SCORE and age alone) to identify women with increased risk of fracture. A population-based prospective study. *Bone*. 2013 Sep;56(1):16-22. doi: 10.1016/j.bone.2013.05.002. PMID: 23669650. Exclusion Code: X4.

384. Rubin KH, Möller S, Holmberg T, et al. A new fracture risk assessment tool (FREM) based on public health registries. *J Bone Miner Res*. 2018 Nov;33(11):1967-79. doi: 10.1002/jbmr.3528. PMID: 29924428. Exclusion Code: X3.

385. Saag KG, Williams SA, Wang Y, et al. Effect of abaloparatide on bone mineral density and fracture incidence in a subset of younger postmenopausal women with osteoporosis at high risk for fracture. *Clin Ther*. 2020 Jun;42(6):1099-107.e1. doi: 10.1016/j.clinthera.2020.04.012. PMID: 32513495. Exclusion Code: X5.

386. Sambrook PN, Flahive J, Hooven FH, et al. Predicting fractures in an international cohort using risk factor algorithms without BMD. *J Bone Miner Res*. 2011 Nov;26(11):2770-7. doi: 10.1002/jbmr.503. PMID: 21887705. Exclusion Code: X9.

387. Sanderson J, Martyn-St James M, Stevens J, et al. Clinical effectiveness of bisphosphonates for the prevention of fragility fractures: a systematic review and network meta-analysis. *Bone*. 2016 Aug;89:52-8. doi: 10.1016/j.bone.2016.05.013. PMID: 27262775. Exclusion Code: X3.

388. Sandhu SK, Nguyen ND, Center JR, et al. Prognosis of fracture: evaluation of predictive accuracy of the FRAX algorithm and Garvan nomogram. *Osteoporos Int*. 2010 May;21(5):863-71. doi: 10.1007/s00198-009-1026-7. PMID: 19633880. Exclusion Code: X9.

389. Satyaraddi A, Shetty S, Kapoor N, et al. Performance of risk assessment tools for predicting osteoporosis in south Indian rural elderly men. *Arch Osteoporos*. 2017 Dec;12(1):35. doi: 10.1007/s11657-017-0332-5. PMID: 28378274. Exclusion Code: X8.

390. Seeto AH, Tadrous M, Gebre AK, et al. Evidence for the cardiovascular effects of osteoporosis treatments in randomized trials of post-menopausal women: A systematic review and Bayesian network meta-analysis. *Bone*. 2022 Nov 10:116610. doi: 10.1016/j.bone.2022.116610. PMID: 36372197. Exclusion Code: X3.

391. Sheer RL, Barron RL, Sudharshan L, Pasquale MK. Validated prediction of imminent risk of fracture for older adults. *Am J Manag Care*. 2020 Mar 1;26(3):e91-e7. doi: 10.37765/ajmc.2020.42641. PMID: 32181621. Exclusion Code: X5.

392. Shen J, Ke Z, Dong S, et al. Pharmacological therapies for osteoporosis: a Bayesian network meta-analysis. *Med Sci Monit*. 2022 Apr 17;28:e935491. doi: 10.12659/msm.935491. PMID: 35430576. Exclusion Code: X3.

393. Shepherd AJ, Cass AR, Carlson CA, Ray L. Development and internal validation of the male osteoporosis risk estimation score. *Ann Fam Med*. 2007 Nov-Dec;5(6):540-6. doi: 10.1370/afm.753. PMID: 18025492. Exclusion Code: X9.

394. Shevroja E, Marques-Vidal P, Aubry-Rozier B, et al. Cohort profile: The OsteoLaus study. *Int J Epidemiol*. 2019 Aug 1;48(4):1046-7g. doi: 10.1093/ije/dyy276. PMID: 30590566. Exclusion Code: X7.

395. Shevroja E, Mo Costabella F, Gonzalez Rodriguez E, et al. The fracture predictive ability of lumbar spine BMD and TBS as calculated based on different combinations of the lumbar spine vertebrae. *Arch Osteoporos*. 2022 Jun 9;17(1):83. doi: 10.1007/s11657-022-01123-8. PMID: 35678937. Exclusion Code: X5.

396. Shi Z, Zhou H, Pan B, et al. Effectiveness of teriparatide on fracture healing: a systematic review and meta-analysis. *PLoS One*. 2016;11(12):e0168691. doi: 10.1371/journal.pone.0168691. PMID: 27997614. Exclusion Code: X3.

397. Shi ZY, Zhang XG, Li CW, et al. Effect of traditional Chinese medicine product, QiangGuYin, on bone mineral density and bone turnover in Chinese postmenopausal osteoporosis. *Evid Based Complement Alternat Med*. 2017;2017:6062707. doi: 10.1155/2017/6062707. PMID: 28512501. Exclusion Code: X6.

398. Shigematsu T, Muraoka R, Sugimoto T, Nishizawa Y. Risedronate therapy in patients with mild-to-moderate chronic kidney disease with osteoporosis: post-hoc analysis of data from the risedronate phase III clinical trials. *BMC Nephrol*. 2017 Feb 15;18(1):66. doi: 10.1186/s12882-017-0478-9. PMID: 28201994. Exclusion Code: X6.

399. Shim JG, Kim DW, Ryu KH, et al. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch Osteoporos*. 2020 Oct 23;15(1):169. doi: 10.1007/s11657-020-00802-8. PMID: 33097976. Exclusion Code: X3.

400. Shimizu T, Arita K, Murota E, et al. Effects after starting or switching from bisphosphonate to romosozumab or denosumab in Japanese postmenopausal patients. *J Bone Miner Metab*. 2021 Apr 13;39(5):1-8. doi: 10.1007/s00774-021-01226-1. PMID: 33847831. Exclusion Code: X6.

401. Shin YH, Shin WC, Kim JW. Effect of osteoporosis medication on fracture healing: an evidence based review. *J Bone Metab*. 2020 Feb;27(1):15-26. doi: 10.11005/jbm.2020.27.1.15. PMID: 32190605. Exclusion Code: X2.

402. Silverman SL, Wang A, Cheng L, et al. Comorbidity indices for clinical trials: adaptation of two existing indices for use with the FREEDOM trial in women with postmenopausal osteoporosis. *Osteoporos Int*. 2016 Jan;27(1):75-80. doi: 10.1007/s00198-015-3215-x. PMID: 26174879. Exclusion Code: X3.

403. Singh S, Dutta S, Khasbage S, et al. A systematic review and meta-analysis of efficacy and safety of romosozumab in postmenopausal osteoporosis. *Osteoporos Int*. 2022 Jan;33(1):1-12. doi: 10.1007/s00198-021-06095-y. PMID: 34432115. Exclusion Code: X1.

404. Siris E, McDermott M, Pannacciulli N, et al. Estimation of long-term efficacy of denosumab treatment in postmenopausal women with osteoporosis: a FRAX- and virtual twin-based post hoc analysis from the FREEDOM and FREEDOM extension trials. *JBMR Plus*. 2020 Apr;4(4):e10348. doi: 10.1002/jbm4.10348. PMID: 32258966. Exclusion Code: X6.

405. Siris ES, Fan CPS, Yang X, et al. Association between gastrointestinal events and compliance with osteoporosis therapy. *Bone Reports*. 2016 Jun;4:5-10. doi: 10.1016/j.bonr.2015.10.006. PMID: 28326336. Exclusion Code: X6.

406. Skjødt MK, Möller S, Hyldig N, et al. Validation of the Fracture Risk Evaluation Model (FREM) in predicting major osteoporotic fractures and hip fractures using administrative health data. *Bone*. 2021 Jun;147:115934. doi: 10.1016/j.bone.2021.115934. PMID: 33757901. Exclusion Code: X3.

407. Soen S, Arai Y, Matsuda S, et al. A 3-year postmarketing study on the safety and effectiveness of once-monthly risedronate in Japanese patients with osteoporosis. *Osteoporosis and Sarcopenia*. 2020 Dec;6(4):191-8. doi: 10.1016/j.afos.2020.11.002. PMID: 33426308. Exclusion Code: X5.

408. Sone T, Kon N, Gaither KW, et al. Effects of 3-year denosumab treatment on hip structure in Japanese postmenopausal women and men with osteoporosis. *Bone Rep*. 2017 Dec;7:164-71. doi: 10.1016/j.bonr.2017.11.002. PMID: 29188222. Exclusion Code: X7.

409. Sontag A, Wan X, Krege JH. Benefits and risks of raloxifene by vertebral fracture status. *Curr Med Res Opin*. 2010 Jan;26(1):71-6. doi: 10.1185/03007990903427082 [doi]. PMID: 19908937. Exclusion Code: X5.

410. Sorensen HT, Christensen S, Mehnert F, et al. Use of bisphosphonates among women and risk of atrial fibrillation and flutter: population based case-control study. *BMJ*. 2008 Apr 12;336(7648):813-6. doi: 10.1136/bmj.39507.551644.BE. PMID: 18334527. Exclusion Code: X3.

411. Söreskog E, Borgström F, Shepstone L, et al. Long-term cost-effectiveness of screening for fracture risk in a UK primary care setting: the SCOOP study. *Osteoporos Int*. 2020 Aug;31(8):1499-506. doi: 10.1007/s00198-020-05372-6. PMID: 32239237. Exclusion Code: X3.

412. Steiner ML, Crotti GP, Teodoro JD, et al. Best laboratory screening in diagnosing secondary osteoporosis and fracture risk assessment tool and the national osteoporosis guideline group performance in determining clinical risk: a cross-sectional evaluation of the bone health in postmenopausal Brazilian women. *J Bone Metab*. 2023 Feb;30(1):47-57. doi: 10.11005/jbm.2023.30.1.47. PMID: 36950840. Exclusion Code: X8.

413. Ström O, Lauppe R, Ljunggren Ö, et al. Real-world effectiveness of osteoporosis treatment in the oldest old. *Osteoporos Int*. 2020 Aug;31(8):1525-33. doi: 10.1007/s00198-020-05380-6. PMID: 32232509. Exclusion Code: X3.

414. Su FM, Liu DH, Chen JF, et al. Development and validation of an Osteoporosis Self-Assessment Tool for Taiwan (OSTAi) postmenopausal women-a sub-study of the Taiwan OsteoPorosis Survey (TOPS). *PLoS One.* 2015;10(6):e0130716. doi: 10.1371/journal.pone.0130716. PMID: 26086766. Exclusion Code: X4.

415. Su Y, Kwok TCY, Cummings SR, et al. Can Classification and regression tree analysis help identify clinically meaningful risk groups for hip fracture prediction in older American men (The MrOS Cohort Study)? *JBMR Plus*. 2019 Oct;3(10):e10207. doi: 10.1002/jbm4.10207. PMID: 31687643. Exclusion Code: X9.

416. Su Y, Leung J, Hans D, et al. The added value of trabecular bone score to FRAX® to predict major osteoporotic fractures for clinical use in Chinese older people: the Mr. OS and Ms. OS cohort study in Hong Kong. *Osteoporos Int*. 2017 Jan;28(1):111-7. doi: 10.1007/s00198-016-3741-1. PMID: 27565645. Exclusion Code: X8.

417. Su Y, Leung J, Kwok T. The role of previous falls in major osteoporotic fracture prediction in conjunction with FRAX in older Chinese men and women: the Mr. OS and Ms. OS cohort study in Hong Kong. *Osteoporos Int*. 2018 Feb;29(2):355-63. doi: 10.1007/s00198-017-4277-8. PMID: 29067485. Exclusion Code: X8.

418. Su Y, Woo JW, Kwok TCY. The added value of SARC-F to prescreening using FRAX for hip fracture prevention in older community adults. *J Am Med Dir Assoc*. 2019 Jan;20(1):83-9. doi: 10.1016/j.jamda.2018.08.007. PMID: 30314676. Exclusion Code: X4.

419. Subasinghe H, Lekamwasam S, Ball P, et al. Performance of Sri Lankan FRAX algorithm without bone mineral density and with quantitative ultrasound data input. *Ceylon Med J*. 2019 Mar 31;64(1):17-24. doi: 10.4038/cmj.v64i1.8836. PMID: 31055903. Exclusion Code: X6.

420. Subasinghe H, Lekamwasam S, Ball P, et al. Estimating regional bone mineral density-based t-scores using clinical information; tools validated for postmenopausal women in Sri Lanka. *Osteoporos Sarcopenia*. 2020 Sep;6(3):122-8. doi: 10.1016/j.afos.2020.08.004. PMID: 33102805. Exclusion Code: X3.

421. Subramaniam S, Chan CY, Soelaiman IN, et al. Development of osteoporosis screening algorithm for population aged 50 years and above in Klang Valley, Malaysia. *Int J Environ Res Public Health*. 2020 Apr 7;17(7)doi: 10.3390/ijerph17072526. PMID: 32272697. Exclusion Code: X3.

422. Subramaniam S, Chan CY, Soelaiman IN, et al. The performance of osteoporosis self-assessment tool for Asians (OSTA) in identifying the risk of osteoporosis among Malaysian population aged 40 years and above. *Arch Osteoporos*. 2019 Nov 28;14(1):117. doi: 10.1007/s11657-019-0666-2. PMID: 31781876. Exclusion Code: X6.

423. Sugimoto T, Matsumoto T, Hosoi T, et al. Efficacy of denosumab co-administered with vitamin D and Ca by baseline vitamin D status. *J Bone Miner Metab*. 2020 Nov;38(6):848-58. doi: 10.1007/s00774-020-01119-9. PMID: 32671481. Exclusion Code: X3.

424. Sugimoto T, Yoshimura T, Uzawa T. Efficacy of once-weekly and twice-weekly injections of teriparatide by patient characteristics: a post hoc analysis of the TWICE study. *Osteoporos Sarcopenia*. 2021 Mar;7(1):11-6. doi: 10.1016/j.afos.2020.12.001. PMID: 33869800. Exclusion Code: X2.

425. Sun LM, Liang JA, Lin CL, et al. Cancer risk in patients with osteoporosis: a population-based cohort study. *Curr Med Res Opin*. 2017 Apr;33(4):733-9. doi: 10.1080/03007995.2017.1278681. PMID: 28044464. Exclusion Code: X5.

426. Sundh V, Hange D, Ahlqwist M, et al. FRAX and mandibular sparse trabeculation as fracture predictors: a longitudinal study from 1980 to 2002. *Eur J Oral Sci*. 2017 Apr;125(2):135-40. doi: 10.1111/eos.12341. PMID: 28263008. Exclusion Code: X4.

427. Tamaki J, Iki M, Sato Y, et al. Does Trabecular Bone Score (TBS) improve the predictive ability of FRAX(®) for major osteoporotic fractures according to the Japanese Population-based OSteoporosis (JPOS) cohort study? *J Bone Miner Metab*. 2019 Jan;37(1):161-70. doi: 10.1007/s00774-018-0910-7. PMID: 29468478. Exclusion Code: X9.

428. Tanaka S, Kuroda T, Saito M, Shiraki M. Urinary pentosidine improves risk classification using fracture risk assessment tools for postmenopausal women. *J Bone Miner Res*. 2011 Nov;26(11):2778-84. doi: 10.1002/jbmr.467 [doi]. PMID: 21773990. Exclusion Code: X7.

429. Tang JL, Pan BB, Shu Y, et al. Evaluation of two tools for the early screening of osteoporosis in postmenopausal Chinese women with type 2 diabetes mellitus. *J Int Med Res*. 2020 Mar;48(3):300060520903889. doi: 10.1177/0300060520903889. PMID: 32208853. Exclusion Code: X8.

430. Tang Y, Liu J, Tian C, et al. A novel primary osteoporosis screening tool (POST) for adults aged 50 years and over. *Endocrine*. 2023 Oct;82(1):190-200. doi: 10.1007/s12020-023-03442-3. PMID: 37450217. Exclusion Code: X8.

431. Tang Y, Liu Z, Wang S, et al. Development and validation of a novel screening tool for osteoporosis in older US adults: The NHANES cross-sectional study. *Endocrine*. 2022 May;76(2):446-56. doi: 10.1007/s12020-022-03001-2. PMID: 35122626. Exclusion Code: X5.

432. Thadani SR, Ristow B, Blackwell T, et al. Relationship of bisphosphonate therapy and atrial fibrillation/flutter: outcomes of Sleep Disorders in Older Men (MrOS Sleep) study. *Chest*. 2016 May;149(5):1173-80. doi: 10.1016/j.chest.2015.11.022. PMID: 26836889. Exclusion Code: X10.

433. Tinago W, Cotter AG, Sabin CA, et al. Predictors of longitudinal change in bone mineral density in a cohort of HIV-positive and negative patients. *AIDS*. 2017 Mar 13;31(5):643-52. doi: 10.1097/qad.0000000000001372. PMID: 28060010. Exclusion Code: X2.

434. Tourolle DC, Dempster DW, Ledoux C, et al. Ten-year simulation of the effects of denosumab on bone remodeling in human biopsies. *JBMR Plus*. 2021 Jun;5(6):e10494. doi: 10.1002/jbm4.10494. PMID: 34189383. Exclusion Code: X3.

435. Tran T, Bliuc D, Pham HM, et al. A risk assessment tool for predicting fragility fractures and mortality in the elderly. *J Bone Miner Res*. 2020 Oct;35(10):1923-34. doi: 10.1002/jbmr.4100. PMID: 32460361. Exclusion Code: X3.

436. Tripto-Shkolnik L, Fund N, Rouach V, et al. Fracture incidence after denosumab discontinuation: real-world data from a large healthcare provider. *Bone*. 2020 Jan;130:115150. doi: 10.1016/j.bone.2019.115150. PMID: 31715340. Exclusion Code: X3.

437. Tsai WH, Tsai MC, Chang CY, et al. Comparison of Cardiovascular Events among Users of Different Classes of Anti-Osteoporosis Medications. *International Journal of Gerontology*. 2023;17(4):275-9. doi: 10.6890/IJGE.202310_17(4).0012. Exclusion Code: X4.

438. Tsuda T, Hashimoto Y, Okamoto Y, et al. Meta-analysis for the efficacy of bisphosphonates on hip fracture prevention. *J Bone Miner Metab*. 2020 Sep;38(5):678-86. doi: 10.1007/s00774-020-01096-z. PMID: 32236684. Exclusion Code: X3.

439. Turner DA, Khioe RFS, Shepstone L, et al. The cost-effectiveness of screening in the community to reduce osteoporotic fractures in older women in the UK: economic evaluation of the SCOOP study. *J Bone Miner Res*. 2018 May;33(5):845-51. doi: 10.1002/jbmr.3381. PMID: 29470854. Exclusion Code: X7.

440. Ullah KA, Rehman F, Anwar M, et al. Machine learning-based prediction of osteoporosis in postmenopausal women with clinical examined features: A quantitative clinical study. *Health Sci Rep*. 2023 Oct;6(10):e1656. doi: 10.1002/hsr2.1656. PMID: 37900094. Exclusion Code: X4.

441. Um MJ, Cho EA, Jung H. Combination therapy of raloxifene and alendronate for treatment of osteoporosis in elderly women. *J Menopausal Med*. 2017 Apr;23(1):56-62. doi: 10.6118/jmm.2017.23.1.56. PMID: 28523260. Exclusion Code: X5.

442. Urano T, Shiraki M, Kuroda T, et al. Bisphosphonates prevent age-related weight loss in Japanese postmenopausal women. *J Bone Miner Metab*. 2018 Nov;36(6):734-40. doi: 10.1007/s00774-017-0891-y. PMID: 29236162. Exclusion Code: X7.

443. Valero C, Olmos JM, Llorca J, et al. Osteoporotic patients treated with bisphosphonates do not show the increased mortality observed in those untreated. *J Bone Miner Metab*. 2021 Apr 13;39(5):876-82. doi: 10.1007/s00774-021-01228-z. PMID: 33847832. Exclusion Code: X3.

444. van Geel TA, Eisman JA, Geusens PP, et al. The utility of absolute risk prediction using FRAX(R) and Garvan fracture risk calculator in daily practice. *Maturitas*. 2014 Feb;77(2):174-9. doi: 10.1016/j.maturitas.2013.10.021. PMID: 24287178. Exclusion Code: X9.

445. van Staa TP, Geusens P, Kanis JA, et al. A simple clinical score for estimating the long-term risk of fracture in post-menopausal women. *QJM*. 2006 Oct;99(10):673-82. doi: 10.1093/qjmed/hcl094. PMID: 16998210. Exclusion Code: X7.

446. Vestergaard P. Occurrence of gastrointestinal cancer in users of bisphosphonates and other antiresorptive drugs against osteoporosis. *Calcif Tissue Int*. 2011 Dec;89(6):434-41. doi: 10.1007/s00223-011-9539-4. PMID: 22002678. Exclusion Code: X10.

447. Vestergaard P, Fischer L, Mele M, et al. Use of bisphosphonates and risk of breast cancer. *Calcif Tissue Int*. 2011 Apr;88(4):255-62. doi: 10.1007/s00223-011-9463-7. PMID: 21253712. Exclusion Code: X7.

448. Villa P, Lassandro AP, Moruzzi MC, et al. A non-invasive prevention program model for the assessment of osteoporosis in the early postmenopausal period: a pilot study on FRAX(®) and QUS tools advantages. *J Endocrinol Invest*. 2016 Feb;39(2):191-8. doi: 10.1007/s40618-015-0341-4. PMID: 26141076. Exclusion Code: X6.

449. Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2018 Jun 26;319(24):2532-51. doi: 10.1001/jama.2018.6537. PMID: 29946734. Exclusion Code: X9.

450. Waltman N, Kupzyk KA, Flores LE, et al. Bone-loading exercises versus risedronate for the prevention of osteoporosis in postmenopausal women with low bone mass: a randomized controlled trial. *Osteoporos Int*. 2022 Feb;33(2):475-86. doi: 10.1007/s00198-021-06083-2. PMID: 34519832. Exclusion Code: X10.

451. Wang CC, Lu HT, Dusetzina SB, Wu CH. The association between long-term bisphosphonate use and the risk of fracture among women aged 50 or older with osteoporosis. *J Womens Health (Larchmt)*. 2016 Jul;25(7):738-46. doi: 10.1089/jwh.2015.5617. PMID: 27096405. Exclusion Code: X3.

452. Wang G, Sui L, Gai P, et al. The efficacy and safety of vertebral fracture prevention therapies in post-menopausal osteoporosis treatment: which therapies work best? a network meta-analysis. *Bone Joint Res*. 2017 Jul;6(7):452-63. doi: 10.1302/2046-3758.67.Bjr-2016-0292.R1. PMID: 28765269. Exclusion Code: X3.

453. Wang JC, Chien WC, Chung CH, et al. Adverse cardiovascular effects of nitrogen-containing bisphosphonates in patients with osteoporosis: a nationwide population-based retrospective study. *Int J Cardiol*. 2016 Jul 15;215:232-7. doi: 10.1016/j.ijcard.2016.04.088. PMID: 27128537. Exclusion Code: X10.

454. Wang L, Zhang D, Xu J. Association between the Geriatric Nutritional Risk Index, bone mineral density and osteoporosis in type 2 diabetes patients. *J Diabetes Investig*. 2020 Jul;11(4):956-63. doi: 10.1111/jdi.13196. PMID: 31828962. Exclusion Code: X8.

455. Wang WY, Chen LH, Ma WJ, You RX. Drug efficacy and safety of denosumab, teriparatide, zoledronic acid, and ibandronic acid for the treatment of postmenopausal osteoporosis: a network meta-analysis of randomized controlled trials. *Eur Rev Med Pharmacol Sci*. 2023 Sep;27(17):8253-68. doi: 10.26355/eurrev_202309_33586. PMID: 37750653. Exclusion Code: X3.

456. Wang Y, Mehrabi S, Sohn S, et al. Natural language processing of radiology reports for identification of skeletal site-specific fractures. *BMC Med Inform Decis Mak*. 2019 Apr 4;19(Suppl 3):73. doi: 10.1186/s12911-019-0780-5. PMID: 30943952. Exclusion Code: X5.

457. Watts NB, Brown JP, Papapoulos S, et al. Safety observations with 3 years of denosumab exposure: comparison between subjects who received denosumab during the randomized FREEDOM trial and subjects who crossed over to denosumab during the FREEDOM extension. *J Bone Miner Res*. 2017 Jul;32(7):1481-5. doi: 10.1002/jbmr.3119. PMID: 28277603. Exclusion Code: X6.

458. Watts NB, Grbic JT, Binkley N, et al. Invasive oral procedures and events in postmenopausal women with osteoporosis treated with denosumab for up to 10 years. *J Clin Endocrinol Metab*. 2019 Jun 1;104(6):2443-52. doi: 10.1210/jc.2018-01965. PMID: 30759221. Exclusion Code: X6.

459. Watts NB, Hattersley G, Fitzpatrick LA, et al. Abaloparatide effect on forearm bone mineral density and wrist fracture risk in postmenopausal women with osteoporosis. *Osteoporos Int*. 2019 Jun;30(6):1187-94. doi: 10.1007/s00198-019-04890-2. PMID: 30899994. Exclusion Code: X5.

460. Wei FL, Gao QY, Zhu KL, et al. Efficacy and safety of pharmacologic therapies for prevention of osteoporotic vertebral fractures in postmenopausal women. *Heliyon*. 2023 Feb;9(2):e11880. doi: 10.1016/j.heliyon.2022.e11880. PMID: 36852077. Exclusion Code: X3.

461. Wells GA, Hsieh SC, Zheng C, et al. Risedronate for the primary and secondary prevention of osteoporotic fractures in postmenopausal women. *Cochrane Database Syst Rev*. 2022 May 3;5(5):Cd004523. doi: 10.1002/14651858.CD004523.pub4. PMID: 35502787. Exclusion Code: X3.

462. Wen Z, Ding N, Chen R, et al. Comparison of methods to improve fracture risk assessment in Chinese diabetic postmenopausal women: a case-control study. *Endocrine*. 2021 Jul;73(1):209-16. doi: 10.1007/s12020-021-02724-y. PMID: 33932202. Exclusion Code: X8.

463. Whittier DE, Samelson EJ, Hannan MT, et al. A fracture risk assessment tool for high resolution peripheral quantitative computed tomography. *J Bone Miner Res*. 2023 Sep;38(9):1234-44. doi: 10.1002/jbmr.4808. PMID: 37132542. Exclusion Code: X3.

464. Wingood M, Criss MG, Irwin KE, et al. Screening for osteoporosis risk among community-dwelling older adults: a scoping review. *J Geriatr Phys Ther*. 2023 Oct-Dec 01;46(4):E137-e47. doi: 10.1519/jpt.0000000000000381. PMID: 36827688. Exclusion Code: X3.

465. Wolverton D, Elliott DP. Evaluating the evidence behind treating osteoporosis in the oldest adults. *Consult Pharm*. 2018 Jun 1;33(6):308-16. doi: 10.4140/TCP.n.2018.308. PMID: 29880092. Exclusion Code: X3.

466. Wu Q, Xiao X, Xu Y. Evaluating the performance of the WHO international reference standard for osteoporosis diagnosis in postmenopausal women of varied polygenic score and race. *Journal of Clinical Medicine*. 2020 Feb 12;9(2)doi: 10.3390/jcm9020499. PMID: 32059423. Exclusion Code: X9.

467. Wu ST, Chen JF, Tsai CJ. The impact of bisphosphonates on mortality and cardiovascular risk among osteoporosis patients after cardiovascular disease. *J Formos Med Assoc*. 2021 Nov;120(11):1957-66. doi: 10.1016/j.jfma.2020.12.014. PMID: 33358773. Exclusion Code: X8.

468. Wu Y, Chao J, Bao M, Zhang N. Predictive value of machine learning on fracture risk in osteoporosis: a systematic review and meta-analysis. *BMJ Open*. 2023 Dec 9;13(12):e071430. doi: 10.1136/bmjopen-2022-071430. PMID: 38070927. Exclusion Code: X5.

469. Xiao X, Wu Q. Validation of a genome-wide polygenic score in improving fracture risk assessment beyond the FRAX tool in the Women's Health Initiative study. *PLoS One*. 2023;18(6):e0286689. doi: 10.1371/journal.pone.0286689. PMID: 37262069. Exclusion Code: X9.

470. Xie D, Zhou Y, Zhang Y, et al. Osteoporosis screening based on body mass index, years since menopause and age among postmenopausal women in south central China. *Int J Clin Exp Med*. 2018;11(3):2543-50. Exclusion Code: X3.

471. Xie Z, Chen Y, Gurbuz S, et al. Effects of teriparatide in Chinese and Caucasian women with osteoporosis: bridging study on efficacy. *Clin Interv Aging*. 2019;14:959-68. doi: 10.2147/cia.S181929. PMID: 31213783. Exclusion Code: X3.

472. Xu Z. Alendronate for the treatment of osteoporosis in men: a meta-analysis of randomized controlled trials. *Am J Ther*. 2017 Mar/Apr;24(2):e130-e8. doi: 10.1097/mjt.0000000000000446. PMID: 27058577. Exclusion Code: X3.

473. Yamamoto T, Tsujimoto M, Sowa H. Safety of daily teriparatide treatment: a post hoc analysis of a phase III study to investigate the possible association of teriparatide treatment with calcium homeostasis in patients with serum procollagen type I N-terminal propeptide elevation. *Clin Interv Aging*. 2015;10:1101-9. doi: 10.2147/cia.S83549. PMID: 26185429. Exclusion Code: X2.

474. Yang HY, Huang JH, Chiu HW, et al. Vitamin D and bisphosphonates therapies for osteoporosis are associated with different risks of atrial fibrillation in women: a nationwide population-based analysis. *Medicine (Baltimore)*. 2018 Oct;97(43):e12947. doi: 10.1097/md.0000000000012947. PMID: 30412111. Exclusion Code: X10.

475. Yang L, Kang N, Yang JC, et al. Drug efficacies on bone mineral density and fracture rate for the treatment of postmenopausal osteoporosis: a network meta-analysis. *Eur Rev Med Pharmacol Sci*. 2019 Mar;23(6):2640-68. doi: 10.26355/eurrev_201903_17414. PMID: 30964193. Exclusion Code: X3.

476. Yang Q, Cheng H, Qin J, et al. A Machine Learning-Based Preclinical Osteoporosis Screening Tool (POST): Model Development and Validation Study. *JMIR Aging*. 2023 Nov 8;6:e46791. doi: 10.2196/46791. PMID: 37986117. Exclusion Code: X5.

477. Yang S, Leslie WD, Morin SN, Lix LM. Administrative healthcare data applied to fracture risk assessment. *Osteoporos Int*. 2019 Mar;30(3):565-71. doi: 10.1007/s00198-018-4780-6. PMID: 30554259. Exclusion Code: X3.

478. Yang S, Leslie WD, Morin SN, et al. Antiresorptive therapy and newly diagnosed diabetes in women: a historical cohort study. *Diabetes Obes Metab*. 2016 Sep;18(9):875-81. doi: 10.1111/dom.12678. PMID: 27097832. Exclusion Code: X7.

479. Yang T, Feng C, Qu Y, et al. Effect of teriparatide on quality of life in patients with postmenopausal osteoporosis: a retrospective cohort study. *J Int Med Res*. 2020 Feb;48(2):300060519876744. doi: 10.1177/0300060519876744. PMID: 31601158. Exclusion Code: X7.

480. Yang XC, Deng ZH, Wen T, et al. Network meta-analysis of pharmacological agents for osteoporosis treatment and fracture prevention. *Cell Physiol Biochem*. 2016;40(3-4):781-95. doi: 10.1159/000453138. PMID: 27915335. Exclusion Code: X3.

481. Yang Y, Luo X, Xie X, et al. Influences of teriparatide administration on marrow fat content in postmenopausal osteopenic women using MR spectroscopy. *Climacteric*. 2016 Jun;19(3):285-91. doi: 10.3109/13697137.2015.1126576. PMID: 26744910. Exclusion Code: X5.

482. Yang Y, Luo X, Yan F, et al. Effect of zoledronic acid on vertebral marrow adiposity in postmenopausal osteoporosis assessed by MR spectroscopy. *Skeletal Radiol*. 2015 Oct;44(10):1499-505. doi: 10.1007/s00256-015-2200-y. PMID: 26130070. Exclusion Code: X8.

483. Yavropoulou MP, Kasdagli MI, Makras P, et al. Nocebo-associated treatment discontinuation with subcutaneous anti-osteoporotic drugs. A systematic review and meta-analysis of placebo-arm dropouts in randomized-controlled trials. *Maturitas*. 2024 Jan;179:107874. doi: 10.1016/j.maturitas.2023.107874. PMID: 37976923. Exclusion Code: X3.

484. Yin MT, Shiau S, Rimland D, et al. Fracture prediction with modified-FRAX in older HIV-infected and uninfected men. *J Acquir Immune Defic Syndr*. 2016 Aug 15;72(5):513-20. doi: 10.1097/qai.0000000000000998. PMID: 27003493. Exclusion Code: X5.

485. Yıldırım Güzelant A, Sarıfakıoğlu AB, Yılmaz İ, et al. Should the cardiac rhythm be monitored following administration of zoledronic acid in elderly osteoporotic women? *Turkiye Fiziksel Tip ve Rehabilitasyon Dergisi*. 2016;62(3):214-21. doi: 10.5606/tftrd.2016.78871. Exclusion Code: X6.

486. Yu J, Goldshtein I, Shalev V, et al. Association of gastrointestinal events and osteoporosis treatment initiation in newly diagnosed osteoporotic Israeli women. *Int J Clin Pract*. 2015 Sep;69(9):1007-14. doi: 10.1111/ijcp.12676. PMID: 26278464. Exclusion Code: X6.

487. Yu R, Leung J, Woo J. Sarcopenia combined with FRAX probabilities improves fracture risk prediction in older Chinese men. *J Am Med Dir Assoc*. 2014 Dec;15(12):918-23. doi: 10.1016/j.jamda.2014.07.011. PMID: 25262197. Exclusion Code: X4.

488. Yu SF, Chen JF, Chen YC, et al. Beyond bone mineral density, FRAX-based tailor-made intervention thresholds for therapeutic decision in subjects on glucocorticoid. *Medicine (united states)*. 2017 Feb;96(5) (no pagination):e5959. doi: 10.1097/MD.0000000000005959. PMID: CN-01340713. Exclusion Code: X3.

489. Yu SF, Chen JF, Chen YC, et al. Beyond bone mineral density, FRAX-based tailor-made intervention thresholds for therapeutic decision in subjects on glucocorticoid: a nationwide osteoporosis survey. *Medicine (Baltimore)*. 2017 Feb;96(5):e5959. doi: 10.1097/md.0000000000005959. PMID: 28151883. Exclusion Code: X9.

490. Yu X, Ye C, Xiang L. Application of artificial neural network in the diagnostic system of osteoporosis. *Neurocomputing*. 2016;214:376-81. doi: 10.1016/j.neucom.2016.06.023. Exclusion Code: X3.

491. Yuh DY, Chang TH, Huang RY, et al. The national-scale cohort study on bisphosphonate-related osteonecrosis of the jaw in Taiwan. *J Dent*. 2014 Oct;42(10):1343-52. doi: 10.1016/j.jdent.2014.05.001. PMID: 24907558. Exclusion Code: X10.

492. Yun H, Delzell E, Ensrud KE, et al. Predicting hip and major osteoporotic fractures using administrative data. *Arch Intern Med*. 2010 Nov 22;170(21):1940-2. doi: 10.1001/archinternmed.2010.410. PMID: 21098356. Exclusion Code: X9.

493. Zagórski P, Tabor E, Martela-Tomaszek K, et al. Five-year fracture risk assessment in postmenopausal women, using both the POL-RISK calculator and the Garvan nomogram: the Silesia osteo active study. *Arch Osteoporos*. 2021 Feb 16;16(1):32. doi: 10.1007/s11657-021-00881-1. PMID: 33594643. Exclusion Code: X6.

494. Zanchetta MB, Boailchuk J, Massari F, et al. Significant bone loss after stopping long-term denosumab treatment: a post FREEDOM study. *Osteoporos Int*. 2018 Jan;29(1):41-7. doi: 10.1007/s00198-017-4242-6. PMID: 28975362. Exclusion Code: X6.

Appendix C. Excluded Studies

495. Zavras AI, Shanmugham JR. Bisphosphonates, osteoporosis, and osteonecrosis of the jaw: a critical review of a large nested case-control study. *J Evid Based Dent Pract*. 2016 Jun;16(2):136-8. doi: 10.1016/j.jebdp.2016.05.007. PMID: 27449847. Exclusion Code: X3.

496. Zebaze R, Libanati C, McClung MR, et al. Denosumab reduces cortical porosity of the proximal femoral shaft in postmenopausal women with osteoporosis. *J Bone Miner Res*. 2016 Oct;31(10):1827-34. doi: 10.1002/jbmr.2855. PMID: 27082709. Exclusion Code: X7.

497. Zechmann S, Scherz N, Reich O, et al. Appropriateness of bone density measurement in Switzerland: a cross-sectional study. *BMC Public Health*. 2018 Apr 2;18(1):423. doi: 10.1186/s12889-018-5305-0. PMID: 29606111. Exclusion Code: X7.

498. Zeng LF, Pan BQ, Liang GH, et al. Does routine anti-osteoporosis medication lower the risk of fractures in male subjects? An updated systematic review with meta-analysis of clinical trials. *Front Pharmacol*. 2019;10:882. doi: 10.3389/fphar.2019.00882. PMID: 31447677. Exclusion Code: X3.

499. Zenke Y, Ikeda S, Fukuda F, et al. Study of atypical femoral fracture cases coupled in a multicenter study. *J uoeh*. 2016 Sep;38(3):207-14. doi: 10.7888/juoeh.38.207. PMID: 27627968. Exclusion Code: X6.

500. Zhang J, Zhou R, Luo X, et al. Routine chest CT combined with the osteoporosis self-assessment tool for Asians (OSTA): a screening tool for patients with osteoporosis. *Skeletal Radiol*. 2023 Jun;52(6):1169-78. doi: 10.1007/s00256-022-04255-7. PMID: 36520217. Exclusion Code: X8.

501. Zhang J, Zhou S, Cai F, Ma Y. The curative effect of the additional administration of ibandronate sodium and calcitriol on elderly patients with osteoporosis. *Int J Clin Exp Med*. 2020;13(4):2371-8. PMID: CN-02164219. Exclusion Code: X6.

502. Zhang K, Wang M, Han W, et al. Construction of a predictive model for osteoporosis risk in men: using the IOF 1-min osteoporosis test. *J Orthop Surg Res*. 2023 Oct 11;18(1):770. doi: 10.1186/s13018-023-04266-7. PMID: 37821993. Exclusion Code: X5.

503. Zhang X, Hamadeh IS, Song S, et al. Osteonecrosis of the jaw in the United States Food and Drug Administration's Adverse Event Reporting System (FAERS). *J Bone Miner Res*. 2016 Feb;31(2):336-40. doi: 10.1002/jbmr.2693. PMID: 26288087. Exclusion Code: X3.

504. Zhang X, Lin J, Yang Y, et al. Comparison of three tools for predicting primary osteoporosis in an elderly male population in Beijing: a cross-sectional study. *Clin Interv Aging*. 2018;13:201-9. doi: 10.2147/cia.S145741. PMID: 29440880. Exclusion Code: X8.

505. Zhang ZL, Liao EY, Xia WB, et al. Alendronate sodium/vitamin D3 combination tablet versus calcitriol for osteoporosis in Chinese postmenopausal women: a 6-month, randomized, open-label, active-comparator-controlled study with a 6-month extension. *Osteoporos Int*. 2015 Sep;26(9):2365-74. doi: 10.1007/s00198-015-3141-y. PMID: 25929192. Exclusion Code: X6.

506. Zhao S, Zhao W, Du D, et al. Effect of bisphosphonate on hip fracture in patients with osteoporosis or osteopenia according to age: a meta-analysis and systematic review. *J Investig Med*. 2022 Mar;70(3):837-43. doi: 10.1136/jim-2021-001961. PMID: 34893517. Exclusion Code: X3.

507. Zheng JQ, Lai HJ, Zheng CM, et al. Association of stroke subtypes with risk of hip fracture: a population-based study in Taiwan. *Arch Osteoporos*. 2017 Nov 22;12(1):104. doi: 10.1007/s11657-017-0390-8. PMID: 29167998. Exclusion Code: X5.

508. Zhou J, Ma X, Wang T, Zhai S. Comparative efficacy of bisphosphonates in short-term fracture prevention for primary osteoporosis: a systematic review with network meta-analyses. *Osteoporos Int*. 2016 Nov;27(11):3289-300. doi: 10.1007/s00198-016-3654-z. PMID: 27273112. Exclusion Code: X3.

509. Zhou J, Wang T, Zhao X, et al. Comparative efficacy of bisphosphonates to prevent fracture in men with osteoporosis: a systematic review with network meta-analyses. *Rheumatol Ther*. 2016 Jun;3(1):117-28. doi: 10.1007/s40744-016-0030-6. PMID: 27747517. Exclusion Code: X3.

510. Zysset P, Pahr D, Engelke K, et al. Comparison of proximal femur and vertebral body strength improvements in the FREEDOM trial using an alternative finite element methodology. *Bone*. 2015 Dec;81:122-30. doi: 10.1016/j.bone.2015.06.025. PMID: 26141837. Exclusion Code: X7.

Appendix D Table 1. Characteristics of Included Studies for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Participant Characteristics	Intervention Groups
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS NTR2430 RCT Fair	**N=**11,032 Women ages 65 to 90 years recruited from general practice registries in the Netherlands Key Inclusion Criteria: ≥1 clinical risk factor for fracture based on questionnaire (previous fracture after age 50, parental hip fracture, BMI <19, rheumatoid arthritis, menopause <45 years, malabsorption syndrome, chronic liver disease, type 1 diabetes, immobility) Key Exclusion Criteria: Age ≥91 years; short life expectancy according to their general practitioner, terminal illness; current use of antiosteoporosis medication or in preceding 5 years, recent densitometry; body weight >135 kg; corticosteroid use >7.5 mg/prednisone equivalent/day Mean (SD) age: 75.0 (6.7) % Female: 100 Mean T-score (site of BMD): NR % Prior Fx: 43% of usual-care group and 44% of screening group had fracture after age 50 years per questionnaire, but not reported if fragility fracture	**Screening:** All patients received onetime FRAX without BMD assessment, DXA, VFA, fall risk assessment, and blood chemistry screening (serum vitamin D, calcium, creatinine, albumin, thyroxine, thyroid stimulating hormone, erythrocyte sedimentation rate) to exclude secondary osteoporosis. U.K. FRAX tool was used but with age-dependent cutoffs derived from data on a representative sample of older Dutch persons. Based on those tests, women with treatment indications were referred to PCP for personalized treatment advice to include anti-osteoporotic therapy, additional evaluation for secondary cause of osteoporosis, fall prevention, and calcium/vitamin D supplementation. Indications for treatment included FRAX with BMD score above age-dependent threshold, T-score <-2, or prevalent vertebral fracture. Age-dependent thresholds reported in Table 6 of Elders et al.[125] Participating general practitioners attended a group education on general aspects of osteoporosis and treatment and received instruction on the study protocol and treatment program. Practitioners could contact the study team for advice as needed. First choice treatment was alendronate 70 mg/week or risedronate 35 mg/week. Deviation from treatment protocol allowed based on professional judgment. **No routine screening:** Participants offered the same screening program after study completion (i.e., put on a wait-list). No routine screening offered; participants had usual care from their PCPs. Participants with an indication for DXA based on national guidelines at the time of the study were notified and advised to contact their PCP as part of usual care. Existing national guidelines suggest DXA or VFA testing based on assessment of clinical risks including history of vertebral fracture or recent fracture (within 2 years) after age 50; age older than 60 years, nonrecent fractures after age 50, parental hip fracture, body weight <60 kg, and severe immobility or 1 fall or more in the past year. **Fidelity/adherence to screening intervention:** 1,347/5,575 randomized (24%) to screening did not receive screening. 1,417/5,575 randomized (25%) to screening had an indication for treatment.

Appendix D Table 1. Characteristics of Included Studies for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Participant Characteristics	Intervention Groups
Merlijn et al, 2019[124] Elders et al, 2017[125] (continued)	Other: Fewer than 1% were using corticosteroids; mean (SD) 10-year FRAX w/o BMD MOF risk 24.3 (10.5) in usual-care group and 24.6 (10.8) in screening group; mean (SD) 10 yr FRAX w/o BMD Hip Fx risk 11.3 (10.2) in usual-care group and 11.6 (10.5) in screening group. Treatment indications: morphometric vertebral fractures on instant vertebral assessment, fracture risk according to FRAX >age-specific threshold, T-score <-2	1,154/5,575 (21%) randomized to screening received treatment over the course of the study. 18% (982/5,575 randomized) reported starting treatment and 11.8% (657/5,575 randomized) reported still being on treatment at 36 months; of those without an indication, 1% (68/5,575 randomized) reported treatment at 36 months. The discussion states that 31% of those with an indication did not start medication. 52/5,457 randomized (1%) to control were lost to followup and not included. 291/5,457 randomized (5%) to control received treatment over the course of the study; 3% (167/5,457) by 18 months.
Rubin et al, 2018[126] Rubin et al, 2015 [127] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE NCT01388244 RCT Fair	**N=**34,229 Women ages 65 to 80 years living in southern Denmark Key Inclusion Criteria: Age 65 to 80 years Key Exclusion Criteria: NR **Mean (SD) age:** Median 71 [IQR 68, 76] % Female: 100 Mean T-score (site of BMD): In the population that returned the initial questionnaire with no missing data in the screening group and who were not already receiving treatment and who had baseline FRAX score ≥15% and who accepted offer of DXA scan (5,009, which was 71% of those invited to DXA). -1.2 (1.0); total hip -1.3 (1.4); lumbar spine A total of 446 (8.8%) and 926 (18.3%) of the scanned women had T-score below -2.5 at total hip or lumbar spine, respectively.	**Screening**: The intervention included two steps: (1) fracture risk assessment via FRAX and (2) invitation to DXA for areal BMD and VFA if 10-year FRAX MOF risk was ≥15%. Results of the DXA were sent to the participant and her general practitioner, which included treatment recommendations based on national guidelines. Final decision about treatment was at the discretion of the patient and provider. **No routine screening:** No contact after completion of baseline data collection; usual care guided by PCP. **Fidelity/adherence to screening intervention:** 7,793/17,072 randomized (45.6%) to screening did not receive screening with FRAX calculation (1,132 already on treatment, 2,894 returned questionnaire blank, 104 returned questionnaire with data missing to calculate FRAX, and the rest did not return the questionnaire). 2,047/17,072 randomized (12%) were high risk but did not have a DXA (830 were not interested in a DXA and 1,217 dropped out); 5,009/17,072 randomized (29%) were high risk and had a DXA. This represents 71% of persons deemed high risk based on FRAX (5,009/7,056). [The authors reported that 48% of those screened had a DXA, which comes from the 10,411 with calculated FRAX scores and not the overall randomized intervention group of 17,072.]. 1,236/17,072 randomized (7%) had a DXA result with an indication for treatment. Eligibility for DXA required a completed questionnaire and high-risk FRAX score (≥15%).

Appendix D Table 1. Characteristics of Included Studies for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Participant Characteristics	Intervention Groups
Rubin et al, 2018[126] Rubin et al, 2015 [127] Rothman et al, 2017[128] Hoiberg et al, 2019[129] (continued)	% Prior Fx: Not available for the ITT population. In the population that was invited to participate and who returned the questionnaire with no missing data (61.1% of those invited): Total: 2,570 (12.3%) Screening: 1,316 (12.6%) Control: 1,254 (12.0%) Participants in the screening group also received VFA; 187 (3.7%) of those scanned had prevalent moderate to severe vertebral fractures. Other: 61.1% of those invited to participate returned the questionnaire with no missing data; 1,994 (9.5%) indicated they were already being treated for osteoporosis, and 20.9% had conditions related to secondary osteoporosis. The incidence of these two were similar between the screening and control groups. Median 10-year FRAX MOF 20 (in both screened and control); median 10-year FRAX hip: 6.7 (screened); 6.6 (control).	986/17,072 randomized (6%) received treatment; this number (986) appears to be based on only those who received DXA through the study and had an indication for treatment based on the study DXA who were then referred back to their GPs for further evaluation and management as part of the study. This is 80% of those eligible for treatment [986/1,236]. The authors stated that 23% of the screening group received medication after the index date (mailing of questionnaire), which we assume includes the 1,132 women who indicated they were already receiving medication on the baseline questionnaire along with women who were randomized to screening but who did not return the questionnaire but who may have been prescribed medication by their GPs through the course of usual care outside of this study. 7831/17,157 randomized (45.6%) did not participate (1,168 were already on treatment, 3,143 returned a blank questionnaire, 111 returned a questionnaire with missing data to calculate FRAX, and the rest did not return the questionnaire). In the control group, 7,026/17,157 randomized (41%) had FRAX ≥15%. The number of participants in the control group who received a DXA was not reported, but the authors report that 25% of women in the control group had a DXA vs. 48% in the screening group. Based on the information in the article, the denominator is likely "Calculated FRAX total" and this gives us a N/10,494=25% such that likely N=2,623.5 or 15% of total control group. The authors note that 18% of the control group received medication after the index date (mailing of the questionnaire); it is unclear whether these were women with FRAX ≥15% and ≤15% or whether they received DXA prior to treatment, and whether this includes the 1,168 women who were excluded from FRAX calculation because they indicated they were taking treatment on the baseline questionnaire.

Appendix D Table 1. Characteristics of Included Studies for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year **Trial Name** **Registry Number** **Study Design** **Study Quality**	**Participant Characteristics**	**Intervention Groups**
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP ISRCTN 55814835 RCT Fair	**N=**12,483 Women ages 70 to 85 years without known osteoporosis and who were recruited through general practitioner offices in the U.K.; 99% White Key Inclusion Criteria: Women ages 70 to 85 years **Key Exclusion Criteria:** Known to be on prescription treatment for osteoporosis (other than calcium and vitamin D), any known comorbidity that would in the general practitioner's opinion make entry to the trial inadvisable (e.g., advanced malignancy), other factors that would make invitation to participate in a research study inappropriate (e.g., recent bereavement). Mean (SD) age: Screening: 75.5 (4.16) Control: 75.5 (4.14) % Female: 100 Mean T-score (site of BMD): Screening (high risk segment): -2.6 (femoral neck) Control: Not measured % Prior Fx: Broken bone since age 50: Screening: 22% Control: 23%	**Screening:** Onetime FRAX assessment with high-risk group invited for femoral neck DXA. High-risk designation was based on comparison of participants 10-year hip fracture risk to an age-based threshold (70–74 years, 5.18%; 75–79 years, 6.81%; 80–84 years, 8.46%; 85 years, 8.39%) derived based on U.K. cost-effectiveness data. Participants deemed low risk were notified of low-risk status by letter to participant and their PCP and no further intervention offered. High-risk persons completing DXA scan had updated FRAX score with BMD information communicated to them and their PCP. Participants with age-specific risks above treatment thresholds were advised to discuss treatment options with their PCP; thresholds as follows: 70–74 years, 5.24%, 75–79 years, 6.87%, 80–84 years, 8.52%, 85 years, 8.99%. **No routine screening:** Letter sent to participant's PCP informing them of their patient's participation in the study, no routine screening offered, usual care as determined by participant's PCP. **Fidelity/adherence to screening intervention:** 6/6,233 randomized (<0.1%) to screening were not screened. 247/6,233 (4%) randomized to screening were high risk but did not have a DXA (157 declined, 81 were unable to have hip BMD measured, and 9 died). 2,817/6,233 randomized (45%) to screening were high risk after FRAX screening and had a DXA. 898/6,233 randomized (14.4%) to screening continued to be high risk after revised FRAX score with BMD and had treatment recommended. 1,486/6,233 randomized (24%) received at least one prescription for treatment over the course of the study; 953/6,233 randomized (15%) received treatment in the first 12 months; of those considered high risk, 703/898 (78%) received treatment in the first 6 months. Adherence among those taking medication at 6 months: 79.2% by 1 year, 65% by 2 years, 34.9% by 5 years. 6/6,250 randomized (<0.1%) to control did not participate. Number randomized to control that received DXA through usual care was NR.

Appendix D Table 1. Characteristics of Included Studies for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Participant Characteristics	Intervention Groups
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] (continued)	Other: Mean (SD) 10-year FRAX MOF risk: 19.3% (8.9%) Screened; 19.3% (8.8%) Control Mean (SD) 10-year FRAX Hip risk: 8.5% (7.4%) Screened; 8.5% (7.3%) Control	982/6,250 randomized (16%) to control received treatment over the course of the study; 264/6,250 randomized (4%) in the first 12 months. Participants with prescriptions for antiosteoporotic medication: End of first year: Screening group, 15%, usual care group, 4% End of fifth year: Overall, 11.5%; screening group, 13%–14%, usual care group, 9.7%

Abbreviations: BMD=bone mineral density; BMI=body mass index; DXA=dual-energy X-ray absorptiometry; FRAX=Fracture Risk Assessment Tool; FRAX MOF=Fracture Risk Assessment Tool: Major Osteoporotic Fracture; Fx=fracture; GP=general practitioner; IQR=interquartile ratio; ISRCTN=International Standard Randomised Controlled Trial Number; ITT=intention to treat; N=number; NCT=National Clinical Trial; NR=not reported; NTR=Netherlands Trial Registry; PCP=primary care provider; RCT=randomized, controlled trial; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study; SD=standard deviation; U.K.=United Kingdom; VFA=vertebral fracture assessment.

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Azagra et al, 2015[168, 310] Poor	FRIDEX; Spain	Women ages 40 to 90 years referred for DXA by their physician	816; N (%) Female: 816 (100)	56.8 (8.2)	Persons with cancer or who were receiving osteoporosis medications were excluded. *Mean T-score at baseline:* NR; 127 (15.6) were classified as osteoporosis based on DXA N (%) with prior fracture at baseline: 166 (20.3)
Azagra et al, 2016[174] Poor	FROCAT; Spain	Retrospective analysis of a cohort of women ages 40 to 90 years from primary care practices managed by a main public provider of health services	1,090; N (%) Female: 1,090 (100)	59.1 (12.4) ≥65 years: 375 (34%)	Persons who developed cancer, lived outside of the study area, died, or were unable to be contacted were all excluded; current or past users of osteoporosis medication were not excluded. *Mean T-score at baseline:* NR; of the 234 women with DXA, 36.3% had osteoporosis N (%) with prior fracture at baseline: Previous fragility 154 (14.1)
Baleanu et al, 2021[179] Poor	FRISBEE; Belgium	Population based cohort of postmenopausal women ages 60 to 85 years recruited from population registers to participate in a study designed to evaluate various risk prediction models	3,030; N (%) Female: 3,030 (100)	NR; 1,347 (44.5%) ≥70 years	Mean T-score at baseline: NR N (%) with prior fracture at baseline: 801 (26.4)
Bolland et al, 2011[166] Poor	None; New Zealand	Healthy menopausal women age ≥55 years who were taking part in a 5-year placebo-controlled trial of calcium supplements; race/ethnicity NR	1,422; N (%) Female: 1,422 (100)	74.2 (4.2)	Normal lumbar spine BMD for their age (Z-score >-2), not taking osteoporosis medication or vitamin D supplements in doses >1,000 IU/day, serum 25 [OH] D levels ≥25 nmol/L *Mean T-score at baseline:* T-score at FN: -1.3 (1.0) % with T-score <-2.5: 11 *N (%) with prior fracture at baseline:* Prior fracture during adult life: NR (33.5)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159] Moller et al, 2022[163] Leslie et al, 2022[164] Poor	Manitoba BMD Registry; Canada	All cohorts retrospectively assembled from the registry; race/ethnicity NR; each article used slightly different criteria for its analysis as follows: • Women age ≥50 who had initial DXA scan between 1996 and 2011[158] • Women age ≥40 who had initial DXA in 1996 or later and had at least 5 years of followup[160] • Persons age 50 years or older with first DXA between January 1990 and March 2007[157] • Persons age 50 years or older with first DXA after 1996 with at least 5 years of observation post-test[161] • Persons age 40 years or older with first DXA of hip and lumbar spine between 1996 and 2013[162] • Women ages 40 to 59 years who underwent DXA scan between 1998 and 2002[159] • Persons age 45 or older who underwent DXA between 1996 and March 2016[163, 164]	68,730 (largest N from the articles)[162] N (%) Female (largest N from the articles): 62,275 (90.6)	From the largest article[162] Women: 64.1 (11.1) Men: 66.0 (12.2)	From the largest analysis[162] One analysis excluded persons taking osteoporosis treatment[161] Mean T-score at baseline: T-score at FN Women: -1.4 (1.0) Men: -1.1 (1.1) using White female reference range N (%) with prior fracture at baseline: Prior fragility fracture Women: 8,833 (14.2) Men: 1,179 (18.3)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Chapurlat et al, 2020[175] Poor	OFELY and QUALYOR; France	Retrospective analysis of 2 population-based cohorts Postmenopausal women with a baseline bone measure obtained during 2006–2008 from OFELY, and women with T-scores at the hip of between -1.0 and -2.5 with clinical risk factors or <-3.0 without risk factors from QUALYOR	2,100; N (%) Female: 2,100 (100)	OFELY: 68 (NR) QUALYOR: 65.9 (NR)	*Mean T-score at baseline:* OFELY: -1.36, 6.7% with osteoporosis QUALYOR: -1.70, 7.8% with osteoporosis N (%) with prior fracture at baseline: NR
Cheung et al, 2012[150] Poor	Hong Kong Osteoporosis Study; Hong Kong	Community-dwelling, ambulatory, postmenopausal women age ≥40 years recruited from different districts of Hong Kong between 1995 and 2009 during health fairs and road shows on osteoporosis	2,266; N (%) Female: 2,266 (100)	62.1 (8.5)	Women taking osteoporosis treatment were excluded. *Mean T-score at baseline:* -1.5 (1.1); 30.1% with osteoporosis *N (%) with prior fracture at baseline:* Low-trauma fracture after age 45: 291 (12.8)
Collins et al, 2011[167] Poor	THIN Database; U.K.	Patients ages 30 to 85 years registered between 1994 and 2008 with records in the THIN database, a database of general practices that use INPS Vision system (20% of U.K. practices); race/ethnicity NR	2,209,451; N (%) Female: 1,136,417 (50.6)	Median (IQR) Women: 48 (37 to 62) Men: 47 (37 to 59)	No previously recorded fracture of hip, distal radius, or vertebra Mean T-score at baseline: NR N (%) with prior fracture at baseline: 0 (0)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Crandall et al, 2014[139] Crandall et al,2018[140] Crandall et al,2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Poor	Women's Health Initiative (WHI); U.S.	Retrospective cohort of postmenopausal women ages 50 to 79 years enrolled in either clinical trial or observational study components of the WHI who were free from serious cardiac, pulmonary, renal, and hepatic conditions with at least 3 years' life expectancy; race/ethnicity: 86.0% White; 7.4% Black; 3.0% Hispanic; 3.7% other/unknown[141]	161,808 overall sample size; after additonal exclusion criteria applied (N=117,707)[141] and after persons with missing covariate data excluded (N=99,413)[142] Three analyses were limited to women ages 50 to 64 years at baseline N=62,492[139] N=63,723[140] N=67,169[143] *N (%) Female:* Varies by analysis (100)	Mean age: 62.7 (7.1) from the largest analysis[141] 50–54: 16,699 (14.2) 55–59: 24,898 (21.2) 60–64: 28,090 (23.9) 65–69: 25,534 (21.7) 70–74: 16,289 (13.8) 75–79: 6,197 (5.3) Mean age 57.9 (4.1) in analysis limited to age 50–64[139]	Postmenopausal, free from serious medical conditions; participants using osteoporosis medication or somatostatin agents at baseline were excluded as were participants with fewer than 10 years of followup time and who contributed incomplete information regarding risk factors *Mean T-score at baseline:* NR for overall analystic sample; but a subset of participants did have BMD at baseline and 1,642/8,134 (20%) had T-score <-2.5[141] *N (%) with prior fracture at baseline:* Self-reported fracture after age 55 years: 10,090 (8.6)[141]
Dagan et al, 2017[169] Poor	None; Israel	Electronic health record data for members ages 30 to 100 years (depending on tool validation) from one of four national healthcare insurer/providers; race/ethnicity NR	1,054,815; N (%) Female: NR (54.6)	50–59: 38.0% 60–69: 28.4% 70–79: 21.1% 80–89: 12.5%	Continuous membership in the health plan for 3 years prior to index date and during followup period Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* Prior fracture after age 50 years: 119,329 (11.3)
Davis et al, 2019[173] Poor	Fremantle Diabetes Study Phase 1; Australia	Retrospective analysis of a longitudinal cohort of persons with known diabetes from an urban community in one region of the country; only cohort members between ages 40 and 89 years with type 2 diabetes were included in this analysis.	1,251; N (%) Female: 641 (51)	65.0 (10.0)	Mean T-score at baseline: NR N (%) with prior fracture at baseline: 19 (1.5)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Desbiens et al, 2020[177] Poor	CARTaGENE; Canada	Retrospective analysis of data from a population-based survey of adults ages 40 to 69 years in a single province (Quebec); persons with history of dialysis or kidney transplant were excluded. Only the persons without chronic kidney disease from this cohort were included for this update review. 90% White.	9,522; N (%) Female: NR (51.9)	Median 51 (IQR 46 to 57)	Persons living in nursing homes, correctional facilities, and First Nation Reserves were excluded. Mean T-score at baseline: NR N (%) with prior fracture at baseline: NR (3.2%)
Ensrud et al, 2009[152] Premaor et al, 2013[153] Poor	Study of Osteoporotic Fractures (SOF); U.S.	White women younger than age 65 years recruited between 1986 and 1988 from population-based listings in 4 U.S. areas.	6,252; N (%) Female: 6,252 (100)	71.3 (5.1)	Black women were excluded because of low incidence of hip fracture; women who were unable to walk without assistance or had a history of bilateral hip replacement were also excluded. Mean T-score at baseline: FN BMD (g/cm^2) Overall: 0.65 (0.11) Obese: 0.66 (0.10) Nonobese: 0.61 (0.10) N (%) with prior fracture at baseline: Overall 2,155 (35) Obese: NR (45.6) Nonobese: NR (45.3)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Ettinger et al, 2012[72] Ettinger et al, 2013[145] Gourlay et al, 2017[146] Poor	MrOs; U.S.	Community-dwelling men age 65 years or older recruited from 6 clinical centers between March 2000 and April 2002; 89.4% White; 4% Black; 3% Asian; 2% Hispanic; 1% other	5,893; N (%) Female: 5,893 (0) 5,200 reported in companion study[146]	73.6 (5.9) Reported in companion study[146] 65–69: 67.1 (1.4) 70–74: 71.9 (1.4) 75–79: 76.8 (1.4) ≥80: 83.0 (2.9)	Men who had used bisphosphonates within 30 days prior to baseline visit were excluded. *Mean T-score at baseline:* T-score at femoral neck: -1.12 (0.91) N (%) by category of T-score Normal: 2,459 (41.7) Osteopenic: 3,151 (53.5) Osteoporosis: 282 (4.8) *N (%) with prior fracture at baseline:* Fracture after age 50 years: 1,302 (22.1);[145] Fracture after age 45 years: 1,247 (21.1)[72] For the analysis in the companion study,[146] only included men without a prior history of hip or clinical vertebral fracture, who had no history of past or current FDA-approved antifracture treatment and did not have osteoporosis by BMD at baseline (for the analysis of fracture risk scores calculated with BMD) or men without a prior history of hip or clinical vertebral fracture and who had no history of past or current FDA-approved antifracture treatment (for the analysis of fracture risk scores calculated without BMD).
Fraser et al, 2011[154] Langsetmo et al, 2011[155] Poor	CaMos; Canada	Data from the CaMos cohort which included persons living within proximity to 1 of 9 Canadian cities randomly selected from residential phone numbers; only persons age 50 years or older were included in one of the analyses.[154] Participants ages 55 to 95 years were included in the other;[155] race/ethnicity NR	From Fraser et al,[154] 6,697; N (%) Female: 4,778 (71.3) From Langsetmo et al[155] 5,758 N (%) Female: 4,152 (72.1)	From Fraser et al,[154] Women: 65.8 (8.8) Men: 65.3 (9.1) From Langsetmo et al[155] Women: 67.7 (7.60) Men: 67.6 (7.6)	From Fraser et al,[154] *Mean T-score at baseline:* FN T-score/ Minimum T-score Women: -1.5 (1.1)/-1.8 (1.1) Men: -0.5 (1.2)/-0.8 (1.2) N (%) with prior fracture at baseline: Prior fragility fracture: Women: 540 (11.3) Men: 94 (4.9) From Langsetmo et al[155] Mean T-score at baseline: FN T-score Women: -1.43 (0.93) Men: -1.0 (1.0) N (%) with no history of fracture after age 50: Women: 3,628 (87.4) Men: 1,518 (94.5)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Garcia-Sempere et al, 2022[181] Poor	ESOVAL; Spain	Men and women age 50 years or older recruited from primary care centers in a single large healthcare system	9,082; N (%) Female: 3,679 (40.5)	64.2 (9.8)	Of the 11.2% of persons who had BMD testing within 2 years of recruitment, 1.8% had osteoporosis N (%) with prior fracture at baseline: 538 (5.9)
Goldshtein et al, 2018[172] Poor	Maccabi Healthcare Services (MHS); Israel	Retrospective cohort assembled from data from the computerized database of Maccabi Healthcare Services; a large government-funded health maintenance organization. This analysis included women ages 50 to 90 years in 2004 with at least 3 years of prior membership.	141,320; N (%) Female: 141,320 (100)	Median 58 (IQR 54 to 67)	Persons with osteoporosis treatment were included (19%) if they were on therapy before the index date. Mean T-score at baseline: NR N (%) with prior fracture at baseline: Prior MOF: 4%
Gonzalez-Macias et al, 2012[149] Poor	ECOSAP; Spain	Caucasian women age 65 years or older recruited from 58 primary care centers of the National Health Services in Spain between March 2000 and June 2001	5,146; N (%) Female: 5,146 (100)	72.3 (5.3)	Excluded women with metabolic bone disease, renal failure, hypercalcemia, therapeutic doses of fluoride for certain duration, life expectancy <3 years Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* Any fracture since age 35 years: NR (20.2)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148] Poor	QResearch Database; U.K.	Hippisley-Cox et al[147] Cohort of primary care patients obtained from the QResearch Database, a database of more than 13 million patients registered at more than 620 general practices using the Egton Medical Information System. For this review; we only considered the "validation" dataset from this cohort. For this analysis, patients ages 30 to 100 years registered with practices between January 1993 and October 2011 were included; 94% White. Hippisley-Cox et al[148] Retrospective cohort of patients ages 30 to 85 years assembled from electronic health record databases of over 11 million patients registered at 574 general practices using the Egton Medical Information System during January 1993 to June 2008; only the validation dataset was included for purposes of this update review; 94% White.	Hippisley-Cox et al[147] 1,583,373; *N (%) Female:* 804,563 (50.8) Hippisley-Cox et al[148] 1,275,917; *N (%) Female:* Women: 642,153 (50.3) Men: 633,764 (49.7)	Hippisley-Cox et al[147] 50 (16) Hippisley-Cox et al[148] Median (IQR) age Women: 49 (37 to 63 years) Men: 46 (37 to 69 years)	Hippisley-Cox et al[147] Mean T-score at baseline: NR N (%) with prior fracture at baseline: 27,907 (1.8%) Hippisley-Cox et al[148] Persons with prior fracture of hip, distal radius, or vertebra were excluded. Mean T-score at baseline: NR N (%) with prior fracture at baseline: 0 (0)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170] Poor	Clinical Practice Research Database; U.K.	Retrospective analysis of data on participants ages 30 to 99 years from the Clinical Practice Research Database, a database of patients from general practices in the U.K.; race/ethnicity NR. One analysis was limited to persons at 357 practices with links to the Office of National Statistics.[171] Other analysis was limited to persons ages 40 to 90 years between January 1987 and December 2013 from medical records of 625 primary care practices.[170]	2,852,381 for analysis of fractures[171] *N (%) Female:* 1,682,709 (51.4) (For the entire database of 3.3 million) 338,755 (24,227 for hip fracture analysis)[170] *N (%) Female:* NR for general population, but in the matched RA cohort the proportion that was female was 67.8%	By age band, % men/% women for the entire database (3.3 million) 25–34 years: 26.9%/27.8% 35–44 years: 25.0%/21.6% 45–54 years: 18.5%/16.5% 55–64 years: 13.4%/12.6% 65–74 years: 9.3%/9.8% 75+ years: 6.9%/11.8% From companion study[170] NR for general population, but in the matched RA cohort the mean (SD) age was 62.9 (11.4)	Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* For the entire database of 3.3 million Men: 24,265 (1.5) Women: 45,752 (2.7) From companion study:[170] Persons exposed to osteoporosis drugs before the index date were excluded; the reported analysis compared persons with RA to the general population; only data for the general population were captured in our review. Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* NR (only reported for the RA population, but the general population was not matched on this characteristic)
Jain et al, 2023[311] Poor	Temple University Hospital; U.S.	Retrospective analysis of EHR data on men and women at least 50 years of age with at least 1 primary care provider visit per year between 2010 and 2018 and at least 2 full years of followup. Excluded participants with missing data for risk calculation or with prescription for osteoporosis medication. Participants with unknown race or who were multiracial were excluded.	24,189; N (%) Female: 13,051 (54.0)	White: 64.5 (9.9) Black: 61.2 (9.3) Hispanic: 60.2 (8.7)	Mean T-score at baseline: NR N (%) with any prior fracture at baseline: 4% to 5% (varied by race) N (%) with secondary osteoporosis: 18.6% to 22.1% (varied by race)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Lo et al, 2011[73] Pressman et al, 2011[165] Poor	Kaiser Permanente Northern California; U.S.	Women ages 50 to 85 years who underwent first DXA scan between 1997 and 2003; 76% White, 14% Asian, 6% Hispanic, 4% Black	94,489; *N (%) Female:* 94,489 (100)	Category of age, N (%) 50–59: 39,138 (41.4) 60–69: 32,831 (34.8) 70–79: 19,098 (20.2) 80 or older: 3,422 (3.6)	Excluded women without coverage 1 year before and after the DXA scan, without accessible data, or missing race/ethnicity. Women with a filled prescription for bisphosphonates in the year prior to DXA were also excluded. Mean T-score at baseline: T-score at FN Above -1.0: NR (39.1) Between -1.0 and -2.5: NR (49.7) -2.5 or below: NR (11.2) N (%) with prior fracture at baseline: Fracture after age 45: NR (10.1)
Lu et al, 2021[178] Poor	5 cohorts (UK Biobank, MrOs US, MrOs Sweden, SOF, CKB); U.K., Sweden, U.S., China	Retrospective analysis using data from 5 cohort studies; these were population-based cohorts with varying inclusion/exclusion criteria.	115,206; *N (%) Female:* Range 0% to 100% across the 5 cohorts	Range 53.7 to 75.4 across the 5 cohorts	Varied Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* Range 9.2% to 35.4% across the 5 cohorts
Marques et al, 2017[176] Poor	3 different Portuguese cohorts (SAOL, IPR, EPIPorto); Portugal	Retrospective analysis using data from 3 Portuguese cohorts (SAOL, IPR, EPIPort) using participants age 40 years or older with complete FRAX data.	2,626; N (%) Female: 1,943 (73)	58.2 (10.2)	*Mean T-score at baseline:* T-score at FN: -1.54 (1.31) N (%) with osteoporosis: 435 (22.9) N (%) with prior fracture at baseline: 512 (19.5)
Pluskiewicz et al, 2023[312] Poor	RAC-POL-OST; Poland	Population-based random sample of postmenopausal women age 55 years or older supplemented with a nonrandom sample of additional volunteers, but it is unclear how these additional volunteers were recruited; however, no differences in baseline history of fracture or FRAX or Garvan fracture risk was observed between participants recruited randomly or nonrandomly.	978 N (%) Female: 978 (100)	65.1 (7.0)	Mean T-score at baseline: -1.24 to (0.9) at FN N (%) with prior fracture at baseline: 195 (30.5) N (%) with fall in the year prior to baseline: 207 (32.3) N (%) with secondary osteoporosis: 30 (4.8)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Tamaki et al, 2011[151] Poor	Japanese Population-Based Osteoporosis Cohort; Japan	Population-based cohort of women ages 15 to 79 years randomly selected in 5-year age groups from resident registrations in municipalities in Japan starting in 1996.	815; N (%) Female: 815 (100)	56.7 (9.6)	Women who were taking osteoporosis drugs or hormone replacement or younger than 40 years were excluded; women older than 75 years were also excluded because of low followup in that age group, women without FN BMD were also excluded. *Mean T-score at baseline:* BMD (g/cm^2) at FN: 0.71 (0.11) N (%) with prior fracture at baseline: Prior fragility fracture: 65 (8.0)
Tanaka et al, 2010[156] Poor	Multiple Japanese Cohorts; Japan	Data from participants enrolled in two Japanese cohort studies (Miyama and Taiji); these cohorts randomly selected participants ages 40 to 79 years for recruitment from resident registration records in December 1988 and the Taiji cohort enrolled participants ages 40 to 79 years randomly selected from resident registration records in June 1992; only women from these cohorts were included in this analysis of the validation dataset.	400; N (%) Female: 400 (100)	59.5 (11.3)	Mean T-score at baseline: T-score at FN: -1.6 (1.8) *N (%) with prior fracture at baseline:* NR (25) [only measured in 1 of the 2 cohorts]
Tebe Cordomi et al, 2013[144] Poor	CETIR cohort; Spain	Random sample of women identified from a database of women ages 40 to 90 years with a first visit for DXA between January 1992 and February 2008.	1,231; N (%) Female: 1,231 (100)	56.8 (7.8)	Mean T-score at baseline: T-score: -1.4 (1.1) 16% with T-scores <-2.5 N (%) with prior fracture at baseline: 185 (15%)

Appendix D Table 2. Characteristics of Included Primary Research Studies for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Study Quality	Cohort Name; Country	Cohort Description	Total N; N (%) Female	Mean Age (SD)	Additional information
Tebe et al, 2022[180] Poor	BIFAP cohort, Spain	Persons from the Base de datos para la Investigacion Farmacoepidemiologica en Atencion Primaria (BIFAP) longitudinal cohort derived from primary care medical records to conduct pharmacoepidemiological studies for the Spanish Agency for Medicines and Health Products, which covers 4 million patients from 7 regions, are included in this cohort. Only persons ages 50 to 85 years who had not been treated with any osteoporosis drugs and who have had at least 1 year of followup were included in this analysis.	1,823,217 (male portion of the cohort) N (%) Female: 0 (0)	61.8 (10.8) (full cohort including women)	Mean T-score at baseline: NR *N (%) with prior fracture at baseline:* 55,540 (1.4) (entire cohort including women)
Zwart et al, 2023[313] Poor	EPIFROS cohort; Spain	Caucasian persons ages 40 to 90 years from different regions of Spain selected from lists of participating general practitioners. Persons who received bone-conserving drugs at baseline or during 10 years of followup were excluded.	156 (54.7)	61.5 (14)	Mean T-score at baseline: NR N (%) with prior fracture at baseline: 13 (4.6)

Abbreviations: BMD=bone mineral density; CaMos=Canadian Multicentre Osteoporosis Study; DXA=dual energy X-ray absorptiometry; ECOSAP=Ecografia Osea en Atencio Primaria cohort; FDA=U.S. Food and Drug Administration; FN=femoral neck; FRAX=fracture risk assessment tool; FRIDEX=Fracture RIsk factors and bone DEnsitometry type central dual X-ray; FROCAT=abbreviation not defined; IQR=interquartile range; IU=international units; MOF=major osteoporotic fracture; MrOs=Osteoporotic Fractures in Men Cohort; N=number; NR=not reported; OFELY= Os des Femmes de Lyon; QUALYOR=QUalité Osseuse LYon Orléans; RA=rheumatoid arthritis; SOF=Study of Osteoporotic Fractures; THIN=The Health Improvement Network; U.K.=United Kingdom; U.S.=United States; WHI=Women's Health Initiative.

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Baleanu et al, 2021[179] Fracture Risk Brussels Epidemiological Enquiry (FRISBEE) Belgium High/poor	Postmenopausal women ages 60 to 85 years who were enrolled in longitudinal, prospective, population-based cohort study between 2007 and 2013 designed to evaluate and develop fracture risk prediction models.	NR; 44.5% were ≥70	3,030 (100)	Mean BMD: NR Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic System 4500 W T-score reference range used: NR
Black et al, 2018[192] Study of Osteoporotic Fractures (SOF) U.S. Some concerns/fair	Participants from a multicenter, prospective, cohort study of risks for fracture that included community-dwelling ambulatory White women age 65 years or older who were enrolled between 1986 and 1988 **Additional inclusion/exclusion criteria:** None specified by the study, but only women with BMD measures were included in the analysis for this update (n=7,959) **Proportion with prior fracture:** Any nonvertebral: 3,118 (38.4) Hip: 184 (2.3)	73.4 (5.1)	8,130 (100)	**Mean BMD:** T-score at FN: -1.4 (NR) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic 1000 **T-score reference range used:** NHANES reference database for White women; age not specified
Bolland et al, 2011[166] New Zealand Some concerns/fair	Postmenopausal women age 55 years or older with no major medical conditions who were taking part in a trial of calcium supplementation; race/ethnicity NR **Additional inclusion/exclusion criteria:** Normal lumbar spine BMD for their age (Z-score >-2), not taking treatment for osteoporosis, hormone replacement therapy, or vitamin D supplementation Proportion with prior fracture: NR (33.5)	74.2	1422 (100)	**Mean BMD:** FN T-score: -1.3 (1.0) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 11	DXA machine/software: NR T-score reference range used: NR

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Chapurlat et al, 2020[175] OFELY and QUALYOR Cohorts (2 population-based cohorts in France) France High/poor	Two population-based cohorts in France; postmenopausal women with baseline bone measurements obtained between 2006 and 2008 from the Os des Femmes de Lyon (OFELY) cohort and women from the QUALYFOR cohort who were recruited from Lyon and Orleans with a T-score between -1.0 and -2.5 with clinical risk factors or women with T-score <-3.0 without risk factors and who were followed for 5 years. **Additional inclusion/exclusion criteria:** Participants with missing FRAX data were excluded. Proportion with prior fracture: NR	OFELY: 68.0 (NR) QUALYOR: 65.9 (NR)	2,100 (100)	Mean BMD: T-score OFELY: -1.36 (NR) QUALYFOR: -1.70 (NR) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 6.8 Site 2 BMD Measurement 7.8	DXA machine/software: OFELY: QDR 4500 QUALYOR: Hologic Discovery A **T-score reference range used:** NHANES III; age and sex information of reference range used NR
Cheung et al, 2012[150] Hong Kong Osteoporosis Study Hong Kong Some concerns/fair	Southern Chinese (Hong Kong) postmenopausal women from an extended cohort of a prospective population-based cohort study. Patients were recruited from different districts between 1995 and 2009 during health fairs and road shows on osteoporosis **Additional inclusion/exclusion criteria:** Excluded if already prescribed treatment for osteoporosis **Proportion with prior fracture:** Past history of low-trauma fracture after age 45 years: 291 (12.8)	62.1 (8.5)	2,266 (100)	**Mean BMD:** T-score LS: -1.6 (1.2); TH: -1.3 (1.2); FN: -1.5 (1.1) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 30.1	DXA machine/software: Hologic QDR 4500 **T-score reference range used:** NHANES database and a local Southern Chinese normative database
Fraser et al, 2011[154] Canadian Multicentre Osteoporosis Study (CaMos) Canada Some concerns/fair	Canadian men and women randomly selected from population-based longitudinal cohort CaMos study; race/ethnicity NR **Additional inclusion/exclusion criteria:** Included if lived within a 50-km radius of one of nine Canadian cities (St John's, Halifax, Quebec City, Toronto, Hamilton, Kingston, Saskatoon, Calgary, and Vancouver) and were able to converse in English, French, or Chinese (Toronto and Vancouver). Proportion with prior fracture: 634 (9.5)	Women: 65.8 (8.8) Men: 65.3 (9.1)	4,778 (71.3)	**Mean BMD:** FN T-score Women: -1.5 (1.1) Men: -0.5 (1.2) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	**DXA machine/software:** Hologic and GE Lunar **T-score reference range used:** FN T-scores were calculated in both men and women using the NHANES III White female reference values

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Goldshtein et al, 2018[172] Maccabi Healthcare Services Israel Some concerns/fair	Women ages 50 to 90 years enrolled in Maccabi Health Care Services Health Maintenance Organization with at least 3 years membership history and a BMD test result and height and weight data. Additional inclusion/exclusion criteria: NR **Proportion with prior fracture:** Among those who sustained a fracture: NR (6.3) Among those who remained fracture free: NR (4.0)	Median 58 (IQR 54–67)	16,578 (100)	**Mean BMD:** T-score Among those who sustained a fracture: -1.8 (0.7) Among those who remained fracture free: -1.4 (0.8) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: GE Lunar Prodigy **T-score reference range used:** White women from NHANES III; age not specified
Gourlay et al, 2017[146] MrOs U.S. Some concerns/fair	Retrospective analysis from participants in the MrOs (U.S.) cohort; community-dwelling men age 65 years or older; 89% White, 4% African American, 2% Hispanic, 1% Other **Additional inclusion/exclusion criteria:** For this study's analysis, men with a prior history of hip or clinical vertebral fracture, were not taking FDA-approved antifracture treatment, and did not have osteoporosis by BMD at baseline (for the analysis of fracture risk scores calculated with BMD) **Proportion with prior fracture:** Previous fracture after age 50 years: 925 (18.7) (Note, men with prior hip or clinical vertebral fracture were not included)	73.4 (5.8)	0 (0)	**Mean BMD:** NR Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	**DXA machine/software:** NR **T-score reference range used:** NHANES III; White women ages 20 to 29 years
Iki et al, 2021[194] Japanese Population-based Osteoporosis Study Japan High/poor	Women ages 40 to 79 years randomly selected from five areas of Japan who were participants in the Japanese Population-based Osteoporosis Study Additional inclusion/exclusion criteria: NR **Proportion with prior fracture:** Osteoporotic fracture among those with hip fracture: 15 (22.1) Osteoporotic fracture among those without hip fracture: 98 (7.8) (supplementary table B)	Among those with fracture: 70.8 (5.9) Among those without fracture: 58.7 (11.0) Overall: 59.3 (11.1)	1,331 (100)	**Mean BMD:** In g/cm^2 at FN Among those with fracture: 0.588 (0.08) Among those without fracture: 0.702 (0.10) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic QDR4500A T-score reference range used: NR

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Kwok et al, 2012[182] MrOs (Hong Kong) Hong Kong Some concerns/fair	Men enrolled in large prospective population-based cohort of older (age 65 years or older) southern Chinese men recruited from local community centers **Additional inclusion/exclusion criteria:** Community dwelling, walk without assistance, did not have bilateral hip replacements Proportion with prior fracture: 267 (13.9)	72.4 (5.0)	0 (0)	Mean BMD: In gm/cm^2 FN: 0.69 (0.11) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic QDR 4500 T-score reference range used: NR

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Leslie, et al, 2010[157] Hans et al, 2011[183] Leslie, et al, 2013[184] Leslie, et al, 2016[161] Leslie, et al, 2018[162] Crandall et al, 2019[160] Agarwal et al, 2022[193] Manitoba BMD Registry Canada Some concerns/fair	From Leslie, et al, 2010[157] Men and women age 50 years or older at the time of DXA testing between January 1990 and March 2007 in the province of Manitoba, Canada; race/ethnicity NR **Additional inclusion/exclusion criteria:** Patients were required to have medical coverage from Manitoba Health during the observation period Proportion with prior fracture: Female: 4,984 (13.6) Male: 431 (15) From Hans et al, 2011[183] Leslie, et al, 2013[184] Postmenopausal women from the Canadian province of Manitoba who were part of the Manitoba Bone Density Program, a targeted case-finding clinical program Race/ethnicity NR **Additional inclusion/exclusion criteria:** Included women were 50 years or older and had medical coverage during the observation period. Proportion with prior fracture: 3,999 (13.6) From Leslie, et al, 2016[161] Men and women age 50 years or older with BMD testing after January 1, 1996, and at least 5 years of followup **Additional inclusion/exclusion criteria:** Persons already receiving osteoporosis therapy were excluded. Proportion with prior fracture: 4,903 (14.4)	Leslie, et al, 2010[157] Female: 65.7 (9.8) Male: 68.2 (10.1) Leslie, et al, 2016[161] 66.6 (9.8) Hans et al, 2011[183] Leslie, et al, 2013[184] 65.4 (9.5) Leslie, et al, 2018[162] Men 66.0 (12.2) Women 64.1 (11.1) Crandall et al, 2019[160] 63.9 (11.2) Agarwal et al, 2022[193] Women: 66.6 (9) Men: 69 (10)	Leslie, et al, 2010[157] 36,730 (92.7) Leslie, et al, 2016[161] 31,007 (91) Hans et al, 2011[183] Leslie, et al, 2013[184] 29,407 (100) Leslie, et al, 2018[162] 62,275 (91.1) Crandall et al, 2019[160] 54,459 (100) Agarwal et al, 2022[193] 16,682 (85)	From Leslie, et al, 2010[157] **Mean BMD:** T-score at FN Female: -1.5 (1.0) Male: -1.2 (1.1) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement From Hans et al, 2011[183] Leslie, et al, 2013[184] 29,407 (100) **Mean BMD:** T-score Lumbar spine: -1.19 (1.50) FN: -1.47 (0.94) TH: -1.03 (1.16) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 13 Site 2 BMD Measurement 9.6 Site 3 BMD Measurement 24.3 From Leslie, et al, 2016[161]	From Leslie, et al, 2010[157] **DXA machine/software:** Lunar DPX and Lunar Prodigy **T-score reference range used:** Hip T-scores were calculated using NHANES III White female reference values. Lumbar spine T-scores were calculated using the manufacturer's U.S. White female reference values after vertebral levels affected by artifact were excluded by experienced clinician

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Leslie, et al, 2010[157] Hans et al, 2011[183] Leslie, et al, 2013[184] Leslie, et al, 2016[161] Leslie, et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry Canada Some concerns/fair (continued)	From Leslie, et al, 2018[162] Men and women age 40 years or older with BMD measurement between 1996 and 2013 Additional inclusion/exclusion criteria: NR Proportion with prior fracture: Men 1,179 (18.3) Women 8,833 (14.2) From Crandall et al, 2019[160] Initial DXA between 1996 and 2011; sample limited to women age 40 years or older; race/ethnicity NR Additional inclusion/exclusion criteria: NR Proportion with prior fracture: Prior MOF 7,570 (13.9) From Agarwal et al, 2022[193] Men and women ages 50 to 95 years with DXA between 2012 and 2018 **Additional inclusion/exclusion criteria:** Missing baseline data for clinical risk assessment **Proportion with prior fracture:** Women: 3,612 (22); Men: 669 (24)			**Mean BMD:** T-score -1.5 (1.0) at FN T-score -2.0 (1.1) at FN or TH or LS Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 31.7 From Leslie, et al, 2018[162] Mean BMD: T-score Men -1.1 (1.1) Women -1.4 (1.0) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 12.2 Site 2 BMD Measurement 8.4	
Leslie, et al, 2010[157] Hans et al, 2011[183] Leslie, et al, 2013[184] Leslie, et al, 2016[161] Leslie, et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry Canada Some concerns/fair (continued)				From Crandall et al, 2019[160] **Mean BMD:** T-score FN: -1.4 (1.0) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement FN 12.2 Agarwal et al, 2022[193] **Mean BMD:** T-score Women: FN -1.5 (1) Men: FN -1.4 (1.2) Prevalence (%) of osteoporosis: NR	

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Marques et al, 2017[176] SAOL, IPR, and EPIPorto (3 Portuguese cohorts) Portugal High/poor	Persons age 40 years or older identified from 3 different Portuguese cohort studies with complete FRAX and FN BMD data available Additional inclusion/exclusion criteria: NR Proportion with prior fracture: 512 (19.5)	58.2 (10.2)	1,943 (73.0) (for the entire cohort including those for whom BMD was not available)	**Mean BMD:** T-score -1.54 (1.31) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 22.9 Site 2 BMD Measurement 20.8 Site 3 BMD Measurement 23.4	DXA machine/software: Hologic QDR 4500/c **T-score reference range used:** NHANES III references rages; age and sex information NR
Nguyen et al, 2004[186] Dubbo Osteoporosis Epidemiology Study (DOES) Australia High/poor	Subset of women from the DOES; race/ethnicity NR Additional inclusion/exclusion criteria: NR Proportion with prior fracture: NR	65.2 (12.3)	549 (100)	Mean BMD: NR Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Lunar DPX-L **T-score reference range used:** Manufacturer's reference ranges
Prince et al, 2019[191] Perth Longitudinal Study of Aging in Women (PLSAW) Australia Some concerns/fair	Women age 70 years or older who were enrolled in an RCT to evaluate the use of oral calcium supplements; participants in this trial were recruited from electoral rolls. Participants were then invited to participation in study followup without intervention for an additional 10 years as part of the study. Only participants with BMD data from this study were included in this update (n=1,057). 99% Caucasian. **Additional inclusion/exclusion criteria:** For enrollment in the initial trial: ambulant, life expectancy ≥5 years; not using any medication known to affect bone metabolism **Proportion with prior fracture:** Among participants with baseline vertebral fracture: 45 (45) Among participants without baseline vertebral fracture: 248 (25.2)	Among participants with baseline vertebral fracture: 75.1 (2.7) Among participants without baseline vertebral fracture: 74.9 (2.6)	1,084 (100)	**Mean BMD:** T-score at FN Among participants with baseline vertebral fracture: -1.59 (0.97) Among participants without baseline vertebral fracture: -1.39 (0.85) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 8.4	**DXA machine/software:** Hologic Acclaim 4500A T-score reference range used: NR

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Robbins et al, 2007[187] Women's Health Initiative (WHI) U.S. Some concerns/fair	A subset of women ages 50 to 79 years from 3 sites in the WHI who had undergone DXA testing; 89% White. Additional inclusion/exclusion criteria: NR Proportion with prior fracture: NR	NR	10,749 (100)	**Mean BMD:** NR Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 4.9	**DXA machine/software:** NR **T-score reference range used:** NR
Sornay-Rendu et al, 2010[188] Os des Femmes de Lyon (OFELY) cohort France Some concerns/fair	Women post and premenopausal, age 40 years or older Additional inclusion/exclusion criteria: NR Proportion with prior fracture: 89 (10.3)	58.8 (10.3)	867 (100)	**Mean BMD:** FN T-score: -1.2 (1.0) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic QDR 2000 **T-score reference range used:** FN T-scores calculated based on NHANES III reference values
Stewart et al, 2006[190] Aberdeen Prospective Osteoporosis Screening Study (APOSS) U.K. Some concerns/fair	Women ages 45 to 54 years who underwent BMD measurement between 1990 and 1994 as part of a population-based cohort study **Additional inclusion/exclusion criteria:** No menses within the prior 6 months; treatment for osteoporosis was allowed Proportion with prior fracture: NR	48.6 (2.4)	3,883 (100)	Mean BMD: In g/cm^2 LS: 1.052 (0.161) FN: 0.881 (0.125) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Norland XR-26 T-score reference range used: NR
Sund et al, 2014[185] Kuopio Osteoporosis Risk Factor and Prevention (OSTPRE) Finland High/poor	Postmenopausal women with clinical risk factors originally recruited from a population-based mail survey; women included in this analysis were a subset that had available clinical information including FN BMD **Additional inclusion/exclusion criteria:** Excluded women with hip fractures before 1994 Proportion with prior fracture: 551 (20.0)	59.1 (2.9)	2,755 (100)	**Mean BMD:** T-score: -1.0 (0.91) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: NR T-score reference range used: NR
Tamaki et al, 2011[151] Japanese Population-Based Osteoporosis Study (JPOS) Japan High/poor	Japanese women ages 40 to 74 years randomly selected from 5-year age groups using resident registrations from three areas in seven municipalities throughout Japan **Additional inclusion/exclusion criteria:** Excluded if no FN BMD, taking osteoporosis drugs or hormone replacement therapy Proportion with prior fracture: 65 (8.0)	56.7 (9.6)	815 (100)	**Mean BMD:** In g/cm^2: 706 (0.111) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	DXA machine/software: Hologic QDR 4500A T-score reference range used: NR

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Name Country ROB/Study Quality	Population	Mean Age (SD)	N (%) Female	Mean BMD and Prevalence of Osteoporosis	Reference Test Details
Tanaka et al, 2010[156] Miyama and Taiji Cohorts Japan High/poor	Women from the Miyama Cohort who were selected from Miyama village's resident registration in 1988 as part of nationwide community-based cohort studies, and women from the Taiji Cohort, a community-based cohort study created in 1992 with participants selected from Taiji town's resident registration **Additional inclusion/exclusion criteria:** Excluded if had metabolic bone disease or secondary osteoporosis (e.g., hyperparathyroidism, hyperthyroidism other than patients on T4 replacement and with euthyroid for more than one year, chronic renal failure or osteomalacia) Proportion with prior fracture: NR (25)	59.5 (11.3)	400 (100)	**Mean BMD:** Lumbar T-score: -1.36 (1.19) FN T-score: -1.61 (1.84) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement NR	**DXA machine/software:** Lunar DPX and Hologic QDR-1000 T-score reference range used: NR
Trajanoska et al, 2018[15] Rotterdam Study The Netherlands High/poor[15]	Dutch persons age 45 years or older enrolled in a population-based prospective cohort study over 3 waves between 1990 and 2006; race/ethnicity NR Additional inclusion/exclusion criteria: NR Proportion with prior fracture: NR	Men 64.7 (9.4) Women 66.5 (10.9)	6,275 (57)	Mean BMD: NR Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 6.7 Site 2 BMD Measurement 11.1	**DXA machine/software:** Lunar DPX-L for waves 1 and 2; GE Lunar Prodigy for wave 3; data calibrated across the 3 waves for comparability. **T-score reference range used:** Sex-specific NHANES III young healthy reference population
Tremollieres et al, 2010[189] Menopause et Os (MENOS) Study France High/poor	Postmenopausal women older than 45 years enrolled in the MENOS cohort study; a study designed to assess whether bone mass at menopause is a predictor of different diseases. **Additional inclusion/exclusion criteria:** Excluded: past/current treatment for osteoporosis for >3 months, HRT use at baseline Proportion with prior fracture: Women with fracture: 12 (8.3) Women without fracture: 43 (2.1)	Women with fracture: 54.8 (4.3) Women without fracture: 53.4 (4.2)	2,196 (100)	**Mean BMD:** In g/cm^2 at LS Women with fracture: 0.96 (0.126) Women without fracture: 1.03 (0.148) Prevalence (%) of osteoporosis based on: Site 1 BMD measurement 8.8	DXA machine/software: DPX-IQ, Lunar GE **T-score reference range used:** Reference ranges from authors' own normative database for women ages 25–35 years

Abbreviations: BMD=bone mineral density; CaMos=Canadian Multicentre Osteoporosis Study; DOES=Dubbo Osteoporosis Epidemiology Study; DXA=dual-energy X-ray absorptiometry; FN=femoral neck; FRAX=Fracture Risk Assessment Tool; HRT=hormone replacement therapy; IQR=interquartile range; LS=lumbar spine; MOF=major

Appendix D Table 3. Characteristics of Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

osteoporotic fracture; N=number; NHANES=National Health and Nutrition Examination Survey; NR=not reported; OFELY=Os des Femmes de Lyon; QUALYOR= QUalité Osseuse LYon Orléans; RCT=randomized, controlled trial; ROB=risk of bias; SD=standard deviation; TH=total hip; U.S.=United States.

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Adler et al, 2003[221] U.S. Fair	Men enrolled in a pulmonary clinic and a rheumatology clinic at a single VA medical center; patients with previous DXA testing ineligible. 69% White, 30% Black, 2% other	64.3 (12.3)	0 (0)	Site of BMD measurement 1 FN or TH or LS 28/181 (15.5%)	DXA machine/software: Hologic QDR 4500 **T-score reference range used:** NHANES reference database for hip. Hologic reference source for spine, age, sex, race of reference group not reported Interval between risk assessment and BMD testing: 1 month
Bansal et al, 2015[220] Pecina et al, 2016[234] U.S. Fair	All women between the ages of 50 and 64.5 years who underwent DXA during a 6-month period and were enrolled in a primary care practice of the Mayo Clinic in Rochester, MN; 97.9% White, 1.4% Asian, 0.3% Hispanic, 0.3% Black.	From Bansal et al, 2015[220] 57.4 (NR) From Pecina et al, 2016[234] 56.6 (3.4)	From Bansal et al, 2015[220] 464 (100) From Pecina et al, 2016[234] 290 (100)	From Bansal et al, 2015[220] Site of BMD measurement 1 FN or LS 120/464 (25.9%) From Pecina et al, 2016[234] Site of BMD measurement 1 Site: FN or LS 50/290 (17.2%) Site of BMD measurement 2 Site: LS 41/290 (14.1%) Site of BMD measurement 3 Site: FN 19/290 (6.6%)	DXA machine/software: NR T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Brenneman et al, 2003[222] OPRA U.S. Fair	Postmenopausal women ages 60–79 years in the Osteoporosis Population-based Risk Assessment (OPRA) study, which enrolled participants from Group Health Cooperative in the U.S.	69.3 (5.5)	416 (100)	Site of BMD measurement 1 FN or TH or LS 126/416 (30.3%)	DXA machine/software: Hologic QDR 2000 **T-score reference range used:** NHANES III, does not specify age or sex of reference group Interval between risk assessment and BMD testing: Concurrent

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Cadarette et al, 2004[223] Canada Fair	Caucasian women age 45 years or younger assembled from either 1) presenting for BMD testing from a university-based ambulatory health center or 2) women with DXA results retrospectively assembled from university-affiliated family practices	62.4 (11.2)	190 (100)	Site of BMD measurement 1 FN or LS 106/644 (16.5%)	**DXA machine/software:** Multiple machines used: Hologic, Lunar, Norland, Unknown T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Cadarette et al, 2001[206] Canadian Multicentre Osteoporosis Study (CaMOS) Canada Fair	Women from the general population recruited from 1996–1997: 96.6% White, 1.8% Asian, 0.3% Black, 1.3% Other.	66.4 (8.8)	2,365 (100)	Site of BMD Measurement 1 FN 239/2,365 (10.1%)	DXA Machine/software: Hologic QDR 4500 Hologic QDR 2000 Hologic QDR 1000 Lunar DPX **T-score reference range used**: Canadian young adult normal values at the FN. (Authors noted the Canadian young adult normal reference at the FN (mean [SD], 0.857 [0.125] g/cm^3) is similar to that reported by NHANES III for non-Hispanic White Americans (mean [SD], 0.858 [0.120] g/cm^3) Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Cass et al, 2016[232] Shepherd et al, 2010[202] NHANES U.S. Fair	From Cass et al[232] Men age 50 years or older in the NHANES III cohort (1988–1994) with a valid DXA scan; 88.5% were non-Hispanic White, 8.5% were African American, and 2.9% were Mexican American. From Shepherd et al[202] Men age 50 years or older with DXA scan in any of the NHANES 1999 to 2000, 2001 to 2002, and 2003 to 2004 datasets; 81% White, 8.2% African American, 3.6% Mexican American, 7.2% Other	From Cass et al[232] 64.2 (9.7) From Shepherd et al[202] 63 (NR)	0 (0)	From Cass et al[232] Site of BMD Measurement 1 TH or FN 68/1,498 (4.5%) From Shepherd et al[202] Site of BMD measurement 1 Site: FN or TH or LS 303/2,944 (10.3%) Site of BMD measurement 2 Site: LS 126/2,944 (4.3%)	From Cass et al[232] DXA Machine/software: NR **T-score reference range used**: NHANES III non-Hispanic White women ages 20–29 years **Interval between risk assessment and BMD testing:** NR but likely reasonably concurrent since NHANES enrolls persons prospectively From Shepherd et al[202] DXA machine/software: Hologic QDR-4500A T-score reference range used: White men ages 20–29 years **Interval between risk assessment and BMD testing:** NR but likely reasonably concurrent since NHANES enrolls persons prospectively
Cass et al, 2006[224] U.S. Fair	Postmenopausal women, age 45 years or older (receiving usual care at university-based family practice clinic in the U.S.). 29% White, 43% African American, 28% Hispanic	60.2 (9.6)	226 (100)	Site of BMD Measurement 1 TH or LS 22/203 (10.8%) Site of BMD Measurement 2 Site: LS 16/203 (7.9%) Site of BMD Measurement 3 Site: TH 2/203 (1%)	DXA Machine/software: Hologic QDR 4500A **T-score reference range used**: Manufacturer's reference ranges **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment because subjects were enrolled prospectively.
Cass et al, 2013[197]	Men age 60 years or older who attended university-based primary care clinics for usual care: 76% non-Hispanic White, 11.8% African American, 10.7% Hispanic, 1.4% Other	70.2 (6.9)	0 (0)	Site of BMD Measurement 1 FN or TH 15/346 (4.3%)	**DXA Machine/software**: Hologic QDR 4500A or GE Lunar iDXA **T-score reference range used**: FN and TH T-scores calculated based on NHANES III non-Hispanic White women ages 20–29 years Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Chan et al, 2006[207] Singapore Fair	Free-living ambulant postmenopausal women (Tanjong Rhu community), age 55 years or older	68.4 (5.5)	135 (100)	Site of BMD Measurement 1 FN 33/135 (24.4%) Site of BMD Measurement 2 Site: LS 37/135 (27.4%)	**DXA Machine/software**: DXA (Hologic QDR 4500A) T-score reference range used: NR **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment since subjects were enrolled prospectively
Chang et al, 2016[240] Taiwan Fair	Men who required BMD examinations at a large teaching hospital between 2009 and 2012	71.9 (13.3)	0 (0)	Site of BMD Measurement 1 FN 321/834 (38.5%)	DXA Machine/software: NR T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Chen et al, 2016[233] Taiwan Fair	Community-dwelling ambulant persons age 60 years or older at community centers between July-December 2012 who had a registered household in Tanzi District without severe cardiopulmonary disease	67.4 (6.4)	367 (66)	Site of BMD Measurement 1 FN 97/553 (17.5%)	**DXA Machine/software**: Hologic Discovery Wi Bone Densitometer T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Christodoulou et al, 2016[242] Greece Fair	Postmenopausal women without prior use of medication for osteoporosis, recruited between October 2012 and October 2014 to a tertiary care center; race/ethnicity NR	63.4 (NR)	1,000 (100)	Site of BMD Measurement 1 NR NR/1,000 (0%)	DXA Machine/software: NR T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Cook et al, 2005[208] U.K. Fair	Postmenopausal women through natural or unnatural causes, referred by GPs or hospital-based clinics because of one or more clinical risk factors for osteoporosis in the U.K.; race not reported	59.7 (NR)	208 (100)	Site of BMD Measurement 1 LS or TH 45/208 (21.6%)	DXA Machine/software: Hologic QDR-4500C **T-score reference range used**: T-scores were computed using the databases supplied with the DXA systems. Interval between risk assessment and BMD testing: Concurrent

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Crandall et al, 2014[195] Crandall et al, 2019[141] Crandall et al, 2023[143]Women's Health Initiative U.S. Fair	Subset of participants from the WHI clinical trial or observational study, which were studies in postmenopausal women ages 50 to 79 years who were free from serious medical conditions and not taking medications known to influence BMD who underwent DXA testing at baseline at 3 of the WHI clinical sites. The 2014 and 2023 analyses included women ages 50 to 64[143, 195] and the 2019 analysis included women of any age who had at least 10 years of followup and had information relevant to the NOF risk algorithm and hormone users were not excluded; however, only the data for women ages 50 to 64 were used because the data for older women are not relevant since the strategies assessd were strictly age-based.[141] The 2023 analysis included women with self-reported race/ethnicity and excluded women with missing covariate information or who were taking medication.[143] Race/ethnicity: From 2014 analysis: 72% White, 17% Black, 8% Hispanic[195] From 2019 analysis: 86% White (based on full WHI population, not the subset used in this analysis)	57.7 (based on 5,165 participants, but only 2,857 non-users of hormone therapy were used in the analysis)[195] 57.8 (4.1)[143] 62.7 (7.1)[141]	N=2,857 (100) (non-users of hormone therapy)[195] N=8,134 (4,805 ages 50 to 64 years) (100)[141] N=4,607[143]	Site of BMD Measurement From Crandall et al[195] Based on FN (among nonusers of hormone therapy) 174/2,857 (6.1%) From Crandall et al[141] Based on any site 1642/8,134 (20.2%) for all ages 682/4,805 (14.2%) for ages 50–64 From Crandall et al[143] Based on FN only 235/4,607 (5.1%) Based on any site: 653/4,607 (14.1%)[143]	**DXA Machine/software**: Hologic QDR2000 or 4500W **T-score reference range used**: FN T-scores calculated based on NHANES III normative reference database (presumably young non-Hispanic White females ages 20–29 years) **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment since subjects were enrolled prospectively.

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
D'Amelio et al, 2005[225] Italy Fair	Postmenopausal Caucasian women referred to university bone metabolic unit for DXA. 13% were noted to have secondary osteoporosis.	Normal BMD: 57.3 (6.6) Osteopenic BMD: 60.2 (7.8) Osteoporotic BMD: 62.2 (6.7)	525 (100)	Site of BMD Measurement 1 LS or FN 249/525 (47.4%)	DXA Machine/software: Hologic QDR 4500 T-score reference range used: NR Interval between risk assessment and BMD testing: NR
D'Amelio et al, 2013[199] Italy Fair	Menopausal women from general practices in Italy	65.0 (8)	995 (100)	Site of BMD Measurement 1 FN or LS 335/995 (33.7%)	**DXA Machine/software**: DXA (Hologic QDR 4500A), software version NR T-score reference range used: NR **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment because subjects were enrolled prospectively.
Diem et al, 2017[235] Lynn et al, 2008[218] MrOS Fair Multicountry (including U.S.)	From Diem et al, 2017[235] Community-dwelling, ambulatory men, age 65 years or older recruited using population-based listings at six settings in Birmingham, AL; Minneapolis, MN; Palo Alto, CA; Pittsburgh, PA; Portland, OR; and San Diego, CA. From Lynn et al, 2008[218] As above plus Hong Kong participants were recruited using a combination of private solicitation and public advertising from community centers, housing estates, and the general community. Men who had bilateral hip replacements or who were unable to walk without the assistance of another person were excluded.	From Diem et al, 2017[235] 76.3 (4.8) From Lynn et al, 2008[218] (NR) all age >65	0 (0)	From Diem et al, 2017[235] Site of BMD Measurement 1 FN or TH or LS 216/4,043 (5.3%) From Lynn et al, 2008[218] Site of BMD measurement 1 Site: FN (U.S.) 233/4,658 (5%) Site of BMD measurement 2 Site: FN (Hong Kong) 96/1,914 (5%) Site of BMD measurement 3 Site: LS (U.S.) 138/4,658 (3%) Site of BMD measurement 4 Site: LS (Hong Kong) 38/1,914 (2%)	From Diem et al, 2017[235] **DXA Machine/software**: Hologic QDR 4500 W T-score reference range used: From Diem et al, 2017[235] White women ages 20–29 years from NHANES III From Lynn et al, 2008[218] U.S.: Caucasian male normative reference database from NHANES Hong Kong: local Chinese reference ranges Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Erjiang et al, 2021[243] Ireland Fair	Caucasian men and women age 40 years who had DXA scans ordered by their clinicians from January 2000 to November 2018 at 3 sites	Female 61.4 (10.9) Male 64.9 (11.7)	15,964 (85.5)	Site of BMD Measurement 1 FN or TH or LS 4,064/18,670 (21.8%) Site of BMD Measurement 2 Site: FN or TH or LS, women 3,467/15,964 (21.7%) Site of BMD Measurement 3 Site: FN or TH or LS, men 597/2,706 (22.1%)	DXA Machine/software: GE Lunar **T-score reference range used**: NHANES III/U.S. White female reference range Interval between risk assessment and BMD testing: NR
Geusens et al, 2002[231] U.S. Fair	Community-dwelling women age 45 years or older and 82% White who were screened from across 11 sites for the Fracture Intervention Trial	61.3 (9.6) (for the U.S. sample)	1,102 (100)	Site of BMD measurement 1 FN 152/1,102 (13.8%)	**DXA machine/software:** The brand of DXA manufacturer varied among centers, and included Norland, Hologic, and Lunar machines. **T-score reference range used:** FN: non-Hispanic female White women ages 20–29 years (NHANES) LS: unclear Interval between risk assessment and BMD testing: NR
Gourlay et al, 2008[230] Study of Osteoporotic Fractures U.S. Fair	Study of Osteoporotic Fractures inception cohort; a population-based cohort of women age 65 years or older recruited from 4 U.S. sites	2,714 (34.9%) ≥75 years 5,065 (65.1%) ages 67–74	7,779 (100)	Site of BMD measurement 1 FN 1,562/7,779 (20.1%)	DXA machine/software: Hologic **T-score reference range used:** FN: non-Hispanic female White women age 20–29 years (NHANES) LS: manufacturers norms for women age 30 years Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Gourlay et al, 2005[209] Richy et al, 2004[210] Ben Sedrine et al, 2001[211] Belgium Fair	Caucasian women either consulting spontaneously or referred for a BMD measurement between January 1996 and September 1999 to an osteoporosis outpatient center in Liege	61.5 (8.8) Age 65–54 years: 2,539 (63%) Age ≥65 years: 1,496 (37%)	4,035 (100)	Site of BMD measurement 1 FN[209] 383/4,035 (9.5%) Site of BMD measurement 2 FN or TH or LS[210] 1,291/4,035 (32%) Site of BMD measurement 3 Site: TH[210] 383/4,035 (9.5%) Site of BMD measurement 4 Site: LS[210] 981/4,035 (24.3%) Site of BMD measurement 5 Site: FN[210] 747/4,035 (18.5%)	**DXA machine/software:** Hologic QDR 1000, 2000, and 45000 densitometers **T-score reference range used:** multiple cited; T-score reference range was NHANES III NHW women ages 20–29 years[209] or reference values for the population of Liege, Belgium[210] Interval between risk assessment and BMD testing: NR
Hamdy et al, 2018[238] U.S. Fair	Caucasian men ages 50 to 70 years referred to an Osteoporosis Center and not currently on anti-osteoporosis medications or had history of secondary osteoporosis, low serum vitamin D levels, or diseases affecting bone metabolism	61.2 (4.8)	0 (0)	Site of BMD measurement 1 FN 86/726 (11.8%)	DXA machine/software: NR **T-score reference range used:** 1) FN T-score compared with young healthy Caucasian female reference population, and 2) lowest T-score of FN, TH, and LS compared with a young healthy male reference population (which we did not abstract) Interval between risk assessment and BMD testing: NR
Harrison et al, 2006[217] U.K. Fair	Caucasian females ages 55 to 70 years who were referred to Clinical Radiology, Imaging Science and Biomedical Engineering, University of Manchester for bone densitometry scans because of suggested osteopenia on radiographs, low-trauma fracture	61 (4)	207 (100)	Site of BMD measurement 1 FN or TH or LS 70/207 (33.8%)	**DXA machine/software:** GE Lunar Prodigy or the Hologic Discovery **T-score reference range used:** Hologic reference data for the LS and NHANES reference data for the proximal femur Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Inderjeeth et al, 2020[239] Australia Fair	Women and men age 70 years or older identified by 3 large outer metropolitan general practices including one co-located within a residential care facility and supported by a tertiary hospital's Fracture Liaison Service; no prior fragility fracture, glucocorticoid use, or rheumatoid arthritis. Race/ethnicity data NR	78.0 (5.7)	238 (44.8)	Site of BMD measurement 1 FN or TH or LS or Forearm 130/531 (24.5%)	DXA machine/software: GE Lunar Prodigy T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Jiang et al, 2016[236] U.S. Fair	Postmenopausal women ages 50 to 64 years presenting for DXA screening at a single health center between January 1, 2007, and March 1, 2009; 95.1% White, 2.0%, Black, 1.8% Hispanic/Latino, and 1.1% Asian/other	57.2 (4.2)	445 (100)	Site of BMD measurement 1 NR 38/445 (8.5%)	**DXA machine/software:** All four testing sites belong to the same institution using bone densitometers of the same make and model that was NR. T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Jimenez-Nunez et al, 2013[203] Spain Fair	Caucasian women age 50 years or older and menopausal 12 months or more, in good general health, without prior diagnosis of osteoporosis. 60% recruits from primary care, 40% from specialty clinics	61 (7)	505 (100)	Site of BMD measurement 1 FN or LS 101/505 (20%)	**DXA machine/software:** GE Lunar Prodigy Advance DXA densitometer **T-score reference range used:** Manufacturer's reference for the Spanish population Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Kung et al, 2005[212] Kung et al, 2003[213] Hong Kong Fair	Men age 50 years or older and postmenopausal women recruited from the community. Postmenopausal women recruited from the community	From Kung et al, 2005[212] Men: 64 (range 50–90) From Kung et al, 2003[213] Women: 62 (8) 62 (8)	722 (67)	From Kung et al, 2005[212] Men Site of BMD measurement 1 Site: FN or LS 56/356 (15.7%) Site of BMD measurement 2 Site: LS 36/356 (10.1%) Site of BMD measurement 3 Site: FN 40/356 (11.2%) From Kung et al, 2003[213] Women Site of BMD measurement 1 Site: FN or LS 272/722 (37.7%) Site of BMD measurement 2 Site: LS 221/722 (30.6%) Site of BMD measurement 3 Site: FN 155/722 (21.5%)	From Kung et al, 2005[212] **DXA machine/software:** QDR 2000 Plus, Hologic **T-score reference range used:** Young healthy males ages 20–39 years from local area Interval between risk assessment and BMD testing: NR From Kung et al, 2003[213] **DXA machine/software:** Sahara ultrasound bone densitometer (Hologic) **T-score reference range used:** Peak young Chinese mean values, source NR Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Leslie, et al, 2013[200] Morin et al, 2009[159] Manitoba BMD Registry Canada Fair	From Leslie, et al, 2013[200] Population-based sample of all women ages 50 to 64 years with medical coverage and valid DXA measurements from the LS and hip in Manitoba, Canada from 1990–March 2007 From Morin et al, 2009[159] Population-based sample of all women ages 40 to 59 years or older who received DXA testing in Manitoba. Note: criteria for BMD testing in women younger than age 65 years include premature ovarian failure, history of steroid use, prior fracture, x-ray evidence of osteopenia	From Leslie, et al, 2013[200] 57 (4) From Morin et al, 2009[159] 52.7 (4.9)	From Leslie, et al, 2013[200] 18,315 (100) From Morin et al, 2009[159] 8,254 (100)	From Leslie, et al, 2013[200] Site of BMD measurement 1 FN or TH or LS 3,437/18,315 (18.8%) From Morin et al, 2009[159] Site of BMD measurement 1 FN or TH or LS 1,226/8,254 (14.9%)	**DXA machine/software:** Lunar DPX prior to 2000; Lunar Prodigy 2000 and later **T-score reference range used:** FN T-scores calculated based on NHANES III White female reference; LS used T-scores from manufacturer's U.S. White female reference values Interval between risk assessment and BMD testing: NR
Machado et al, 2010[201] Portugal Fair	Population-based sample of men age 50 years or older randomly selected from 19,000 registered voters between 1998–1999	63.8 (8.2)	0 (0)	Site of BMD measurement 1 FN or TH or LS 34/202 (16.8%) Site of BMD measurement 2 Site: LS 30/202 (14.9%) Site of BMD measurement 3 Site: FN 10/202 (5%) Site of BMD measurement 4 Site: TH 2/202 (1%)	DXA machine/software: Hologic QDR 4500/c **T-score reference range used:** NHANES III young normal references values (sex unspecified) for FN; manufacturer's database for LS from male Caucasian references values (age unspecified). Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Martinez-Aguila et al, 2007[214] Spain Fair	Postmenopausal women ages 40 to 69 years referred to a local bone densitometry unit from local gynecologists; 24% with history of prior fracture; race/ethnicity NR	54.2 (5.4)	665 (100)	Site of BMD measurement 1 FN or LS 117/665 (17.6%) Site of BMD measurement 2 Site: LS 111/665 (16.7%) Site of BMD measurement 3 Site: FN 25/665 (3.8%)	DXA machine/software: Hologic QDR **T-score reference range used:** Reference ranges peak bone mass from a study conducted in a Spanish population of healthy subjects of same sex **Interval between risk assessment and BMD testing:** NR but study was done retrospectively and subjects were asked to answer questions in relation to the date of DXA scans.
Mauck et al, 2005[226] U.S. Fair[226]	Population-based sample of postmenopausal women age 45 years or older in Rochester, MN, 99% White	69.2 (11.9) Ages 45 to 64: 79 (39%)	202 (100)	Site of BMD measurement 1 FN 69/202 (34.2%) Site of BMD measurement 2 Site: LS 14/202 (6.9%)	**DXA machine/software:** Hologic QDR2000 instrument **T-score reference range used:** References ranges for young healthy women ages 20–29 years in the local community area Interval between risk assessment and BMD testing: Concurrent
McLeod et al, 2015[205] Canada Good	Women referred for screening in the Regina General Hospital, Saskatchewan, between 2010 and 2011 with no prior testing; primarily Caucasian	59 (6.7)	174 (100)	Site of BMD measurement 1 FN or LS 18/174 (10.3%)	**DXA machine/software:** GE Lunar Prodigy densitometer **T-score reference range used:** NHANES III young healthy Caucasian reference values Interval between risk assessment and BMD testing: 3 weeks
Nguyen et al, 2004[227] Dubbo Osteoporo-sis Epidemiology Study Australia Fair	Women from the Dubbo Osteoporosis Epidemiology Study, a population-based cohort; 98.6% White	70.5 (7.5)	410 (100)	Site of BMD measurement 1 FN or TH or LS 170/410 (41.5%) Site of BMD measurement 2 Site: FN 123/410 (30%) Site of BMD measurement 3 Site: LS 107/410 (26.1%)	DXA machine/software: LUNAR DPX-L **T-score reference range used:** Young Australian women at either the FN or LS Interval between risk assessment and BMD testing: concurrent

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Oh et al, 2013[204] Oh et al, 2016[229] Moon et al, 2016[244] KNHANES Republic of South Korea Fair	From Oh et al[204] Postmenopausal women age 50 years or older selected from the KHANES dataset; persons with missing BMD, previously diagnosed osteoporosis or treatment or bed ridden were excluded From Oh et al[229] Population-based sample of men age 50 years or older in the KNHANES dataset; Republic of Korea From Moon et al, 2016[244] Men ages 50 to 69 years who completed the Korea National Health and Nutrition Examination Survey between 2008–2011 excluding those with chronic liver disease, chronic kidney disease, thyroid disease, rheumatoid arthritis, asthma, or any malignancy	From Oh et al[204] 62.3 (8.2) From Oh et al[229] 63.5 (8.3) From Moon et al, 2016[244] 57.6 (0.13)	From Oh et al[204] 1,046 (100) From Oh et al[229] 0 (0) From Moon et al, 2016[244] 0 (0)	From Oh et al[204] Site of BMD measurement 1 FN or TH or LS 310/1,046 (29.6%) Site of BMD measurement 2 Site: FN 155/1,046 (14.8%) Site of BMD measurement 3 Site: LS 252/1,046 (24.1%) From Oh et al[229] Site of BMD measurement 1 FN or LS 91/1,110 (8.2%) Site of BMD measurement 2 Site: FN 35/1,110 (3.2%) Site of BMD measurement 3 Site: LS 73/1,110 (6.6%) From Moon et al, 2016[244] Site of BMD measurement 1 Mean T-score from FN, TH, and LS 139/2,519 (5.5%)	From Oh et al[204] **DXA machine/software:** QDR Discovery fan-beam densitometer (Hologic) **T-score reference range used:** Sex-specific normal values for young Japanese women **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment since subjects were enrolled prospectively From Oh et al[229] DXA machine/software: Hologic **T-score reference range used:** Gender-specific norms for young Japanese men Interval between risk assessment and BMD testing: Subjects were enrolled prospectively From Moon et al, 2016[244] DXA machine/software: Hologic T-score reference range used: NR
Pang et al, 2014[196] Australia Fair	Men and women age 70 years or older who presented to a participating GP, excluded persons with prior h/o fracture or who were taking anti-osteoporosis medications	78.2 (5.8)	282 (45.1)	Site of BMD measurement 1 FN 47/626 (7.5%) Site of BMD measurement 2 Site: TH 34/626 (5.4%) Site of BMD measurement 3 Site: LS 32/626 (5.1%) Site of BMD measurement 4 Site: Any 77/626 (12.3%)	**DXA machine/software:** Lunar Prodigy limited fan-beam machine, NR **T-score reference range used:** Manufacturer's sex-specific normative database and an ethnic database **Interval between risk assessment and BMD testing:** Not specifically indicated but appears to have been done shortly after enrollment since subjects were enrolled prospectively

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Park et al, 2003[215] Republic of Korea Fair	Postmenopausal women at a menopause clinic in Korea who were not currently using hormone replacement therapy	59.1 (7.7)	1,101 (100)	Site of BMD measurement 1 FN 119/1,101 (10.8%)	**DXA machine/software:** GE Lunar Model DPQ-IQ **T-score reference range used:** Reference range for young Korean women Interval between risk assessment and BMD testing: NR
Richards et al, 2014[198] U.S. Fair	Men older than 50 years attending primary care clinics at 4 participating VA Medical Centers. 72.2% Caucasian, 25.1% African American, the remaining 2.7% were Hispanic, Asian, and other ethnic groups	66 (NR)	0 (0)	Site of BMD measurement 1 FN or TH 92/518 (17.8%)	**DXA machine/software:** DXA on either the Hologic (Hologic Inc., Bedford, MA) or the Lunar (GE Healthcare, Madison, WI) scanner, specific to each participating center. To adjust for systematic differences in BMD by DXA, values were standardized to the Hologic BMD using published equations **T-score reference range used:** NHANES III race-specific male reference ranges Interval between risk assessment and BMD testing: NR
Rud et al, 2005[228] Danish Osteoporo-sis Prevention Study Denmark Fair	Peri- and postmenopausal women ages 45 to 58 years from the general population recruited for the Danish Osteoporosis Prevention Study	50.5 (NR)	1,997 (100)	Site of BMD measurement 1 Site: FN or TH or LS 92/2,009 (4.6%)	**DXA machine/software:** Hologic QDR 1000/W and QDR 2000 **T-score reference range used:** T-scores for the FN and TH calculated using NHANES III reference values Hologic references values were used for the LS. Authors do not specify if age-matched reference group was used or young White women. Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Shepherd et al, 2010[202] NHANES U.S. Fair	Men age 50 years or older with DXA scan in any of the NHANES 1999 to 2000, 2001 to 2002, and 2003 to 2004 datasets; 81% White, 8.2% African American, 3.6% Mexican American, 7.2% Other	63 (NR)	0 (0)	Site of BMD measurement 1 Site: FN or TH or LS 303/2,944 (10.3%) Site of BMD measurement 2 Site: LS 126/2,944 (4.3%)	DXA machine/software: Hologic QDR-4500A T-score reference range used: White men ages 20–29 years **Interval between risk assessment and BMD testing:** NR but likely reasonably concurrent since NHANES enrolls persons prospectively
Shuler et al, 2016[241] U.S. Fair	Patients living in rural areas identified from electronic health record at a single academic health center; women age 65 years or older, men age 70 years or older, or patients age 50 years or older with prior fracture, steroid or Lupron use	65.8 (NR)	39 (87)	Site of BMD measurement 1 NR 23/45 (51.1%)	DXA machine/software: NR T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Sinnott et al, 2006[219] U.S. Fair	African American men, age 35 years or older from outpatient general medicine VA clinics at a single site	63.8 (14.8)	0 (0)	Site of BMD measurement 1 FN or TH or LS 9/128 (7%)	DXA machine/software: GE Lunar machine **T-score reference range used:** Manufacturer's reference values, namely a young Caucasian male database for the hip and a Caucasian female database for the spine Interval between risk assessment and BMD testing: NR
Toh et al, 2019[245] Malaysia Fair	Postmenopausal women age 50 years or older who were randomly selected during office visits at a hospital-based primary care clinic without h/o osteoporosis or risk factors; ethnicity Malay 8.0%, Chinese 72.0%, Indian 18.7%, and Eurasian 1.3%	62.0 (7.0)	150 (100)	Site of BMD measurement 1 FN or LS 16/150 (10.7%) Site of BMD measurement 2 Site: FN 6/150 (4%)	DXA machine/software: IDXA, GE Lunar T-score reference range used: NR Interval between risk assessment and BMD testing: NR

Appendix D Table 4. Characteristics of Included Studies for Diagnostic Accuracy of Risk Assessment Instruments (Key Question 2c)

Author, Year Study Name Country ROB	Population	Mean Age (SD)	N (%) Female	Prevalence of Osteoporosis	Reference Test Details
Williams et al, 2017[237] U.S. Fair	Men age 70 years or older assigned to the VA Salt Lake City bone health team between February 1, 2012, and February 13, 2012; majority Caucasian (94.2%)	80.4 (5.8)	0 (0)	Site of BMD measurement 1 FN or TH or LS 112/463 (24.2%) Site of BMD measurement 2 Site: FN 95/463 (20.5%) Site of BMD measurement 3 Site: LS 24/463 (5.2%) Site of BMD measurement 4 Site: TH 36/463 (7.8%)	DXA machine/software: GE Lunar iDXA T-score reference range used: NR Interval between risk assessment and BMD testing: NR
Zimering et al, 2007[216] U.S. Fair	Ambulatory men age 40 years or older attending general medicine clinics, endocrinology clinics, or osteoporosis clinics at veterans health centers; 94% Caucasian, 5% African American, 1% Other in the validation cohort. A separate cohort of 134 African American men representing a convenience sample recruited at the same time as the development and validation cohorts.	Validation cohort: 68.2 (10.2) African American cohort: 60.9 (13)	0 (0)	Site of BMD measurement 1 FN 22/197 (11.2%)	DXA machine/software: Hologic QDR 4500 SL **T-score reference range used:** NHANES III young male, ethnicity/race-specific reference data Interval between risk assessment and BMD testing: NR

Abbreviations: BMD=bone mineral density; CaMos=Canadian Multicentre Osteoporosis Study; DXA=dual-energy X-ray absorptiometry; FN=femoral neck; GP=general practitioner; KNHANES=Korean National Health and Nutrition Examination Survey; KQ=key question; LS=lumbar spine; MN=Minnesota; N=number; NHANES=National Health and Nutrition Examination Survey; NHW=non-Hispanic White; NR=not reported; OR=odds ratio; ROB=risk of bias; SD=standard deviation; TH=total hip; U.S.=United States; VA=Veterans Affairs.

Appendix D Table 5. Characteristics of Included Studies for Evidence on Repeat Screening (Key Question 2d)

Study	Study Cohort, Country	Inclusion/Exclusion Criteria	Mean Length of Followup, Years	N	Participant Characteristics
Berry et al, 2013[246]	Framingham Osteoporosis Study, U.S.	Included participants with at least two BMD measurements. Excluded those with hip fracture prior to second test. Mean time between BMD tests: 3.7 years	9.6 after repeat test	802	Mean age: 74.8 (SD 4.5) % women: 61
Crandall et al, 2020[249]	Women's Health Initiative U.S.	Included women with 2 BMD measurements 3 years apart. Excluded those reporting use of bisphosphonates, calcitonin, or selective estrogen receptor modulators and those reporting MOF at baseline or prior to 3-year BMD. Participants with missing data regarding hormone therapy, fracture history, or BMI, and those without followup visits after the 3-year BMD were also excluded.	9.0 after repeat test	7,419	Mean age: 66.1 (SD 7.2) % women: 100
Ensrud et al, 2022[250]	Osteoporotic Fractures in Men (MrOS), U.S.	Included men age 65 years or older who completed BMD measurements at baseline and year 7. Excluded men who were unable to walk without the assistance of others or who had history of bilateral hip replacement. Time between measurement:7 years	8.2 years after repeat test	3,651	Mean age: 72.3 (SD 5.1) at time of initial BMD % women: 0
Hillier et al, 2007[247]	Study of Osteoporotic Fractures, U.S.	Included participants with at least two BMD measurements. Excluded those with fracture prior to second test. Mean time between BMD tests: 8 years	11.4 total (5 years after repeat test)	4,124	Mean age: 74 (SD 4) % women: 100
Leslie et al, 2017[24]	Manitoba DXA Registry Canada	Included participants with at least two BMD measurements and no osteoporosis treatment. Mean time between BMD tests: 4.0 years	7.7 after repeat test	3,961	Mean age 60.4 (SD 9.6) % women: 100

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; HR=hazard ratio; MOF=major osteoporotic fracture; N=number; NR=not reported; SD=standard deviation; U.S.=United States.

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Adachi et al, 2009[296] RCT Fair KQ 5	**N=**438 Postmenopausal women, at least 6 months after last menses, at least age 40 years (or age 25 years if surgical menopause) with history of osteoporotic fracture or T-score less than -2.0; 89% White, 8% Hispanic, 3% Asian, 1% Black Mean (SD) age: 65.5 (NR) % Female: 100% **T-score inclusion criteria:** T-score <-2.0 (site unspecified) Mean T-score (site of BMD): NR **% Prior Fx:** 6.8% with hlstory of osteoporotic fracture	**Drug:** Alendronate 10 mg/day Comparator: Placebo Duration of intervention: 3 mo
Ascott-Evans et al, 2003[252] RCT Fair KQ 4, KQ 5	**N=**144 Postmenopausal women age younger than age 80 years, with 85% of enrollees younger than age 65 years; 91.7% White, 8.3% other Mean (SD) age: 57.3 (6.6) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <-1.5 and >-3.5 Mean T-score (site of BMD): -2.3 (SD 0.58); lumbar spine % Prior Fx: 0%	**Drug:** Alendronate 10 mg/day Comparator: Placebo Duration of intervention: 1 y
Bell et al, 2002[276] RCT Fair KQ 5	**N=**65 Postmenopausal women, ages 45 to 88 years with T-score range -1.75 or less at LS and no more than 1 prior vertebral fracture; 100% African American **Mean (SD) age**: Alendronate: 66.4 (1.5), Placebo: 65.9 (1.6) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score -1.75 or less Mean T-score (site of BMD): Alendronate: -3.09, Placebo: -3.18; lumbar spine Alendronate: -1.71, Placebo: -1.82; total hip % Prior Fx: NR	**Drug:** Alendronate 10 mg/day Comparator: Placebo Duration of intervention: 2 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Bone et al, 1997[277] RCT Fair KQ 4, KQ 5	N=359 Women ages 60 to 85 years in generally good health with no more than 1 lumbar vertebral fracture; 97% White **Mean (SD) age**: Alendronate, 1 mg: 71.1 (5.8), Alendronate, 2.5 mg: 70.0 (5.8), Alendronate, 5 mg: 70.8 (5.5) Placebo: 71.1 (6.4) % Female: 100% **T-score inclusion criteria:** BMD 0.824 g/cm^2 or less (Hologic) or 0.944 g/cm^2 or less (Lunar); both of which correspond to T score -2.0 Mean BMD (site of BMD): Alendronate, 1 mg Hologic: 0.70 g/cm^2 (0.08); lumbar spine Lunar: 0.79 g/cm^2 (0.12); lumbar spine Alendronate, 2.5 mg Hologic: 0.72 g/cm^2 (0.08); lumbar spine Lunar: 0.81 g/cm^2 (0.11); lumbar spine Alendronate, 5 mg Hologic: 0.73 g/cm^2 (0.07); lumbar spine Lunar: 0.80 g/cm^2 (0.12); lumbar spine Placebo Hologic: 0.71 g/cm^2 (0.09); lumbar spine Lunar: 0.84 g/cm^2 (0.06); LS **% Prior Fx:** Alendronate, 1 mg: 41.9%, Alendronate, 2.5 mg: 36.0%, Alendronate, 5 mg: 37.6%, Placebo: 34.1%	**Drug:** Alendronate, 1 mg/day, Alendronate, 2.5 mg/day, Alendronate, 5 mg/day Comparator: Placebo Duration of intervention: 2 y
Bone et al, 2008[285] RCT Fair KQ 4, KQ 5	**N=**332 Postmenopausal women Mean (SD) age: 59.4 (7.5) % Female: 100% **T-score inclusion criteria:** lumbar spine or total hip T-score between -1 and <-2.5 Mean T-score (site of BMD): -1.61 (SD 0.42); lumbar spine **% Prior Fx:** No fractures since age 25	**Drug:** Denosumab 60 mg every 6 months at baseline, 6, 12, and 18 months Comparator: Placebo Duration of intervention: 3 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Boonen et al, 2012[251] RCT Good KQ 4, KQ 5	**N=**1,199 Men age 50 to 85; 94% White, 0.01% Black, 0.001% Asian, 0.05% other **Mean (SD) age**: Median age 66 % Female: 0% **T-score inclusion criteria:** Total hip or femoral neck T-score ≤-1.5 **Mean T-score (site of BMD):** Zoledronic acid -2.23 (0.677); femoral neck Placebo: -2.24 (0.685); femoral neck Zoledronic acid: -1.70 (0.764); total hip Placebo: -1.72 (0.808); total hip **% Prior Fx:** Zoledronic acid: 31.1% Placebo: 33.1%	**Drug:** Zoledronic acid 5-mg IV at baseline and 1 y Comparator: Placebo Duration of intervention: 2 y
Chapurlat et al, 2013[288] RCT Fair KQ 5	**N=**148 Women who were at least 1 year postmenopausal **Mean (SD) age**: Ibandronate: 62.7 (5.0) Placebo: 62.7 (5.3) % Female: 100% **T-score inclusion criteria:** Spine or hip T-score <-1.0 and >-2.5 Mean T-score (site of BMD): -1.4 (NR); site unspecified % Prior Fx: NR	**Drug:** Ibandronate 150 mg/month Comparator: Placebo Duration of intervention: 2 y
Chesnut et al, 1995[253] RCT Fair KQ 4, KQ 5	**N=**188 Women more than 5 years postmenopausal and ages 42 to 75 years; 97.9% White, 2.1% Asian **Mean (SD) age**: Alendronate 10 mg: 62.9 (6.1) Placebo: 63.6 (7.1) % Female: 100% T-score inclusion criteria: NR Mean T-score (site of BMD): -1.1 (NR); hip % Prior Fx: 0%	**Drug:** Alendronate 5 mg/day alendronate 10 mg/day alendronate 40 mg/day (for 3 months then 2.5 mg/day for 21 months) alendronate 20 mg/day (for 1 y then placebo for 1 y) alendronate 40 mg/day (for 1 y then placebo for 1 y) Comparator: Placebo Duration of intervention: 2 y
Cryer et al, 2005[297] RCT Fair KQ 5	**N=**454 Postmenopausal women at least 6 months after last menses; 91% White, 2% Black, 1% Asian, 5% Hispanic, 1% Native American, 1% other Mean (SD) age: 65 (10) % Female: 100% **T-score inclusion criteria:** T-score <-2.0 below young mean bone mass at one of any of the following sites: total hip, hip trochanter, femoral neck, total spine >-3.5 SD below young mean bone mass at any site **Mean T-score (site of BMD):** -2.52 (NR) to -2.46 (NR); lumbar spine % Prior Fx: NR	**Drug:** Alendronate 70 mg weekly Comparator: Placebo Duration of intervention: 6 mo

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year **Trial Name** **Study Design** **Study Quality** **KQ**	**Participant Characteristics**	**Intervention Groups and Duration**
Cummings et al, 2009[28] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[28] Palacios et al, 2015[422] FREEDOM (Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months) Trial RCT Fair KQ 4, KQ 5	**N=**7,808 Women ages 60 to 90 years **Mean (SD) age**: Denosumab: 72.3 (5.2) Placebo: 72.3 (5.2) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <-2.5 but >-4.0 Total hip T-score <-2.5 but >-4.0 Mean T-score (site of BMD): -2.8 (0.70); lumbar spine -1.9 (0.81); total hip -2.2 (0.72); femoral neck **% Prior Fx:** Denosumab: 51% with prior fracture Placebo: 50% with prior fracture	**Drug:** Denosumab 60 mg/6 months subcutaneously Comparator: Placebo Duration of intervention: 3 y
Cummings et al, 1998[25] Bauer et al, 2000[289] Cummings et al, 2007[29] Quandt et al, 2005[255] Fracture Intervention Trial (FIT) RCT Good KQ 4, KQ 5	**N=**4,432 Women 2 years or more postmenopausal and ages 55 to 80 years; 97% White **Mean (SD) age**: Alendronate: 67.6 (6.2); placebo: 67.7 (6.1) % Female: 100% **T-score inclusion criteria:** Femoral neck T-score: less than -1.6 (approximate T-score of <-2.0) Mean T-score (site of BMD): NR **% Prior Fx:** Population used for this update included 0% with prior vertebral fracture The FIT comprised two arms: FIT 1 had existing vertebral fractures, and FIT II did not. Analysis of FIT 1 alone was ineligible for this review. Cummings (1998) presented outcomes for FIT II alone (N=4,432). Cummings (2007) and Bauer (2000) presented results for all FIT participants. Quandt (2005) looked at the subgroup of all FIT participants with osteopenia.	**Drug:** Alendronate 5 mg/day for 2 y then 10 mg/day for 1 y for those without existing vertebral fractures, and 2 to 2.6 y for those with vertebral fractures at baseline Comparator: Placebo Duration of intervention: 3 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Devogelaer et al, 1996[302] RCT Fair KQ 5	**N=**516 Women at least 5 years postmenopausal and ages 45 to 80 years; race/ethnicity NR Mean (SD) age: Alendronate 5 mg: 61.2 (6.8) Alendronate 10 mg: 63.2 (6.6) Alendronate 20 mg/5 mg: 63.0 (6.6) Placebo: 62.7 (7.2)**%** Female: 100% **T-score inclusion criteria:** Lumbar spine BMD T-score ≤-2.5 **Mean BMD):** Lumbar spine (g/cm^2; Lunar) Alendronate 5 mg: 0.80 (0.09) Alendronate 10 mg: 0.80 (0.08) Alendronate 20 mg/5 mg: 0.79 (0.11) Placebo: 0.80 (0.09) Lumbar spine (g/cm^2; Hologic) Alendronate 5 mg: 0.72 (0.08) Alendronate 10 mg: 0.70 (0.09) Alendronate 20 mg/5 mg: 0.72 (0.08) Placebo: 0.70 (0.08) **% Prior Fx:** NR; exclusion criteria were more than one lumbar vertebral fracture and/or any fracture of the proximal femur due to osteoporosis; could not assume that all participants with prior fracture were excluded.	Drug: Alendronate 5 mg/d Alendronate 10 mg/d Alendronate 20 mg/d for 2 y, then 5 mg/d for 1 y With 500 mg calcium carbonate qd **Comparator:** Placebo with 500 mg calcium carbonate qd **Duration of intervention:** 3 y
Eisman et al, 2004[298] RCT Good KQ 5	**N=**449 Postmenopausal women and men with osteoporosis (as determined by investigators); 65.7% White, 18% Asian, 12% Hispanic, 5% other Mean (SD) age: 63.6 (NR) % Female: 93%–96% T-score inclusion criteria: NR Mean T-score (site of BMD): NR % Prior Fx: NR	**Drug:** Alendronate 70 mg weekly Comparator: Placebo Duration of intervention: 3 mo
Greenspan et al, 2002[29] RCT Fair KQ 5	**N=**450 Postmenopausal women or men with osteoporosis; 96% White Mean (SD) age: 67 (NR) % Female: 92% T-score inclusion criteria: NR Mean T-score (site of BMD): NR % Prior Fx: NR	**Drug:** Alendronate 70 mg weekly Comparator: Placebo Duration of intervention: 3 mo

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year **Trial Name** **Study Design** **Study Quality** **KQ**	**Participant Characteristics**	**Intervention Groups and Duration**
Greenspan et al, 2003[30] RCT Good KQ 5	**N=**186 Women ages 65 to 90 years Mean (SD) age: 71.5 (NR) % Female: 100% T-score inclusion criteria: NR Mean T-score (site of BMD): -1.7 (NR); femoral neck % Prior Fx: 0%	**Drug:** Alendronate 10 mg/day Study included 2 other arms not relevant to this update: 1) conjugated equine estrogen (CEE) 0.625 mg/day with or without medroxyprogesterone 2.5 mg daily based on uterus presence and 2) alendronate + CEE Comparator: Placebo Duration of intervention: 3 y
Grey et al, 2010[262] Grey et al. 2009[263] RCT Fair KQ 5	**N=**50 Women 5 years or more postmenopausal; age range not specified **Mean (SD) age**: Zoledronate: 62 (8) Placebo: 65 (8) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <-1 and >-2 Total hip T-score <-1 and >-2 **Mean T-score (site of BMD):** Zoledronic acid: -1.0 (0.7); lumbar spine Placebo: -1.3 (0.7); lumbar spine Zoledronic acid: -1.3 (0.6); total hip Placebo: -1.2 (0.5); total hip **% Prior Fx:** 42% with prior fracture, 28% in zoledronate arm and 56% in placebo arm	**Drug:** Zoledronic acid 5-mg IV (onetime dose) Comparator: Placebo Duration of intervention: 3 y
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274] RCT Fair KQ 4, KQ 5	**N=**180 Women more than 5 years postmenopausal with osteopenia **Mean (SD) age**: Zolenronate 1 mg: 64 (8); zoledronate 2.5 mg: 66 (9); zoledronate 5 mg: 66 (8); placebo: 65 (9) % Female: 100% **T-score inclusion criteria:** T-score at either lumbar spine or total hip between -1 and -2.5 **Mean T-score (site of BMD):** Zoledronic acid 1 mg: -1.4 (0.7); zoledronic acid 2.5 mg: -1/2 (0.9); zoledronic acid 5 mg: -1.1 (1), placebo: -1.3 (0.8); lumbar spine Zoledronic acid 1 mg: -1.2 (0.7); zoledronic acid 2.5 mg: -1.3 (0.5); zoledronic acid 5 mg: -1.3 (0.7), placebo: 1.1 **% Prior Fx:** Zoledronate 1 mg: 16%, zoledronate 2.5 mg: 21%. zoledronate 5 mg: 14%, placebo: 19% with prior fracture during adulthood	**Drug:** Zoledronic acid 1 mg, 2.5 mg, and 5 mg (single-dose IV) Comparator: Placebo Duration of intervention: NA, single dose

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Hosking et al, 2003[267] RCT Fair KQ 4, KQ 5	**N=**549 Women ages 60 to 90 years at least 3 years postmenopausal with osteoporosis; 99.5% Caucasian Mean (SD) age: 69 (NR) % Female: 100% **T-score inclusion criteria:** Lumbar spine or total hip T-score <-2.5 or both <-2.0 Mean T-score (site of BMD): In g/cm^2 Placebo: 0.72 (0.10); total hip 0.73 (0.07); lumbar spine Risedronate: 0.69 (0.08); total hip 0.72 (0.08); lumbar spine Alendronate: 0.70 (0.10); total hip 0.71 (0.08); lumbar spine **% Prior Fx:** 48.5% with history of fracture	**Drug:** Alendronate 70 mg weekly Risedronate 5 mg daily Comparator: Placebo Duration of intervention: Risedronate: 3 mo Alendronate: 1 y
Johnell et al, 2002[294] RCT Fair KQ 5	**N=**331 (165 if limited to only the alendronate and placebo groups; study also included a raloxifene and a raloxifene + alendronate groups) Postmenopausal women younger than age 75 years and more than 2 years since their last menstrual period 95% White Mean (SD) age: 63.6 (NR) % Female: 100% T-score inclusion criteria: Femoral neck T-score <-2.0 Mean T-score (site of BMD): In g/cm^2 Alendronate: 0.62 (0.08); femoral neck Placebo: 0.62 (0.09); femoral neck % Prior Fx: NR	**Drug:** Alendronate, 10 mg/day Comparator: Placebo Duration of intervention: 1 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Koh et al, 2016[286] RCT Fair KQ 4, KQ 5	**N=**135 Postmenopausal women ages 60 to 90 years with osteoporosis **Mean (SD) age**: Denosumab: 67.0 (4.86) Placebo: 66.0 (4.77) % Female: 100 **T-score inclusion criteria:** Total hip or lumbar spine T-score <-2.5 and ≥-4.0 **Mean T-score (site of BMD):** Denosumab: -2.5 (0.56); femoral neck Placebo: -2.4 (0.61); femoral neck Denosumab: -2.0 (0.64); total hip Placebo: -1.9 (0.65); total hip Denosumab: -3.0 (0.59); total spine Placebo: -2.9 (0.58); total spine Denosumab:-2.2 (0.63); trochanter Placebo: -2.2 **% Prior Fx:** Denosumab: 30% with previous fracture Placebo: 23% with prior fracture	**Drug:** Denosumab 60 mg (single-dose IV) Comparator: Placebo Duration of intervention: 6 mo
Lewiecki et al, 2007[284] McClung et al, 2006[304] RCT Fair KQ 4, KQ 5	**N=**365 Postmenopausal women up to age 80 years; of the entire study population, 86.2% were White, 9.5% were Hispanic, 2.9% were Black, and 1.5% were other Mean (SD) age: 62.5 (8.1) % Female: 100% **T-score inclusion criteria:** lumbar spine T-scores of -1.8 to -4.0 or femoral neck/total hip T-scores of -1.8 to -3.5 **Mean T-score (site of BMD):** Denosumab (all doses): mean ranged from -2.0 (NR) to -2.3 (NR); lumbar spine Placebo: -2.2 (NR); lumbar spine % Prior Fx: 0% Study also included an open-label alendronate arm (N=70) that was not used in our synthesis because the comparison between alendronate and placebo would be considered high ROB, and this review is not concerned with comparative effectiveness between alendronate and denosumab.	**Drug:** Denosumab 6 mg, 14 mg, or 30 mg every 3 months or denosumab 14 mg, 60 mg, 100 mg, or 210 mg every 6 months, alternating with placebo to maintain blinding Comparator: Placebo Duration of intervention: 2 y
Liberman et al, 1995[256] RCT Fair KQ 4, KQ 5	**N=**994 Women ages 45 to 80 years who were more than 5 years postmenopausal; 87.4% White, 0.4% Black, 12.2% other Mean (SD) age: 64 (NR) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <-2.5 **Mean T-score (site of BMD):** In g/cm^2: range 0.60 to 0.74 at femoral neck across the various densitometers **% Prior Fx:** 21% with prior vertebral fracture	**Drug:** Alendronate 5 or 10 mg/day for 3 years or 20 mg/day for 2 years followed by 5 mg/day for 1 year Comparator: Placebo Duration of intervention: 3 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
McClung et al, 2001[257] RCT Fair KQ 4, KQ 5	**N=**9,331 Women postmenopausal, age 70 years or older; 98% White Mean (SD) age: Overall NR N (%): Age 70–79: 5,445 (58.4); Age ≥80: 3,886 (41.6) % Female: 100% **T-score inclusion criteria:** Femoral neck T-score <-4 or <-3 with risk factor for hip fracture Mean T-score (site of BMD): -3.7 (0.6); femoral neck **% Prior Fx:** Women ages 70 to 79 years: 39% with prior vertebral fracture (1,703/4,351 with information available) Women age 80 years or older: 44% with prior vertebral fracture (1,137/2,566 with information available)	**Drug:** Risedronate 2.5 mg/d, risedronate 5 mg/d Comparator: Placebo **Duration of intervention:** Planned therapy: 3 y (mean therapy: 2 y)
McClung et al, 2009[287] RCT Fair KQ 5	**N=**581 Postmenopausal women age 45 years or older **Mean (SD) age**: 59.6 to 60.5 % Female: 100% **T-score inclusion criteria:** T-score between -1.0 and -2.5 at the lumbar spine and T-score greater than -2.5 at the femoral neck **Mean T-score (site of BMD):** -1.47 (NR) to -1.40 (NR); femoral neck **% Prior Fx:** Persons with previous Grade 2 or 3 vertebral fractures were excluded.	**Drug:** Zoledronic acid 5-mg IV, at baseline and at 1 y Zoledronic acid 5-mg IV, at baseline only Comparator: Placebo Duration of intervention: 2 y
McClung et al, 2009[275] RCT Fair KQ 4, KQ 5	**N=**160 Women postmenopause and ages 45 to 60; race/ethnicity NR Mean (SD) age: Ibandronate: 53.7 (3.6) Placebo: 53.4 (3.8) % Female: 100% **T-score inclusion criteria:** Lumbar spine BMD T-score between -1.0 and -2.5 and baseline BMD T-score >-2.5 at the total hip, trochanter, and femoral neck Mean T-score (site of BMD): Lumbar spine BMD T-score Ibandronate: -1.6 (0.4) Placebo: -1.6 (0.4) **% Prior Fx:** 0% (excluded from enrollment)	**Drug:** Ibandronate 150 mg monthly; daily vitamin D (400 IU and calcium (500 mg) supplements **Comparator:** Placebo; daily vitamin D (400 IU) and calcium (500 mg) supplements Duration of intervention: 1 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
McClung et al, 2004[291] RCT Fair KQ 5	**N=**653 Women more than 1 year postmenopausal; age range unspecified **Mean (SD) age**: Ibandronate 0.5 mg: 58.8 (8.9) Ibandronate 1.0 mg: 57.6 (8.0) Ibandronate 2.5 mg: 58.2 (8.6) Placebo: 57.9 (8.6) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <-1.0 and >-2.5 **Mean T-score (site of BMD):** Ibandronate 0.5 mg: -1.0 (1.1); lumbar spine Ibandronate 1.0 mg: -1.0 (1.0); lumbar spine Ibandronate 2.5 mg: -1.1 (0.9); lumbar spine Placebo: -1.0 (1.2); lumbar spine % Prior Fx: 0%	**Drug:** Ibandronate 0.5 mg/d Ibandronate 1.0 mg/d Ibandronate 2.5 mg/d Comparator: Placebo Duration of intervention: 2 y
Mortensen et al, 1998[25] RCT Fair KQ 4, KQ 5	**N=**111 Women 6–60 months postmenopausal and ages 40 to 61 years; 100% White **Mean (SD) age**: Risedronate 5 mg cyclic: 51.3 (3.4) Risedronate 5 mg daily: 52.1 (3.9) Placebo: 51.2 (4.2) % Female: 100% **T-score inclusion criteria:** Women with normal lumbar spine bone mass (within 2 SD of age-matched mean bone mass) (i.e., Z-score >-2.0) Mean T-score (site of BMD): -1.1 (NR); lumbar spine % Prior Fx: 0%	**Drug:** Risedronate 5 mg cyclic (daily for first 2 weeks of every month, then placebo daily for the rest of the month) Risedronate 5 mg/d Comparator: Placebo Duration of intervention: 2 y
Nakamura et al, 2012[279] RCT Fair KQ 4, KQ 5	**N=**226 Ambulatory Japanese postmenopausal women age 80 years or younger who had osteoporosis and a BMD T-score of -2.5 to -4.0 at the lumbar 1 to lumbar 4 spine or -2.5 to -3.5 at either the femoral neck or total hip Mean (SD) age: 65.1 (6.8) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score of -2.5 to -4.0 Femoral neck or total hip T-score of -2.5 to -3.5 Mean T-score (site of BMD): -3.08 (0.41); lumbar spine -1.85 (0.69); total hip **% Prior Fx:** 34% with prior vertebral fracture	**Drug:** Denosumab 14 mg subcutaneously every 6 months for 12 months Denosumab 60 mg subcutaneously every 6 months for 12 months Denosumab 100 mg subcutaneously every 6 months for 12 months or placebo every 6 months for 12 months Comparator: Placebo Duration of intervention: 1 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Orwoll et al, 2012[278] ADAMO RCT Fair KQ 4, KQ 5	**N=**242 Men ages 30 to 85 years; 94.2% White Mean (SD) age: 65.0 (9.8) % Female: 0% **T-score inclusion criteria:** LS or FN BMD T-score between -2.0 and -3.5 OR LS or FN BMD T-score between -1.0 and -3.5 with prior major osteoporotic fracture; all T-scores based on male reference range **Mean T-score (site of BMD):** All based on male reference range: -2.0 (1.1); lumbar spine -1.4 (0.6); total hip -1.9 (0.6); femoral neck **% Prior Fx:** 39.3% with any prior fracture 24.8% with prior osteoporotic fracture 14.9% with prior major osteoporotic fracture 22.7% with prevalent vertebral fracture	**Drug:** Denosumab 60 mg subcutaneously every 6 months Comparator: Placebo **Duration of intervention:** 1 year (blinded), 2nd year (open-label)
Pols et al, 1999[259] Fosamax International Trial (FOSIT) RCT Fair KQ 4, KQ 5	**N=**1,908 Women 3 years or more postmenopausal and ages 39 to 84 years; 94% White **Mean (SD) age:** Alendronate: 62.8 (7.5) Placebo: 62.8 (7.4) % Female: 100% T-score inclusion criteria: NR Mean T-score (site of BMD): -2.2 (NR); site unspecified % Prior Fx: NR	**Drug:** Alendronate 10 mg/day Comparator: Placebo Duration of intervention: 1 y
Ravn et al, 1996[264] RCT Fair KQ 4, KQ 5	**N=**180 Women more than 10 years past menopause and younger than age 75 years; 100% White (Denmark) Mean (SD) age: 65 (NR) % Female: 100% T-score inclusion criteria: NR Mean T-score (site of BMD): -1.72 (NR); lumbar spine -1.5 (NR); proximal femur % Prior Fx: 0%	**Drug:** Ibandronate 0.25 mg/day; 0.50 mg/day; 1.0 mg/day; 2.5 mg/day; 5.0 mg/day Comparator: Placebo Duration of intervention: 1 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Reginster et al, 2005[266] Monthly Oral Pilot Study (MOPS) RCT Fair KQ 4, KQ 5	**N=**144 Women more than 3 years postmenopausal and ages 55 to 80 years **Mean (SD) age**: Ibandronate 50 mg: 65.7 (61 to 74) Ibandronate 50/100 mg: 61.7 (55 to 77) Ibandronate 100 mg: 64.1 (56 to 77) Ibandronate 150 mg: 63.3 (55 to 79) Placebo: 63.9 (55 to 79) % Female: 100% T-score inclusion criteria: No specific BMD criteria **Mean T-score (site of BMD):** Ibandronate 50 mg: -1.9 (NR); lumbar spine Ibandronate 50/100 mg: -0.3 (NR); lumbar spine Ibandronate 100 mg: -1.1 (NR); lumbar spine Ibandronate 150 mg: -0.8 (NR); lumbar spine % Prior Fx: NR	**Drug:** Ibandronate 50 mg/per month; Ibandronate 50 mg for the first month/100 mg for months 2-3; Ibandronate 100 mg/per month; Ibandronate 150 mg/per month Comparator: Placebo Duration of intervention: 3 mo
Reid et al, 2002[260] RCT Fair KQ 4, KQ 5	**N=**351 Women 5 years or more postmenopausal and ages 45 to 80 years; 95% White **Mean (SD) age**: Zoledronic acid 0.25-mg IV: 64 (6) 0.5-mg IV: 64 (7) 1-mg IV: 65 (7) 2-mg IV: 63 (7) 4-mg IV: 65 (7) Placebo: 64 (6) % Female: 100% T-score inclusion criteria: Lumbar spine T-score <-2.0 Mean T-score (site of BMD): -2.9 (NR); lumbar spine % Prior Fx: 0%	**Drug:** Zoledronic acid IV 0.25 mg/3 m 0.5 mg/3 m 1 mg/3 m 4 mg/1 y 2 mg/6 m Comparator: Placebo Duration of intervention: 1 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] RCT Good KQ 4, KQ 5	**N=**2,000 Postmenopausal women with osteopenia age 65 years or older; 95% European, 0.02% Maori, 0.01% Pacific Islander, 0.02% East Asian, 0.005% Indian, 0.002% other **Mean (SD) age**: 5 mg zoledronic acid: 71 (5.1) Placebo: 71 (5.0) % Female: 100% **T-score inclusion criteria:** Total hip or femoral neck T-score of -1.0 to -2.5 on either side Mean T-score (site of BMD): 5 mg zoledronic acid Lumbar spine: -0.91 (1.12) Total hip: -1.27 (0.59) Femoral neck: -1.64 (0.47) Total body: -0.81 (0.86) Placebo Lumbar spine: -0.87 (1.16) Total hip: -1.24 (0.60) Femoral neck: -1.63 (0.47) Total body: -0.80 (0.90) **% Prior Fx:** 5 mg zoledronic acid 23.7% with prior fracture after age 45 years 13.7% with prevalent vertebral fracture at baseline Placebo 23.8% with prior vertebral fracture after age 45 years 12.6% with prevalent vertebral fracture at baseline	**Drug:** Zoledronic acid 5-mg IV every 18 months Comparator: Placebo Duration of intervention: 6 y
Riis et al, 2001[265] RCT Fair KQ 4, KQ 5	**N=**240 Women more than 5 years menopausal and ages 55 to 76 years **Mean (SD) age**: Ibandronate 2.5 mg: 66.8 (4.9) Ibandronate 20 mg: 67.0 (5.0) Placebo: 66.3 (4.8) % Female: 100% T-score inclusion criteria: Spine T-score <-2.5 Femoral neck T-score <-2.5 **Mean T-score (site of BMD):** Ibandronate 2.5 mg: -3.206 (0.485); spine Ibandronate 2.5 mg: -2.941 (0.487); femoral neck Ibandronate 20 mg: -3.232 (0.573); spine Ibandronate 20 mg: -3.083 (0.425); femoral neck Placebo: -3.264 (0.579); spine Placebo: -2.987 (0.630); femoral neck % Prior Fx: NR	**Drug:** Ibandronate 2.5 mg/d Ibandronate 20 mg every other day for the first 24 days out of every 3 months, followed by a 9-week period without active drug (intermittent cyclical therapy) Comparator: Placebo Duration of intervention: 2 y

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Shiraki et al, 2003[295] RCT Fair KQ 5	**N=**211 Women and men ages 40 to 75 years with senile and postmenopausal osteoporosis; 100% Japanese (implied) Mean (SD) age: 60.3 (NR) % Female: 99% **T-score inclusion criteria:** Lumbar spine T-score <-2.5 without vertebral fracture; <-1.5 with vertebral fracture Mean T-score (site of BMD): -2.9 (NR); lumbar **% Prior Fx:** Mean number of prevalent vertebral fractures 0.3 (SD 0.8)	**Drug:** Risedronate 1 mg, 2.5 mg, 5 mg/day Comparator: Placebo Duration of intervention: 8 mo
Tanko et al, 2003[292] RCT Fair KQ 5	**N=**630 Women 1 to 10 years postmenopausal Mean (SD) age: 55 (NR) % Female: 100% T-score inclusion criteria: T-score ≥-2.5 Mean T-score (site of BMD): In g/cm^2 Ibandronate 5 mg: 1.00 (0.13); lumbar spine Ibandronate 10 mg: 0.98 (0.11); lumbar spine Ibandronate 20 mg: 0.99 (0.12); lumbar spine % Prior Fx: 0%	**Drug:** Ibandronate 5 mg weekly 10 mg weekly 20 mg weekly Comparator: placebo Duration of intervention: 2 y
Thiebaud et al, 1997[293] RCT Fair KQ 5	**N=**126 Women at least 5 years postmenopausal Mean (SD) age: 64 (NR) % Female: 100% T-score inclusion criteria: T-score <-2.5 Mean T-score (site of BMD): 0.71 (NR); lumbar spine % Prior Fx: 0%	**Drug:** Ibandronate 0.25, 0.5 mg, 1.0, or 2.0 mg/3 months 1 g calcium/day Comparator: Placebo Duration of intervention: 1 y
Tucci et al, 1996[268] RCT Fair KQ 4, KQ 5	**N=**478 Women ages 42 to 82 years, postmenopausal for at least 5 years with osteoporosis; 91% White, 8% Asian Mean (SD) age: 64 (NR) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score <2.5 SD below mean BMD of young White women Mean T-score (site of BMD): NR % Prior Fx: NR	**Drug:** Alendronate 5 mg/day Alendronate 10 mg/day Alendronate 20 mg/day for 2 years followed by 5 mg/day Comparator: Placebo Duration of intervention: 3y (5-10 mg) OR 2y (20 mg) + 1y (5 mg day)

Appendix D Table 6. Characteristics of Included Randomized, Controlled Trials for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design Study Quality KQ	Participant Characteristics	Intervention Groups and Duration
Valimaki et al, 2007[261] RCT Fair KQ 5	**N=**170 Women 5 years or more postmenopausal, age range unspecified; 100% White Mean (SD) age: 65.9 (6.8) % Female: 100% **T-score inclusion criteria:** Lumbar spine T-score >-2.5 and <-1 Proximal femur T-score ≤-1 Mean T-score (site of BMD): -1.82 (0.42); lumbar spine -1.23 (0.58); proximal femur % Prior Fx: NR	**Drug:** Risedronate 5 mg/d Comparator: Placebo Duration of intervention: 2 y

Abbreviations: BMD=bone mineral density; CEE=conjugated equine estrogen; FIT=Fracture Intervention Trial; FREEDOM=Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months; Fx=fracture; IV=intravenous; m=month; N=number; NA=not applicable; NR=not reported; OR=odds ratio; RCT=randomized, controlled trial; ROB=risk of bias; SD=standard deviation; y=year.

Appendix D Table 7. Study Characteristics of Included Cohort Studies for Harms of Treatment (Key Question 5)

First Author Year Cohort Title Study Quality	Sample Size and Study Population	Exposure and/or Intervention Comparator Duration
Lee, 2019[307] Korean National Health Insurance Data Fair	**Cohort Size:** 697,126 (analytic cohort) **Population:** New users (women and men) of oral or IV BP for osteoporosis, age 50 years or older, without previous hip fracture, cancer, or metabolic bone disorders. Age: BP users: 69.0 (8.8) Nonusers: 69.0 (8.8) N % Female: BP users: 316,472 (90.9) Nonusers: 316,671 (90.9) Race/Ethnicity NR	Exposed Group: BP users Oral or IV BP, switching within the drug class was allowed Comparator Group: Nonusers Non-BP users Duration Mean duration of BP use in exposed group: 1.02 ± 1.25 years
Pazianas, 2012[305] Danish National Prescription Database and Cause of Death Registry Fair	Cohort Size: 153,030 **Population:** Women age 50 years or older in Denmark with no prior cancer hospitalizations and receiving first prescription of alendronate (or no prescription) between 1996 and 2005 Age: Alendronate users: 71.9 (10.0) Nonusers: 71.9 (10.0) N % Female: 100 Race/Ethnicity NR	Exposed Group: Alendronate Oral alendronate, 67% used weekly dose Comparator Group: Nonusers No alendronate use Duration Duration of use NR Duration of followup: 5 years

Appendix D Table 7. Study Characteristics of Included Cohort Studies for Harms of Treatment (Key Question 5)

First Author Year Cohort Title Study Quality	Sample Size and Study Population	Exposure and/or Intervention Comparator Duration
Rubin, 2020[306] Swedish and Danish National Health Registries Fair	**Cohort Size:** 34,655 for full cohort; N for treatment-naïve cohort NR **Population:** Treatment-naïve zoledronic acid users (not receiving zoledronic acid as part of an oncology regimen) or nonusers living in Denmark or Sweden for 12 months prior to cohort entry during 2007–2012. Nonusers were identified through propensity-score matching. Age: Zoledronic acid users, median (LQ, UQ): 71.9 (64.3, 79.1) Nonusers, median (LQ, UQ): 72.0 (64.5, 79.2) N % Female: Zoledronic acid users, n (%): 7,476 (85.6) Nonusers, n (%): 22,243 (85.8) Race/Ethnicity NR	Exposed Group: Treatment-naïve zoledronic acid users New users of zoledronic acid identified based on prescription claims in national registries. Comparator Group: Treatment-naïve cohort with no osteoporosis treatment No prescription claims for any osteoporosis treatments Duration 3 to 7 years, mean followup time 800 days in Swedish sample and 1,000 days in Danish sample

Abbreviations: BP=bisphosphonate; IV=intravenous; LQ=lower quartile; N/n=number; NR=not reported; ROB=risk of bias; RR=risk ratio; UQ=upper quartile.

Appendix D Table 8. Outcomes from Included Trials for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Outcomes	Harms Outcomes
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS NTR2430 RCT Fair	**Hip fracture** (prespecified secondary endpoint) Screening: After mean followup of 3.7 years 133/5,516 (2.4%) 0.7 cases/100 person-years No screening: 143/5,405 (2.6%) 0.7 cases/100 person-years Adjusted HR: 0.91 (95% CI, 0.71 to 1.15) Other fractures After mean followup of 3.7 years All fractures (primary study endpoint) Screening: 626/5,516 (11.3%); 3.1 cases/100 person-years No screening: 632/5,405 (11.7%); 3.2 cases/100 person-years Adjusted HR: 0.97 (95% CI, 0.87 to 1.08) Osteoporotic fractures (all fractures except skull, finger, hand, toe, and foot) Screening: 547/5,516 (9.9%); 2.7 cases/100 person-years No screening: 578/5,405 (10.7%); 2.9 cases/100 person-years Adjusted HR: 0.91 (95% CI, 0.81 to 1.03) MOFs (hip, vertebral, wrist, humerus) Screening: 427/5,516 (7.7%) 2.1 cases/100 person-years No screening: 452/5,405 (8.3%); 2.3 cases/100 person-years Adjusted HR: 0.91 (95% CI, 0.80 to 1.04) All-cause mortality Screening: After mean followup of 3.7 years 499/5,516 (9.0%); 2.5 cases/100 person-years No screening: 479/5,405 (8.9%) 2.4 cases/100 person-years Adjusted HR: 1.03 (95% CI, 0.91 to 1.17) Subgroup analyses No interaction effects with age, history of fracture after age 50, or recent fracture for the primary outcome of all fractures (p=0.60, 0.48, and 0.34, respectively) Recent fracture association (<2 years before baseline) (screening n=493 and usual care n=473) MOF HR 0.65; 95% CI, 0.44 to 0.96 (screening n=43 vs. usual care n=60) Hip fractures HR 0.38; 95% CI, 0.18 to 0.79 (screening n=10 vs. usual care n=25)	NR

Appendix D Table 8. Outcomes from Included Trials for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Outcomes	Harms Outcomes
Rubin, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE NCT01388244 RCT Fair	**Hip fracture** (prespecified secondary endpoint) Screening: After median followup of 5.0 years (prespecified secondary endpoint, ITT analysis) 534/17,072 (3.1%) Per-protocol 1 analysis 169/9,279 (1.8%) No screening: ITT analysis: 532/17,157 (3.1%) Per-protocol 1 analysis: 202/9,326 (2.2%) Adjusted subhazard ratio (SHR) ITT analysis: 1.002 (95% CI, 0.889 to 1.130), p=0.972 Per-protocol 1 analysis: 0.82 (95% CI, 0.670 to 1.007), p=0.059 Other fractures After median followup of 5.0 years MOF (primary study endpoint, ITT analysis) Screening: 1,697/17,072 (9.9%) No screening: 1,719/17,157 (10.0%) aSHR: 0.986 (95% CI, 0.922 to 1.055), p=0.682 Per-protocol 1 analysis Screening: 725/9,279 (7.8%) No screening: 786/9,326 (8.4%) aSHR: 0.914 (95% CI, 0.827 to 1.011); p=0.082 All osteoporotic fractures (excluding fingers, toes, skull or face, prespecified secondary endpoint) Screening: 2,238/17,072 (13.1%) No screening: 2,233/17,157 (13.0%) aSHR: 1.004 (95% CI, 0.946 to 1.064), p=0.906 Per-protocol 1 analysis Screening: 996/9,279 (10.7%) No screening: 1,025/9,326 (11.0%) aSHR: 0.968 (95% CI, 0.887 to 1.056), p=0.465 All-cause mortality Screening: 1,968/17,072 (11.5%) No screening: 2,038/17,157 (11.9%) RR: 0.97 (95% CI, 0.92 to 1.03)	NR

Appendix D Table 8. Outcomes from Included Trials for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year **Trial Name** **Registry Number** **Study Design** **Study Quality**	**Outcomes**	**Harms Outcomes**
Rubin, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] (continued)	Subgroup analyses Per-protocol 2 analyses comparing DXA-scanned vs. control participants with FRAX MOF ≥15% Median followup of 5.0 years, aSHR (95% CI) Hip fx: 0.741 (0.553 to 0.909) MOF: 0.870 (0.769 to 0.985) All fx: 0.892 (0.801 to 0.993) Analyses stratified by age (65 to 69 years, 70 to 74 years, 75 years or older) showed no significant differences (authors did not specify whether this was ITT, per-protocol 1, or per-protocol 2 or all of them) In per-protocol-analyses controlling for differences in baseline characteristics such as BMI, smoking status, and prior fracture, showed no significant differences compared to the main analysis. In per-protocol analysis 2, when authors excluded hip fractures from the MOF outcome, the significant group differences for MOF became insignificant (unadjusted SHR=0.912 [95% CI, 0.794 to 1.047]) p=0.191 and adjusted SHR=0.924 (95% CI, 0.804 to 1.062) p=0.264	
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP ISRCTN 55814835 RCT Fair [120-123]	Hip fracture Screening: After 5 years followup (prespecified secondary endpoint): 164/6,233 (2.6%) No screening: 218/6,250 (3.5%) HR, 0.72 (95% CI, 0.59 to 0.89), p=0.002 Other fractures After 5 years followup All clinical fractures without regard to trauma excluding hands, feet, nose, skull, or cervical vertebrae (primary endpoint) Screening: 805/6,233 (12.9%) No screening: 852/62,50 (13.6%) HR, 0.94 (95% CI, 0.85 to 1.03), p=0.178 All clinical fractures at any site (prespecified secondary endpoint) Screening: 951/6,233 (15.3%) No screening: 1,002/6,250 (16.0%) HR: 0.94 (95% CI, 0.86 to 1.03), p=0.183 All-cause mortality Screening: After 5 years followup (prespecified secondary endpoint) 550/6,233 (8.8%) No screening: 525/6,250 (8.4%) RR: HR, 1.05 (95% CI, 0.93 to 1.19), p=0.436	Screening harms: After at least 5 years: Anxiety (State-Trait Anxiety Inventory-Short Form) Repeated measures analysis over 5 years, no difference between screening (both low-risk and high-risk groups) and no screening groups (p=0.515). Authors also reported the following "No serious adverse events related to screening were observed."

Appendix D Table 8. Outcomes from Included Trials for Direct Benefits and Harms of Screening (Key Questions 1 and 3)

Author, Year Trial Name Registry Number Study Design Study Quality	Outcomes	Harms Outcomes
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] (continued)	Subgroup analyses HR (95% CI) based on baseline 10-year hip fracture probability without BMD 10th percentile (FRAX hip 2.6%) Any fracture: 0.96 (0.86 to 1.08) Any fracture (selected sites excluded): 0.97 (0.85 to 1.09) Hip fracture: 0.93 (0.71 to 1.23) 25th percentile (FRAX hip 3.8%) Any fracture: 0.96 (0.86 to 1.07) Any fracture (selected sites excluded): 0.96 (0.86 to 1.08) Hip fracture: 0.91 (0.70 to 1.17) 50th percentile (FRAX hip 6.3%) Any fracture: 0.96 (0.87 to 1.05) Any fracture (selected sites excluded): 0.96 (0.86 to 1.06) Hip fracture: 0.85 (0.68 to 1.08) 75th percentile (FRAX hip 10.5%) Any fracture: 0.95 (0.87 to 1.04) Any fracture (selected sites excluded): 0.95 (0.86 to 1.04) Hip fracture: 0.77 (0.63 to 0.95) 90th percentile (FRAX hip 16.8%) Any fracture: 0.94 (0.84 to 1.05) Any fracture (selected sites excluded): 0.93 (0.83 to 1.05) Hip fracture: 0.67 (0.53 to 0.84) P for interaction with baseline FRAX hip risk (as a continuous measure) p>0.30 for any fracture p>0.30 for any fracture (selected sites excluded) p=0.021 for hip fracture	

Abbreviations: aSHR=adjusted subhazard ratio; BMD=bone mineral density; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; FRAX=Fracture Risk Assessment Tool; FRAX MOF=Fracture Risk Assessment Tool: Major Osteoporotic Fracture; Fx=fracture; HR=hazard ratio; ISRCT**N=**International Standard Randomised Controlled Trial Number; ITT=intention to treat; MOF=major osteoporotic fracture; NCT=National Clinical Trial; NR=not reported; NTR=Netherlands Trial Registry; RCT=randomized, controlled trial; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; RR=risk ratio; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study; vs.=versus.

Appendix D Table 9. Included Systematic Reviews for Direct Benefits and Harms of Screening (Key Question 1 and 3)

Author, Year Study Quality	Review Description	Outcomes
Auais et al, 2023[133] Good	**Search dates:** Through November 6, 2022 **Data sources:** Embase, MEDLINE, CINAHL, reference lists **Inclusion/exclusion criteria:** Studies assessing fracture risk assessment using a validated screening tool for preventing fractures or other patient health outcomes; both qualitative and quantitative studies were included. Articles from prior to 2000 were excluded. **Number of included studies:** 14 studies total; but only 3 RCTs pertaining to the primary question of impact of screening tools on fractures and health outcomes were included.	Quantitative synthesis not conducted. 1 RCT (SCOOP) observed no difference on all osteoporotic fractures but a significant decrease in hip fractures. 1 RCT (ROSE) observed no differences in the ITT analysis for MOF, but significant differences in the per protocol analysis for MOF, hip, and all fractures. 1 RCT (SALT-SOS) observed no differences for any fracture types.
Gates et al, 2023[131] Good	**Search dates:** Through April 4, 2022 **Data sources:** Embase, MEDLINE, Cochrane Library, trial registries, reference lists **Inclusion/exclusion criteria:** Varied by KQ. For the KQ concerning direct benefits and harms of screening, included RCTs or CCTs in community-dwelling adults age 40 years or older without diagnosis of osteoporosis or on treatment. Eligible interventions included fracture risk assessments, BMD alone or with VFA, or sequential fracture risk assessment following by BMD with or without VFA, with comparisons to no screening or another screening strategy. Eligible outcomes included hip fractures, clinical fragility fractures, fracture-related mortality, functionality and disability, quality of life or wellbeing, all-cause mortality, and serious adverse events including AFF, ONJ. In addition, nonserious adverse events, discontinuations due to adverse events, and overdiagnosis were also eligible. **Number of included studies:** 5 for KQ 1 (fractures/mortality); 2 for KQ 3 (overdiagnosis)	Clinical fragility fractures (3 RCTs) Pooled RR: 0.93 (95% CI, 0.87 to 0.99) ARD: 5.9 fewer per 1,000 (95% CI, 10.9 fewer to 0.8 fewer) GRADE certainty: Moderate for reduction Hip fractures (3 RCTs + 1 CCT*) Pooled RR: 0.80 (95% CI, 0.71 to 0.91) ARD: 6.2 fewer per 1,000 (95% CI, 9.0 fewer to 2.8 fewer) GRADE certainty: Moderate for reduction All-cause mortality (2 RCTs + 1 CCT*) Pooled RR: 1.00 (95% CI, 0.92 to 1.09) ARD: No difference in 1,000 (95% CI, 7.1 fewer to 5.3 more) GRADE certainty: Moderate for no reduction Overdiagnosis (2 RCTs) Among women ages 70 to 85 years: 11.8% overdiagnosed in the offer to screen population; 24.1% overdiagnosed among those considered at high risk. Among women ages 65 to 90 years: 19.3% overdiagnosed

Appendix D Table 9. Included Systematic Reviews for Direct Benefits and Harms of Screening (Key Question 1 and 3)

Author, Year Study Quality	Review Description	Outcomes
Merlijn et al, 2020[130] Good	**Search dates:** Inception to June 20, 2019 **Data sources:** Embase, MEDLINE **Inclusion/exclusion criteria:** RCTs in general population that used at least bone densitometry for screening and used anti-osteoporosis medication including bisphosphonates, denosumab, or strontium ranelate for any subsequent treatment with fractures as a reported outcome and usual care as a comparator group. Number of included studies: 3	All fractures (3 RCTs) Pooled HR, 0.95 (95% CI, 0.89 to 1.02) Osteoporotic fractures (3 RCTs) Pooled HR, 0.95 (95% CI, 0.89 to 1.00) MOF (3 RCTs) Pooled HR, 0.91 (95% CI, 0.84 to 0.98) Hip fractures (3 RCTs) Pooled HR, 0.80 (95% CI, 0.71 to 0.91) All-cause mortality (2 RCTs) Pooled HR, 1.04 (95% CI, 0.95 to 1.14)

* The review authors describe this study as a controlled clinical trial; however, the primary study design is described as a nonconcurrent cohort study.

Abbreviations: AFF=atypical femur fracture**;** ARD=absolute risk difference**;** CCT= controlled clinical trial; CI=confidence interval; VFA=vertebral fracture assessment; HR=hazard ratio; KQ=key question; MOF=major osteoporotic fracture; ONJ=osteonecrosis of the jaw; RCT=randomized, controlled trial; RR=relative risk ratio.

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Azagra et al, 2015[168, 310] FRIDEX	Osteoporotic fracture: 9.3% MOF: 6% Hip fracture: 1.8%	Reported in one or more of the included SRs
Azagra et al, 2016[174] FROCAT	Hip fracture: 2.2% All ages MOF: 11.7% All ages Hip fracture: 0.4% Age <65 years MOF: 7.0% Age <65 years	FRAX MOF without BMD/≥5%/MOF/10 years AUC: NR Sensitivity: 52.8% (95% CI, NR) Specificity: NR (95% CI, NR) Excluding women taking osteoporosis medication FRAX MOF without BMD/≥5%/MOF/10 years AUC: NR (95% CI, NR) Sensitivity: 60.6% (95% CI, NR) Specificity: 71.5% (95% CI, NR) NOT excluding women taking osteoporosis medication
Baleanu et al, 2021[179]	MOF: 9.3% (5 yrs) Garvan defined OF: 11.7% (5 yrs) Hip: 1.5% (5 yrs)	FRAX Hip with BMD/3%/Hip/5 years AUC: 0.841 (95% CI, 0.795 to 0.887) Sensitivity: 77% (95% CI, NR) Specificity: 72% (95% CI, NR) Garvan Hip with BMD/3%/Hip/5 years AUC: 0.769 (95% CI, 0.702 to 0.836) Sensitivity: 81% (95% CI, NR) Specificity: 59% (95% CI, NR) *FRAX MOF with BMD/NR/MOF/5 years* AUC: 0.708 (95% CI, 0.675 to 0.741) Sensitivity: 26% (95% CI, NR) Specificity: 93% (95% CI, NR) Garvan OF with BMD/20%/Any OF/5 years AUC: 0.721(95% CI, 0.693 to 0.749) Sensitivity: 27% (95% CI, NR) Specificity: 93% (95% CI, NR)
Bolland et al, 2011[166]	MOF: 16.1% FRAX defined MOF Hip fracture: 4.0% OF: 19.6% Garvan defined OF	Reported in one or more of the included SRs

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159] Moller et al, 2022[163] Leslie et al, 2022[164] Manitoba BMD Registry	Brennan et al[158] MOF: 11.0% Kaplan-Meier 10-year estimate Hip: 3.2% Kaplan-Meier 10-year estimate Leslie et al, 2010[157] Hip fracture: Women, 1.4%; Men, 1.5% MOF: Women, 6.5%, Men, 5.7% Leslie et al, 2018[162] Hip Fracture: Women, 2.4%; Men, 1.7% MOF: Women, 8.6%, Men, 6.3% Leslie et al, 2016[161] MOF Men and women combined, 11.5%	Brennan et al, 2014[158]; Leslie et al, 2010[157]; Leslie et al, 2018[162]; Leslie et al, 2016[161]; Morin et al, 2009[159] Discrimination results reported in one or more of the included SRs Crandall et al, 2019[160] FRAX Hip with BMD/≥3%/hip/10 years AUC: NR (95% CI, NR) Sensitivity: 62.2% (95% CI, NR) Specificity: NR (95% CI, NR) NNS overall: 4 By age group Sn/Sp/NNS 40–49: 9.7%/99.3%/137 50–59: 12.0%/98.1%/50 60–69: 31.7%/89.7%/9 70–79: 66.1%/55.5%/2 80+: 94.0%/15.9%/1 FRAX MOF with BMD/≥20%/MOF/10 years AUC: NR (95% CI, NR) Sensitivity: 20.3% (95% CI, NR) Specificity: 92.7% (95% CI, NR) NNS overall: 11 Age-group-specific data Sn/Sp/NNS 40–49: 0%/99.9%/761 50–59: 1.5%/99.4%/159 60–69: 6.7%/97.1%/30 70–79: 23.6%/86.9%/7 80+: 58.5%/58.6%/2

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159] Moller et al, 2022[163] Leslie et al, 2022[164] Manitoba BMD Registry (continued)	Crandall et al, 2019[160] Hip fracture: 3.5% MOF: 11.4% All fractures: 14.9% Morin et al, 2009[159] MOF: 2.7% Hip: 0.23%	FRAX MOF with BMD/≥20%/any fragility/10 years AUC: NR (95% CI, NR) Sensitivity: 18.6% (95% CI, NR) Specificity: 92.9% (95% CI, NR) NNS overall: 11 By age group Sn/Sp/NNS 40–49: 0.2%/99.9%/761 50–59: 1.7%/99.5%/159 60–69: 6.4%/97.2%/30 70–79: 22.8%/87.2%/7 80+: 57.3%/59.1%/2 From Moller et al[163] and Leslie et al[164] FREM/NR/MOF/2 years AUC: 0.64 (95% CI, 0.60 to 0.69) Sensitivity: NR Specificity: NR Subgroup: Men FREM/NR/MOF/2 years AUC: 0.67 (95% CI, 0.65 to 0.68) Sensitivity: NR Specificity: NR Subgroup: Women FREM/NR/MOF/2 years AUC: 0.60 (95% CI, 0.57 to 0.63) Sensitivity: NR Specificity: NR Subgroup: Women age <65 years

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
157-164		FREM/NR/Hip/2 years AUC: 0.66 (95% CI, 0.57 to 0.76) Sensitivity: NR Specificity: NR Subgroup: Men FREM/NR/Hip/2 years AUC: 0.83 (95% CI, 0.81 to 0.86) Sensitivity: NR Specificity: NR Subgroup: Women FREM/NR/Hip/2 years AUC: 0.71 (95% CI, 0.62 to 0.80) Sensitivity: NR Specificity: NR Subgroup: Women age <65 years From Agarwal et al, 2023[423] FRISBEE model/X/MOF/2.5 years
Chapurlat et al, 2020[175] OFELY and QUALYOR	Vertebral and nonvertebral: Cannot determine MOF: Can't determine	FRAX MOF with BMD/20%/MOF/8 years AUC: 0.562 (95% CI, 0.49 to 0.63) Sensitivity: NR (95% CI, NR) Specificity: NR (955 CI, NR)
Cheung et al, 2012[150] Hong Kong Osteoporosis Study	MOF: 4.7% Hip fracture: 0.93%	Reported in one or more of the included SRs
Collins et al, 2011[167] THIN Database	Hip fracture: 1.37% Women MOF minus humerus: 3.0% Women Hip fracture: 0.47% Men MOF minus humerus: 1.0% Men	Reported in one or more of the included SRs

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative	Crandall et al, 2019[141] MOF: 14,105/115,257=12.2% Crandall et al, 2019[142] MOF: 17,435/99,413=17.5% Crandall et al, 2014[139] (limited to ages 50 to 64) MOF: 18.5% Hip: 2.1% Crandall et al, 2018[140] (limited to ages 50 to 64) Hip, MOF Age 50-54 years: 0.3%, 6.3% Age 55-59 years: 0.6%, 8.0% Age 60-64 years: 1.1%, 9.9% Crandall et al, 2023[143] (limited to ages 50 to 64) MOF: 8.3% By self-identified race/ethnicity: Asian: 5.3% Black: 4.6% Hispanic: 8.0% White: 8.8% P<0.001	FRAX MOF without BMD/NR/MOF/10 years[142] AUC: 0.65 (95% CI, 0.65 to 0.66) Sensitivity: NR Specificity: NR Results stratified by race, AUC White: 0.64 (95% CI, 0.64 to 0.65) Black: 0.61 (95% CI, 0.58 to 0.64) FRAX MOF with BMD/NR/MOF/10 years[142] Subset with BMD information (n=5,722) AUC: 0.70 (95% CI, 0.68 to 0.72) Sensitivity: NR Specificity: NR Results stratified by race, AUC White: 0.69 (95% CI, 0.66 to 0.71) Black: 0.66 (95% CI, 0.56 to 0.76) FRAX Hip without BMD/NR/Hip/10 years[142] AUC: 0.76 (95% CI, 0.75 to 0.77) Sensitivity: NR Specificity: NR Results stratified by race, AUC White: 0.75 (95% CI, 0.74 to 0.77) Black: 0.81 (95% CI, 0.75 to 0.88)

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		FRAX Hip with BMD/NR/Hip/10 years[142] Subset with BMD information (n=5,541) AUC: 0.78 (95% CI, 0.74 to 0.82) Sensitivity: NR Specificity: NR Results stratified by race, AUC White: 0.77 (95% CI, 0.73 to 0.81) Black: 0.85 (95% CI, 0.69 to 1.0) FRAX MOF without BMD/≥9.3%/MOF/10 years[139] Ages 50–64 years AUC: 0.56 (95% CI, 0.55 to 0.57) Sensitivity: 25.8% (95% CI, 24.6% to 27.0%) Specificity: 83.0% (95% CI, 83.0% to 83.6%) Results stratified by age (95% CI) Ages 50–54 years (n=14,679): AUC 0.54 (0.52 to 0.55); Sn 4.7% (3.3% to 6.0%); Sp 97.0% (96.8% to 97.3%) Ages 55–59 years (n=22,363): AUC 0.55 (0.53 to 0.56); Sn 20.5% (18.6% to 22.3%); Sp 86.3% (85.8% to 86.7%) Ages 60–64 years (n=25,450): AUC 0.56 (0.55 to 0.57); Sn 37.3% (35.4% to 39.1%); Sp 72.3% (71.7% to 72.9%) FRAX MOF without BMD/≥8.4%/MOF/10 years[141] Ages 50–54 years AUC: NR Sensitivity: 6.7% (95% CI, 5.2% to 8.2%) Specificity: 95.7% (95% CI, 95.4% to 96.0%) Ages 55–59 years AUC: NR Sensitivity: 21.7% (95% CI, 19.9% to 23.5%) Specificity: 85.7% (95% CI, 85.2% to 86.1%) Ages 60–64 years AUC: NR Sensitivity: 49.5% (95% CI, 47.6% to 51.4%) Specificity: 59.4% (95% CI, 58.8% to 60.0%)

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		For the FRAX 8.4% threshold; sensitivity higher and specificity lower for White women compared with other racial/ethnic groups[141] FRAX MOF without BMD/continuous/MOF/10 years[140] Ages 50–64 years only AUC: 0.58 (95% CI, 0.57 to 0.59) Sensitivity: NR Specificity: NR AUC (95% CI) by race White: 0.57 (0.56 to 0.58) African American: 0.53 (0.49 to 0.57) Hispanic: 0.57 (0.53 to 0.62) Other/Unknown: 0.61 (0.56 to 0.65) FRAX MOF without BMD/Varies see below/MOF/10 years[140] Results for age groups presented as 50–54/55–59/60–64 years AUC: 0.55/0.56/0.56 (95% CI, 0.53/0.54/0.55 to 0.57/0.57/0.57) Sensitivity: 26.7/33.6/37.5 (95% CI, 23.9/31.4/35.6 to 29.5/35.7/39.4) Specificity: 79.3/75.4/72.2 (95% CI, 78.7/74.8/71.6 to 80.0/76.0/72.8) Thresholds for the various age groups that maximize the AUC: 50–54: ≥5.10% 55–59: ≥7.04% 60–64: ≥9.27% FRAX MOF without BMD/NR/MOF/10 years[143] Ages 50–64 years AUC (95% CI) by race/ethnicity All: 0.59 (0.59 to 0.60) Asian: 0.65 (0.58 to 0.71) Black: 0.55 (0.52 to 0.59), P=0.01 vs. Asian, P=0.06 vs. Hispanic and vs. White Hispanic: 0.61 (0.56 to 0.65), P=0.31 vs. Asian, P=0.39 vs. White White: 0.59 (0.58 to 0.59), P=0.08 vs. Asian

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		FRAX Hip without BMD/continuous/hip/10 years[140] Ages 50–64 years AUC: 0.68 (95% CI, 0.65 to 0.70) Sensitivity: NR Specificity: NR AUC (95% CI) by race White: 0.66 (0.64 to 0.69) African American: 0.54 (0.36 to 0.73) Hispanic: 0.53 (0.30 to 0.76) Other/Unknown: 0.74 (0.58 to 0.89) FRAX Hip without BMD/>0.706/Hip/10 years[140] Ages 50–64 years AUC: 0.65 (95% CI, 0.62 to 0.67) Sensitivity: 59.2% (95% CI, 54.7% to 63.7%) Specificity: 67.6% (95% CI, 67.2% to 67.9%) AUC (95% CI) by race White: 0.64 (0.61 to 0.66) African American: 0.63 (0.51 to 0.75) Hispanic: 0.71 (0.61 to 0.81) Other/unknown: 0.74 (0.62 to 0.86) Garvan Garvan Hip without BMD/NR/Hip/10 years[140] Ages 50–64 years AUC: 0.62 (95% CI, 0.59 to 0.65) Sensitivity: NR Specificity: NR AUC (95% CI) by race[140] White: 0.61 (0.59 to 0.64) African American: 0.58 (0.39 to 0.76) Hispanic: 0.53 (0.33 to 0.73) Other/unknown: 0.61 (0.42 to 0.80)

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		Garvan Hip without BMD/>0.462/Hip/10 years[140] AUC: 0.58 (95% CI, 0.55 to 0.60) Sensitivity: 16.0% (95% CI, 12.7% to 19.4%) Specificity: 93.5% (95% CI, 93.3% to 93.7%) AUC (95% CI) by race White: 0.57 (0.55 to 0.60) African American: 0.61 (0.48 to 0.74) Hispanic: 0.71(0.58 to 0.83) Other/unknown: 0.67 (0.56 to 0.78) Garvan MOF without BMD/NR/MOF/10 years[140] Ages 50–64 years AUC: 0.57 (95% CI, 0.57 to 0.58) Sensitivity: NR Specificity: NR AUC (95% CI) by race White: 0.57 (0.56 to 0.58) African American: 0.54 (0.50 to 0.58) Hispanic: 0.57 (0.53 to 0.62) Other/Unknown: 0.56 (0.51 to 0.60) Garvan MOF without BMD/Varies see below/MOF/10 years[140] Results for age groups presented as: 50–54/55–59/60–64 years AUC: 0.56/0.56/0.56 (95% CI, 0.54/0.54/0.55 to 0.58/0.57/0.57) Sensitivity: 33.2/46.8/27.1 (95% CI, 30.2/44.5/25.4 to 36.2/49.1/28.9) Specificity: 74.7/63.1/81.6 (95% CI, 74.0/62.4/81.1 to 75.4/63.7/82.1) Thresholds for the various age groups that maximize the AUC: 50–54: ≥7.2% 55–59: ≥8.95% 60–64: ≥13.58

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		Other Instruments SCORE/>7/MOF/10 years[139] Ages 50–64 years AUC: 0.53 (95% CI, 0.53 to 0.54) Sensitivity: 38.6 (95% CI, 37.3 to 39.9) Specificity: 65.8 (95% CI, 65.4 to 66.2) Results stratified by age (95% CI) Ages 50–54 years (n=14,679): AUC 0.54 (0.52 to 0.56); Sn 18.5% (16.0% to 21.0%); Sp 78.8% (78.1% to 79.5%) Ages 55–59 years (n=22,363): AUC 0.53 (0.51 to 0.54); Sn 22.1% (20.2% to 24.0%); Sp 81.1% (80.5% to 81.6%) Ages 60–64 years (n=25,450): AUC 0.53 (0.52 to 0.54); Sn 57.6% (55.7% to 59.5%); Sp 44.4% (43.7% to 45.0%) OST/<2/MOF/10 years[139] Ages 50–64 years AUC: 0.52 (95% CI, 0.52 to 0.53) Sensitivity: 39.8 (95% CI, 38.5 to 41.1) Specificity: 60.7 (95% CI, 60.3 to 61.1) Results stratified by age (95% CI) Ages 50–54 years (n=14,679): AUC 0.54 (0.52 to 0.56); Sn 22.9% (20.1% to 25.6%); Sp 74.2% (73.5% to 74.9%) Ages 55–59 years (n=22,363): AUC 0.52 (0.51 to 0.53); Sn 36.7% (34.5% to 39.0%); Sp 63.9% (63.3% to 64.6%) Ages 60–64 years (n=25,450): AUC 0.54 (0.52 to 0.55); Sn 48.1% (46.2% to 50.1%); Sp 49.6% (48.9% to 50.2%) OST/continous/MOF/10 years[143] Ages 50–64 years AUC (95% CI) by race/ethnicity All: 0.55 (0.54 to 0.56) Asian: 0.62 (0.56 to 0.69) Black: 0.53 (0.50 to 0.57), P=0.02 vs. Asian, P=0.12 vs. Hispanic, P=0.34 vs. White Hispanic: 0.58 (0.54 to 0.62), P=0.27 vs. Asian, P=0.24 vs. White White: 0.55 (0.54 to 0.56), P=0.04 vs. Asian

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143] Women's Health Initiative (continued)		Discrimination results also reported for other score thresholds for FRAX, OST, and SCORE in Crandall et al[139]
Dagan et al, 2017[169]	Hip fracture: 2.7% MOF: 7.7%	Reported in one or more of the included SRs
Davis et al, 2019[173] Fremantle Diabetes Study Phase 1	Hip fracture: 4.0%	QFracture hip/≥3%/hip/10 years AUC: 0.82 (95% CI, 0.77 to 0.85) Sensitivity: 83.3% (95% CI, NR) Specificity: NR FRAX hip without BMD/NR/hip/10 years AUC: 0.80 (95% CI, 0.74 to 0.85) Sensitivity: NR Specificity: NR
Desbiens et al, 2020[177] CARTaGENE	Hip fracture: NR MOF: 1.6%	FRAX MOF without BMD/NR/MOF/5 years AUC: 0.66 (95% CI, 0.61 to 0.71) Sensitivity: NR Specificity: NR QFracture MOF/NR/MOF/5 years AUC: 0.66 (95% CI, 0.61 to 0.71) Sensitivity: NR Specificity: NR Garvan any fracture without BMD/NR/MOF/5 years AUC: 0.59 (95% CI, 0.56 to 0.62) Sensitivity: NR Specificity: NR

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Ensrud et al, 2009[152] Premaor et al, 2013[153] Study of Osteoporotic Fractures (SOF)	Hip fracture: 6.0% MOF: 17% Any clinical fracture: 30%	Reported in one or more of the included SRs
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146] MrOs	From Ettinger et al[145] Hip fracture: 2.7% MOF: 6.4% From Ettinger et al[72] Hip fracture: 2.6% MOF: 5.7% From Gourlay et al[146] Hip fracture: 4.5% MOF: 10.9%	Reported in one or more of the included SRs[72, 145] From Gourlay et al[146] FRAX hip with BMD/≥1.0%/hip/unclear AUC: 0.77 (95% CI, 0.73 to 0.82) Sensitivity: 90% (95% CI, 84% to 95%) Specificity: 43% (95% CI, 43% to 46%) Threshold selected to achieve 90% sensitivity Garvan hip with BMD/≥0.85%/hip/Unclear AUC: 0.78 (95% CI, 0.74 to 0.82) Sensitivity: 0.90 (95% CI, 0.84 to 0.95) Specificity: 0.43 (95% CI, 0.41 to 0.44) Threshold selected to achieve 90% sensitivity

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146] MrOs (continued)		QFracture hip/≥1.44%/hip/unclear AUC: 0.69 (95% CI, 0.64 to 0.74) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 36% (95% CI, 35% to 37%) Threshold selected to achieve 90% sensitivity FRAX hip without BMD/≥1.60%/hip/unclear AUC: 0.70 (95% CI, 0.66 to 0.73) Sensitivity: 90% (95% CI, 86% to 94%) Specificity: 36% (95% CI, 35% to 37%) Threshold selected based on sensitivity of 90% Garvan hip without BMD/≥2.14%/hip/unclear AUC: 0.71 (95% CI, 0.67 to 0.74) Sensitivity: 90% (95% CI, 86% to 95%) Specificity: 35% (95% CI, 33% to 36%) Threshold selected based on sensitivity of 90% QFracture hip/≥1.48%/hip/unclear AUC: 0.69 (95% CI, 0.66 to 0.73) Sensitivity: 90% (95% CI, 86% to 95%) Specificity: 36% (95% CI, 35% to 38%) Threshold selected based on sensitivity of 90%; also this is for the larger set of men without BMD FRAX MOF with BMD/≥5.28%/MOF/unclear AUC: 0.72 (95% CI, 0.68 to 0.76) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 40% (95% CI, 38% to 41%) Threshold selected based on sensitivity of 90% Garvan MOF with BMD/≥9.78%/MOF/unclear AUC: 0.74 (95% CI, 0.70 to 0.78) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 42% (95% CI, 41% to 43%) Threshold selected based on sensitivity of 90%

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146] MrOs (continued)		QFracture MOF/≥2.30%/MOF/unclear AUC: 0.65 (95% CI, 0.61 to 0.70) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 27% (95% CI, 26% to 28%) Threshold selected based on sensitivity of 90% FRAX MOF without BMD/≥6.03%/MOF/unclear AUC: 0.66 (95% CI, 0.62 to 0.69) Sensitivity: 90% (95% CI, 85% to 94%) Specificity: 33% (95% CI, 32% to 34%) Threshold selected based on sensitivity of 90% Garvan MOF without BMD/≥4.15%/MOF/unclear AUC: 0.66 (95% CI, 0.62 to 0.70) Sensitivity: 90% (95% CI, 85% to 94%) Specificity: 25% (95% CI, 24% to 26%) Threshold selected based on 90% sensitivity QFracture MOF/≥2.49%/MOF/unclear AUC: 0.64 (95% CI, 0.61 to 0.68) Sensitivity: 90% (95% CI, 85% to 94%) Specificity: 30% (95% CI, 29% to 32%) Threshold selected based on 90% sensitivity FRAX hip >3% or MOF >20% has sensitivity of 72.6% for predicting hip fracture FRAX hip >3% or MOF >20% has sensitivity of 8.5% for predicting MOF FRAX hip with BMD and Garvan hip with BMD and BMD alone were equivalent and were statistically better than QFracture.

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Fraser et al, 2011[154] Langsetmo et al, 2011[155] Canadian Multicentre Osteoporosis Study Fraser et al, 2011[154] Langsetmo et al, 2011[155] Canadian Multicentre Osteoporosis Study (continued)	From Fraser et al[154] MOF: 12.0% Women (Kaplan-Meier estimates) MOF: 6.4% Men (Kaplan-Meier estimates) Hip fracture: 2.4% Men (Kaplan-Meier estimates) Hip fracture: 2.7% Women (Kaplan-Meier estimates) From Langsetmo et al[155] Women Combined men and women: 97 hip fractures, 174 forearm, 100 upper arm, 89 spine.	Reported in one or more of the included SRs
Garcia-Sempere et al, 2022[181]	MOF: NR Hip: 1.5%	FRAX Hip without BMD/≥3%/8 years AUC: 0.836 (95% CI, 0.805 to 0.866) Sensitivity: 60.0% Specificity: 85.5%

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Goldshtein et al, 2018[172] Maccabi Healthcare Services	Hip fracture: 2.9% MOF: 13.5%	FRAX MOF without BMD/NR/MOF/10 years AUC: 0.65 (95% CI, NR) Sensitivity: NR Specificity: NR AUC by age ≥70 years: 0.57 <70 years: 0.59 p=0.01 for difference in AUC by age FRAX hip without BMD/NR/hip/10 years AUC: 0.82 (95% CI, NR) Sensitivity: NR Specificity: NR AUC by age ≥70 years: 0.64 <70 years: 0.72 p<0.001 for difference in AUC by age *FRAX hip with BMD/NR/hip/10 years* AUC: 0.82 (95% CI, 0.79 to 0.84) Sensitivity: NR Specificity: NR FRAX MOF with BMD/NR/MOF/10 years AUC: 0.67 (95% CI, 0.66 to 0.68) Sensitivity: NR Specificity: NR
Gonzalez-Macias et al, 2012[149] ECOSAP	MOF (minus vertebral): 3.9% Hip: 0.97%	Reported in one or more of the included SRs
Hippsley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148] QResearch Database	From Hippisley-Cox et al[147] MOF: 28,865 events, crude rate 245 per 100,000 person-years (95% CI, 242 to 247) From Hippisley-Cox et al[148] MOF minus humerus: Women: 13,952 Men: 4,519 Hip fracture: Women: 5,424 Men: 1,738	Reported in one or more of the included SRs

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170] Clinical Practice Research Datalink (CPRD)	From Hippisley-Cox et al[171] Women MOF: 34,528 events; 2.89 per 1,000 person-years (95% CI, 2.58 to 3.20) Hip: 17,533 events; 1.32 per 1,000 person-years (95% CI, 1.30 to 1.34) Men MOF: 11,169 events; 1.29 per 1,000 person-years (95% CI, 1.05 to 1.52) Hip: 5,707 events; 0.65 per 1,000 person-years (95% CI, 0.63 to 0.67) From Klop et al[170] MOF: 6.2% Hip fracture: 2.4% Subset of participants with hospital-linked data; estimated incidence (differs from crude)	Reported in one or more of the included SRs

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Jain et al, 2023[311] Temple University Hospital	MOF rate White women: 7.7 per 1,000 PY Black women: 3.9 per 1,000 PY Hispanic women: 6.3 per 1,000 PY White men: 4.9 per 1,000 PY Black men: 3.2 per 1,000 PY Hispanic men: 3.9 per 1,000 PY Hip rate White women: 1.08 per 1,000 PY Black women: 0.51 per 1,000 PY Hispanic women: 0.41 per 1,000 PY White men: 0.75 per 1,000 PY Black men: 0.56 per 1,000 PY Hispanic men: 0.68 per 1,000 PY	FRAX MOF without BMD/threshold NR/MOF/adjusted for followup time (median 4 years) AUC: 0.69 (95% CI, 0.67 to 0.71) By race White: 0.69 Black: 0.70 Hispanic: 0.68 Authors report no significant difference by race. Sensitivity and specificity at various threshold selected based on distribution (top 1%, top 10%, top 20%) 10.1% risk: Sn 3.8%; Sp 99.1% 4.2% risk: Sn 24.7%; Sp 90.3% 2.9% risk: Sn 41.3%; Sp 80.4% QFracture/threshold NR/MOF/adjusted for followup time (median 4 years) AUC: 0.70 (95% CI, 0.68 to 0.73) P=0.08 for FRAX vs. QFracture By race White: 0.70 Black: 0.70 Hispanic: 0.68 Authors report no significant differences by race Sensitivity and specificity at various threshold selected based on distribution (top 1%, top 10%, top 20%) 6.7% risk: Sn 4.5%; Sp 99.1% 2.3% risk: Sn 28.5%; Sp 90.4% 1.4% risk: Sn 42.8%; Sp 80.5% FRAX Hip without BMD/threshold NR/Hip/adjusted for followup time (median 4 years) AUC: 0.77 (95% CI, 0.72 to 0.83) By race White: 0.75 Black: 0.81 Hispanic: 0.77 Authors report no significant differences by race Sensitivity and specificity at various threshold selected based on distribution (top 1%, top 10%, top 20%) 4.5% risk: Sn 7.6%; Sp 99.0% 1.2% risk: Sn 37.9%; Sp 57.6% 0.6% risk: Sn 57.6%; Sp 80.2% QFracture/threshold NR/Hip/adjusted for followup time (median 4 years) AUC: 0.79 (95% CI, 0.74 to 0.84) P=0.21 for FRAX vs. QFracture By race White: 0.77 Black: 0.83 Hispanic: 0.77 Authors report no significant differences by race

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
		Sensitivity and specificity at various threshold selected based on distribution (top 1%, top 10%, top 20%) 4.0% risk: Sn 9.1%; Sp 99.0% 0.8% risk: Sn 37.9%; Sp 90.1% 0.4% risk: Sn 59.1%; Sp 80.1%
Lo et al, 2011[73] Pressman et al, 2011[165] Kaiser Permanente Northern California	Hip fracture: 1.7%	Reported in one or more of the included SRs
Lu et al, 2021[178] 5 cohorts (UK Biobank, MrOs US, MrOs Sweden, SOF, CKB)	Hip: Range 0.3% to 15.6% across the 5 cohorts MOF: Range 1.3% to 20.6% across the 5 cohrots	FRAX MOF with BMD/NR/MOF/unclear followup time AUC: 0.756 (95% CI, 0.749 to 0.813) Sensitivity: NR Specificity: NR FRAX Hip with BMD/NR/hip/unclear followup time AUC: 0.806 (95% CI, 0.799 to 0.813) Sensitivity: NR Specificity: NR
Marques et al, 2017[176] 3 different Portuguese cohorts (SAOL, IPR, EPIPorto)	Hip fracture: 1.1% MOF: 6.8%	FRAX hip without BMD/threshold NR/hip/10 years AUC: 0.72 (95% CI, 0.69 to 0.87) Sensitivity: NR Specificity: NR Women FRAX hip with BMD/threshold NR/hip/10 years AUC: 0.75 (95% CI, 0.62 to 0.87) Sensitivity: NR Specificity: NR Women FRAX hip without BMD/threshold NR/hip/10 years AUC: 0.93 (95% CI, 0.89 to 0.95) Sensitivity: NR Specificity: NR Men

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Marques et al, 2017[176] 3 different Portuguese cohorts (SAOL, IPR, EPIPorto) (continued)		FRAX hip with BMD/threshold NR/hip/10 years AUC: 0.90 (95% CI, 0.86 to 0.93) Sensitivity: NR Specificity: NR Men FRAX MOF without BMD/threshold NR/MOF/10 years AUC: 0.75 (95% CI, 0.73 to 0.77) Sensitivity: NR Specificity: NR Women FRAX MOF with BMD/threshold NR/MOF/10 years AUC: 0.76 (95% CI, 0.74 to 0.78) Sensitivity: NR Specificity: NR Women FRAX MOF without BMD/threshold NR/MOF/10 years AUC: 0.80 (95% CI, 0.76 to 0.84) Sensitivity: NR Specificity: NR Men FRAX MOF with BMD/threshold NR/MOF/10 years AUC: 0.85 (95% CI, 0.81 to 0.88) Sensitivity: NR Specificity: NR Men

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Pluskiewicz et al, 2023[312] RAC-OST-POL	MOF: 97 fractures in 978 women Hip: 15 fractures in 978 women	FRAX MOF with BMD/no threshold/MOF/10 years AUC: 0.65 (95% CI, 0.59 to 0.71) 6% risk threshold (empirically derived) Sn: 70.2% Sp: 51.5% 10% risk threshold Sn: 20% (95% CI, 12% to 29%) Sp: 91% (95% CI, 88% to 93%) FRAX Hip with BMD/no threshold/Hip/10 years AUC: 0.75 (95% CI, 0.61 to 0.88) 1.4% risk threshold (empirically derived) Sn: 69.0% Sp: 71.4% 3% risk threshold Sn: 50% (95% CI, 24% to 76%) Sp: 90% (95% CI, 88% to 93%) Garvan with BMD/no threshold/Any Fracture/10 years AUC: 0.64 (0.58 to 0.69) 14.4% risk threshold (empirically derived) Sn: 56.9% Sp: 68.8% 10% risk threshold Sn: 86% (95% CI, 79% to 91%) Sp: 29% (95% CI, 25% to 34%) Garvan with BMD/no threshold/Hip/10 years AUC: 0.75 (95% CI, 0.60 to 0.90) 8.8% risk threshold (empirically derived) Sn: 86.7% Sp: 64.3% 3% risk threshold Sn: 71% (95% CI, 42% to 90%) Sp: 60% (95% CI, 56% to 64%)
Tamaki et al, 2011[151] Japanese Population-Based Osteoporosis Cohort	MOF: 5.3% Hip: 0.5%	Reported in one or more of the included SRs

Appendix D Table 10. Discrimination Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Author, Year; Cohort Name	Incidence of Fractures	Discrimination Results
Tanaka et al, 2010[156] Multiple Japanese Cohorts	MOF: 15% Hip fracture: 2% Vertebral fracture: 12%	Reported in one or more of the included SRs
Tebe Cordomi et al, 2013[144] CETIR	MOF: 18.1% Hip fracture: 1.1%	Reported in one or more of the included SRs
Tebe et al, 2022[180]	MOF: unclear Hip: unclear	EPIC/NR/MOF/5 years AUC: 0.706 (95% CI, NR) Sensitivity: NR Specificity: NR
Zwart et al, 2023[313]	MOF: 15 (5.3) Hip: 2 (0.7)	NR

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; ECOSAP=Ecografia Osea en Atencio Primaria; EPIC= Escala dePredicci´on de fracturas Implementable en historia Clínica electronica; FRAX=Fracture Risk Assessment Tool; FREM=Fracture Risk Evaluation Model; FRIDEX=Fracture RIsk factors and bone DEnsitometry type central dual X-ray; FROCAT=abbreviation not defined; MOF=major osteoporotic fracture; MrOs= Osteoporotic Fractures in Men Cohort; NNS=number needed to screen; NR=not reported; OF=osteoporotic fracture; OFELY=Os des Femmes de Lyon; OST=Osteoporosis Self-Assessment Tool; QUALYOR=QUalité Osseuse LYon Orléans; SCORE=Simple Calculated Osteoporosis Risk Estimation; SOF=Study of Osteoporotic Fractures; SR=systematic review; THIN=The Health Improvement Network; U.K.=United Kingdom; U.S.=United States.

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313]	***Crandall et al, 2014***[139] Observed/expected ratios over deciles of risk No BMD, MOF: Range 0.76 to 1.15; calibration slope=1.04; overall O/E ratio=1.0; plot shows slight overprediction at lowest risk categories, slight underprediction at mid- to higher risk categories, except for the highest risk category, which was 0.97 No BMD, hip: range 0.27 to 1.63; calibration slope 1.59; overall O/E ratio=1.0; plot shows significant overprediction at lowest risk categories and significant underprediction at 3 highest risk categories ***Crandall et al, 2023***[143] Among younger women and by race/ethnicity O/E ratio range 0.95 to 1.06 across quantiles of risk for all participants, but wider ranges within each race/ethnicity O/E range; Calibration slope Asian: 0.88 to 1.20; 1.12 Black: 0.92 to 1.08; 1.26 Hispanic: 0.87 to 1.10; 1.00 White: 0.95 to 1.03; 1.08 ***Ettinger et al, 2013***[145] Hip without BMD: Average O/E ratio across the 5 quintiles of risk was 1.0 (range from 0.9 to 1.1) Hip with BMD: Average O/E ratio across the 5 quintiles was 1.3 (range 0.4 to 2.0, risk underestimated in 4 of the 5 quintiles) MOF without BMD: Average O/E ratio across the 5 quintiles was 0.8 (range 0.7 to 0.9, suggesting overestimation of risk)	***Ettinger et al, 2013***[145] Calibration plots are reported in Figures 1A and 1B of the study report ***Cheung et al, 2012***[150] Calibration plots for hip and MOF (with or without BMD) were shown in Figures 1 and 2 of the manuscript depicting percentage of subjects sustaining fracture across quartiles of risk; however, no O/E ratios or other measures of fit were reported ***Tamaki et al, 2011***[151] Calibration plot depicted in Figure 2 of the manuscript; absolute 10-year risk plotted by quartile of risk; p<0.05 for trend for FRAX MOF and hip (with or without BMD) ***Ensrud et al, 2009***[152] Calibration for the U.S. FRAX tool calibrated for Caucasians. Calibration plots are presented in Figures 2a and 2b. The proportion of women with fracture in each quartile of risk is depicted. Observed vs. expected proportions NR; a visual linear trend is observed but no statistical test of trend reported. Rather, the authors compared the observed and expected proportions in each quartile between	***Goldshtein et al, 2018***[172] FRAX MOF with BMD: 13.6% vs. 7.0% FRAX hip with BMD: 2.0% vs. 1.8% (Hosmer-Lemeshow p<0.001 for both) ***Gourlay et al, 2017***[146] FRAX MOF: 0.0046 with BMD; 0.0001 without BMD ***Garcia-Sempere et al, 2022***[181] FRAX hip without BMD Hosmer-Lemeshow Goodness of Fit: P=0.52

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	MOF with BMD: average O/E ratio across the 5 quintiles was 0.9 (range 0.7 to 1.1, suggesting overestimation of risk) ***Fraser et al, 2011***[154] With BMD predictions in quintiles vs. observed risk Women MOF: All predictions except the middle quintile were within 95% CI of the observed risks, regression slope 1.07, observed risks were at or above predicted risks across all quintiles Men MOF: All predictions except the second lowest quintile were within 95% CI of the observed risks, regression slope 1.26, observed risks were at, above, or below predicted risks across all quintiles Women hip: All predictions were within 95% CI of the observed risks, regression slope 0.93, observed risks were at, above, or below expected across all quintiles Men hip: Predictions from the lowest 3 quintiles were within 95% CI of the observed risks; the number of observed fractures in the highest two quintiles was much greater than expected, regression slope 1.83 ***Leslie et al, 2010***[157] Observed vs. expected risks by quintiles of risk Women hip: All predicted risks were within 95% CI of observed risks and very close to predicted risks; regression slope 1.03 (95% CI, 1.02 to 1.04) Men hip: All predicted risks within 95% CI of observed risks, but wide CIs observed for the top 3 quintiles, risks were underestimated in the 3rd quintile and overestimated in the 5th quintile, regression slope 0.92 (95% CI, 0.57 to 1.27)	FRAX and models with age and BMD or age and prior fracture history, all of which had no statistical difference from each other. In a companion article, calibration was reported separately for the N=285 obese women compared with the nonobese women. FRAX for hip fracture was underestimated in obese women compared with nonobese women in the lower two risk quartiles; FRAX MOF performed well in all risk quartiles for both obese and nonobese women. ***Fraser et al, 2011***[154] Calibration plots depicted in Figure 2. Women MOF: All predictions except the middle quintile were within 95% CI of the observed risks, regression slope 1.07, observed risks were at or above predicted risks across all quintiles Men MOF: All predictions except the second lowest quintile were within 95% CI of the observed risks, regression slope 1.26, observed risks were at, above, or below predicted risks across all quintiles Women hip: All predictions were within 95% CI of the observed risks, regression slope 0.93, observed risks were at, above, or below across all quintiles Men hip: Predictions from the lowest 3 quintiles were within 95% CI of the observed risks, the number of observed fractures in the highest two quintiles was much greater than expected, regression slope 1.83	

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	Women MOF: Predicted risks were within 95% CI of observed risks for the lowest 3 quintiles but underestimated risk on the top 2 quintiles; regression slope 1.13 (95% CI, 1.08 to 1.19) Men MOF: Predicted risks were within 95% CI of all quintiles except the middle one, and results were more variable due to wider CIs, predicted risks underestimated in the top 3 quintiles, underestimated in the 2nd quintile and was reasonably in agreement at the lowest quintile; regression slope 1.24 (95% CI, 1.00 to 1.48) ***Leslie et al, 2016***[161] Observed incidence by categories of risk MOF (men and women reported together) Low (<10%): 6.0% Moderate (10%–20%): 13.8% High (>20%): 25.1% ***Azagra et al, 2015***[168] O/E ratio (95% CI) FRAX MOF with BMD: 1.61 (1.19 to 2.12) FRAX MOF without BMD: 1.72 (1.27 to 2.27) ***Tamaki et al, 2011***[151] Observed/expected factures; p-value MOF with BMD: 43/49.6; p=0.550 MOF without BMD: 43/49.2; p=0.577 Hip with BMD: 4/8; p=0.382 Hip without BMD: 4/9; p=0.263	***Leslie et al, 2010***[157] Calibration plots in the manuscript are not presented by categories of predicted risk, therefore not considered by our team ***Brennan et al, 2014***[158] Calibration of FRAX for Canada, for both hip and MOF (with and without BMD) across three predicted risk groups (low <10%, moderate 10%-19%, and high ≥20%) in five income quintiles. Results are depicted in Figures 1, 2, and 3. Authors state "good concordance" between observed and predicted risk, but no O/E ratios or statistics reported. Models that did not account for competing mortality risk generally underestimated risk in the highest risk category across all income levels and for both MOF and hip with and without BMD. ***Bolland et al, 2011***[166] Calibration plots were depicted in Figure 2 of the manuscript. MOF with BMD consistently underestimated risk across all deciles, with the worst underestimation at the two highest deciles (p<0.01)	

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	***Pressman et al, 2011***[165] Hip by age group; model appears to underestimate risk in older age groups 50–59 Observed: product limit estimate from proportional hazards model (PLE) 0.41 (95% CI, 0.35 to 0.47) Predicted (with BMD): PLE 0.25 Predicted (without BMD): PLE 0.34 60–69 Observed: PLE 2.00 (95% CI, 1.85 to 2.15) Predicted (with BMD): PLE 0.68 Predicted (without BMD): PLE 1.11 70–79 Observed: PLE 8.00 (95% CI, 7.62 to 8.38) Predicted (with BMD): PLE 2.80 Predicted (without BMD): PLE 4.03 80+ Observed: PLE 20.0 (18.66 to 21.34) Predicted (with BMD): PLE 4.90 Predicted (without BMD): PLE 9.21 ***Crandall et al, 2019***[141] Observed % vs. expected % Hip without BMD expected vs. observed Overall (N=62,723): 0.7% vs. 0.7% By age Ages 50–54 (N=14,768): 0.3% vs. 0.3% Ages 55–59 (N=22,442): 0.5% vs. 0.6% Ages 60–64 (N=25,513): 1.1% vs. 1.1% By Race/Ethnicity White (N=52,536): 0.8% vs. 0.8% African American (N=5,475): 0.2% vs. 0.2% Hispanic (N=2,262): 0.3% vs. 0.4% Other (N=2,450): 0.5% vs. 0.4% No calibration data for MOF.	MOF without BMD also consistently underestimated risk in all but 1 decile (p<0.01) Hip with BMD underestimated risk at 8 lowest deciles, underestimated the 9th and overestimated the 10th (p<0.01) Hip without BMD reasonably calibrated across lowest 7 deciles, overestimation at 3 highest deciles of predicted risk (p=0.18) ***Crandall et al, 2019***[160] Calibration plots were depicted in Figure 2 of the manuscript using quintiles of risk and reported by authors as "good overall" for FRAX with and without BMD. MOF with BMD: Observed and predicted are close to the line of identity MOF without BMD: Risk appears overestimated in the 2 highest quintiles of risk Hip with BMD: Risk appears slightly underestimated in the 2nd highest quintile and significantly underestimated in the highest quintile of risk	

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	***Zwart et al, 2023***[313] FRAX without BMD MOF Women: 1.12 (95% CI, 0.95 to 1.29) Men: 0.57 (95% CI, 0.01 to 1.14) All persons <65 years: 1.23 (95% CI, 0.01 to 2.47) All persons ≥65 years: 1.30 (95% CI, 1.10 to 1.50) Hip Women: 0.47 (95% CI, 0 to 0.94) Men: 0 fractures observed All persons <65 years: 0 fractures observed ≥65 years: 0.48 (95% CI, 0 to 0.97) ***Pluskiewicz et al, 2023***[312] FRAX with BMD MOF: 97 observed/69 predicted; 1.41 Hip: 14 observed/67 predicted: 0.21 ***Garcia-Sempere et al, 2022***[181] FRAX Hip without BMD O/E ratios by quintiles of risk for hip fracture for FRAX Hip without BMD Quintile 1: 0.48 (95% CI, 0.47 to 0.48) Quintile 2: 0.77 (95% CI, 0.77 to 0.77) Quintile 3: 1.12 (95% CI, 1.11 to 1.13) Quintile 4: 1.13 (95% CI, 1.12 to 1.15) Quintile 5: 0.97 (95% CI, 0.94 to 1) ***Goldshtein et al, 2018***[172] Observed vs. expected MOF without BMD: 13.5% vs. 6.9% ≥70 years: 26.0% vs. 16.1% <70 years: 10.6% vs. 4.8% Hip without BMD: 2.9% vs. 2.2% ≥70 years: 10.1% vs. 8.1% <70 years: 1.2% vs. 0.9%	Hip without BMD: risk appears slightly overestimated in the 2nd highest quintile and significantly overestimated in the highest quintile of risk. Observed vs. expected was also plotted by age group in Figure 3 of the manuscript and also reported as "good across age groups" by study authors. Risk for both hip and MOF is overpredicted by FRAX without BMD in the 80+ age group, and to a lesser extent in the 70–79 age group. ***Goldshtein et al, 2018***[172] Calibration plots depicted by decile of risk in Figure 1 of manuscript, shows underestimation of risk in all deciles for MOF, similar finding for hip but degree of underestimation is less compared to MOF (Hosmer-Lemeshow goodness of fit P <0.001 for both) Hip with BMD: Risks were overestimated at 5 lowest deciles of risk and underestimated at 5 highest deciles of risk ***Gonzalez-Macias et al, 2012***[149] Calibration plots presented in Figure 1 of the manuscript. ***Dagan et al, 2017***[169] Calibration plots were presented in Figure 3 of the manuscript. Slope, calibration in the large Women: 0.94; 0.39 Men: 0.94; 0.39	

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	***Marques et al, 2017***[176] Observed vs. estimated Without BMD Women MOF: 145 vs. 97.5 (95% CI, 78.6 to 116.3) Hip: 20 vs. 30.9 (95% CI, 20.1 to 41.8) Men MOF: 33 vs. 24.9 (95% CI, 15.3 to 34.6) Hip: 8 vs. 9.9 (95% CI, 3.82 to 16.1) With BMD Women MOF: 116 vs. 91.3 (95% CI, 73.1 to 109.4) Hip: 17 vs. 35.8 (95% CI, 24.2 to 44.4) Men MOF: 23 vs. 18.9 (95% CI, 10.6 to 27.2) Hip: 7 vs. 10.3 (4.1 to 16.5) In women, MOF was underestimated and hip was overestimated with and without BMD. In men, observed values were within 95% CI of predicted. ***Dagan et al, 2017***[169] O/E ratio hip fracture across ages (note these were adjusted for 5-year risk predictions) Women: Range 1.6 to 1.9 (no pattern across age groups) Men: Range 1.6 to 3.0 (no pattern across age groups) O/E ratio hip fracture across deciles of risk (note these were adjusted for 5-year risk predictions Women: 1.6 (highest decile) to 2.7 (lowest decile) Men: 1.8 (highest 4 deciles) to 4.1 (lowest decile)	***Klop et al, 2016***[170] Calibration plot for hip fracture was depicted in Figure 2b of the manuscript; observed fractures were close to predicted in all but the highest decile of risk, which was overestimated. ***Garcia-Sempere et al, 2022***[181] FRAX hip without BMD: Calibration plot reported in Figure 1. Slope=1.025; intercept=-0.0003. ***Desbiens et al, 2020***[177] Calibration plots are depicted in Figure 4 of the manuscript. FRAX globally overestimated risk across all predicted risk levels.	

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Primary studies)[139, 141, 143-146, 149-152, 154, 156-158, 160, 161, 165, 166, 168-170, 172, 176, 177, 311-313] (continued)	***Klop et al, 2016***[170] Observed vs. expected FRAX without BMD MOF: 6.2% (95% CI, 5.9 to 6.4) vs. 8.6% predicted Hip: 2.4% (95% CI, 2.2 to 2.7) vs. 2.7% predicted (in subset with hospital linked data) ***Tebe Cordomi et al, 2013***[144] With BMD measurement O/E ratio: 4.0 (95% CI, 3.4 to 4.5) OE ratio by age groups 40 to <55: 5.5 (95% CI, 4.4 to 6.9) 55 to <65: 4.1 (95% CI, 3.3 to 5.0) 65 to <75: 2.7 (95% CI, 2.1 to 3.6) ≥75: 3.2 (95% CI, 1.3 to 7.6) Observed fracture risks were higher at all deciles of risk compared to expected (predicted) fracture risk ***Tanaka et al, 2010***[156] O/E ratio across 10 deciles of risk With BMD MOF: 1.59 (underestimation, unclear which version of FRAX was used); p<0.01 ***Jain et al, 2023***[311] O/E Ratio (calibration in the large) FRAX: 0.97 (Overprediction in the highest quantile risk group for MOF and hip; otherwise, reasonable predictions across quantiles; underestimation in youngest age group 50-59) By race: adjustment factors to convert MOF risk for White men and women were accurate for Black women, but underestimated risk in Black men, Hispanic women and men Adjustment factors to convert Hip fracture risk for White men and women used by FRAX were accurate for Black women, but underestimated risk in Black and Hispanic men; Hispanic women NR.		

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRAX (Systematic review summary)[131]	High amount of heterogeneity that was not explained by age, sex, or baseline risk. FRAX without BMD. Hip fractures: O:E ratios ranged from 0.26 to 3.87 for the 13 high ROB studies and 0.93 to 1.71 for the 3 unclear ROB studies (all Canadian FRAX) MOF: O:E ratios ranged from 0.33 to 3.34 for the 12 high ROB studies and from 1.06 to 1.19 for the 3 unclear ROB studies (all Canadian FRAX) GRADE: very low certainty for poor performance for the high ROB studies for hip and MOF GRADE: low (hip) or moderate (MOF) certainty that may be well-calibrated (all Canadian FRAX) FRAX with BMD Hip fractures: O:E ratio ranged from 0.24 to 3.33 for 13 high ROB studies and from 1.00 to 1.85 for 3 unclear ROB studies (all Canadian FRAX) MOF: O:E ratios ranged from 0.44 to 3.90 for 16 high ROB studies and from 1.11 to 1.19 for 3 unclear ROB studies (all Canadian FRAX) GRADE: very low certainty for poor performance for the high ROB studies for hip and MOF GRADE: low certainty that may perform poorly for unclear ROB studies for hip and moderate certainty that probably well-calibrated for MOF (all Canadian FRAX)	NR	NR

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FREM[163, 164] (Primary Studies)	2 year Observed vs. Predicted Incidence MOF Men by age: (underprediction or overprediction depending on age group) <65 years: 1.36% vs. 0.52% (SD 0.73%) 65–79 years: 1.87% vs. 1.00% (SD 1.22%) 80+ years: 2.53% vs. 2.99% (SD 2.39%) Women by age: (overprediction) <65 years: 0.73% vs. 0.76% (SD 0.48%) 65–79 years: 1.29% vs. 1.61% (SD 0.91%) 80+ years: 2.8% vs. 3.88% (SD 2.17%) Hip (overprediction) Men by age <65 years: 0.10% vs. 0.17% (SD 0.36%) 65–79 years: 0.30% vs. 0.62% (SD 1.22%) 80+ years: 0.58% vs. 3.0% (SD 3.57%) Women: <65 years: 0.03% vs. 0.11% (SD 0.14%) 65–79 years: 0.19% vs. 0.49% (SD 0.54%) 80+ years: 0.90% vs. 2.36% (SD 1.78%)	***Leslie et al, 2022***[164] Supplement Table 4 reports fracture incidence per 1,000 person-years by FREM quintile for both MOF and Hip, stratified by sex, an observable increase in fracture rate is observed with increasing quintile of risk HR (reported in Supplemental Table 5) MOF and Hip; Women and Men: compared to middle quintile (HR 1.00); all other quintiles of FREM risk were significantly different from the middle category. However, when stratified by age (Supplemental Table 7), some quintiles no longer significantly different (pattern varied by men vs. women and by hip vs. MOF)	NR
FRC[72, 73, 149] (Primary studies)	***Lo et al, 2011***[73] Unclear whether with or without BMD O/E ratio by tertile of predicted risk Lowest (<1%): 1.3 Middle (1% to 2.9%): 1.3 Highest (3% to 4.9%): 1.4 Fractures were underestimated across all tertiles of risk ***Gonzalez-Macias et al, 2012***[149] ***Gonzalez-Macias et al, 2012***[149] Hip without BMD: E-O ratio 1.10; p<0.001 (indicating significant difference between E and O and limited predictive ability, higher underestimation at lower risk deciles compared with higher risk deciles) MOF without BMD: E-O ratio 0.66; p<0.001 (indicating significant difference between E and O and limited predictive ability; no clear pattern because of rarer events)	***Ettinger et al, 2012***[72] Calibration plots (with and without BMD) reported as Figures 2A and 2B in the article. 5 quintiles were examples for each model. Over the total of 20 quintiles evaluated for MOF (14 quintiles for hip), O/E ratios were within 20% of 1.0. O/E ratios between models with and without BMD did not vary, except for the highest quintile of risk where models with BMD overestimated the risk for both hip and MOF.	NR

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
FRC[131] (Systematic review summary)	FRC without BMD Hip: inconsistent findings in 2 studies (O/E 1.44 in women, 0.97 in men) GRADE: very low certainty for poor performance MOF: 1 study (O/E 0.95) GRADE: very low certainty for acceptable calibration in men FRC with BMD Hip: inconsistent findings in 2 studies (1.50 in women, 1.0 in men) GRADE: very low certainty for poor performance MOF: 1 study; O/E 0.96 GRADE: very low certainty for acceptable calibration	NR	NR
Garvan Fracture Risk Calculator[141, 146, 166, 169, 177, 312] (Primary studies)	***Dagan et al, 2017***[169] Hip fracture across ages (note these were adjusted for 5-year risk predictions) Women: 2.7 (highest age group) to 6.9 (lowest age group) Men: Range 0.6 (highest age group) to 5.2 (lowest age group) O/E ratio hip fracture across deciles of risk (note these were adjusted for 5-year risk predictions) Women: 2.4 (highest decile) to 21.2 (lowest decile) Men: 0.8 (highest decile) to 4.7 (lowest decile) Pluskiewicz et al, 2023[312] With BMD Any fracture: 129 observed/472 predicted; 0.27 Hip: 14 observed/258 predicted; 0.05	***Gourlay et al, 2017***[146] Calibration plots were depicted in Figure 1 of the manuscript by deciles of risk Hip with BMD: risks were overestimated at 5 lowest deciles of risk, in agreement with 6th and 7th decile, and underestimated at 3 highest deciles of risk. No calibration plots for MOF. ***Bolland et al, 2011***[166] Calibration plots were depicted in Figure 2 of the manuscript. Garvan hip overestimated risk at two highest deciles of predicted risk (p<0.01) ***Dagan et al, 2017***[169] Calibration plots were presented in Figure 3 of the manuscript. Slope, calibration in the large Women: 0.64; 0.18 Men: 0.68; -0.95	***Gourlay et al, 2017***[146] Hip: 0.0001 with BMD, <0.0001 without BMD MOF: 0.0001 with BMD; 0.0104 without BMD ***Bolland et al, 2011***[166] Garvan OF underestimated risk at lower predicted risk deciles and overestimated risk at higher predicted risk deciles (p<0.01) Garvan hip overestimated risk at two highest deciles of predicted risk (p<0.01)

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
Garvan Fracture Risk Calculator[141, 146, 166, 169, 177, 312] (Primary studies) (continued)	***Crandall et al, 2019***[141] Hip without BMD expected vs. observed Overall (N=62,723): 0.2% vs. 0.7% By age Ages 50–54 (N=14,768): 0.1% vs. 0.3% Ages 55–59 (N=22,442): 0.2% vs. 0.6% Ages 60–64 (N=25,513): 0.3% vs. 1.1% By race/ethnicity White (N=52,536): 0.2% vs. 0.8% African American (N=5,475): 0.1% vs. 0.2% Hispanic (N=2,262): 0.2% vs. 0.4% Other (N=2,450): 0.2% vs. 0.4% No data for MOF	FRAX MOF with BMD consistently underestimated risk across all deciles, with the worst underestimation at the two highest deciles (p<0.01) FRAX MOF without BMD also consistently underestimated risk in all but 1 decile (p<0.01) FRAX hip with BMD underestimates risk at 8 lowest deciles, underestimates the 9th and overestimates the 10th (p<0.01) FRAX Hip without BMD reasonably calibrated across lowest 7 deciles, overestimation at 3 highest deciles of predicted risk (p=0.18) ***Desbiens et al, 2020***[177] Calibration plots are depicted in Figure 4 of the manuscript. FRAX globally overestimated risk across all predicted risk levels while QFracture globally underestimated risk across all risk levels. Garvan underestimated risk at lower risk levels and overestimated risk at high risk levels	
Garvan[131] (Systematic review summary)	Garvan without BMD Hip: O/E 3.63; 1 study GRADE: Very low certainty for poor performance Garvan with BMD Hip: inconsistent across 5 studies, O/E 0.10 to 0.66 GRADE: Very low certainty for poor performance MOF: inconsistent across 4 studies; O/E 0.34 to 1.65 GRADE: very low certainty for poor performance	NR	Garvan without BMD Hip: 1 study; P<0.0001 indicating poor calibration MOF: 1 study; P=0.01014 GRADE: Very low certainty for poor performance Garvan with BMD Hip: NR MOF: 1 study, P=0.0001 indicating poor calibration GRADE: Very low certainty for poor performance

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
OST[424]	Takano et al, 2022[424] Among younger women (ages 50 to 64) and by race/ethnicity OE range across quantiles of risk All: 0.94 to 1.08 Asian: 0.81 to 1.18; 0.87 Black: 0.95 to 1.04; 1.31 Hispanic: 0.89 to 1.13; 0.92 White: 0.94 to 1.04; 0.92 Note: OST treated as a continuous variable.	NR	NR
QFracture (2009 version)[148, 167] (Primary studies)	***Hippsley-Cox et al, 2009***[148] MOF Women 0.92 to 1.09 Men 0.92 to 1.11 Hip (ages 40 to 85) Women 0.81 to 2.47 Men 0.84 to 1.53 ***Collins et al, 2011***[167] Brier Score (lower score means greater accuracy MOF (minus humerus): Women 0.027 (0.025 to 0.029) Men 0.010 (0.008 to 0.012) Hip Women 0.013 (0.012 to 0.015) Men 0.005 (0.003 to 0.007) Observed vs. predicted were also evaluated by age group; MOF and hip risk were slightly underestimated at the oldest age groups (>75) for women, but not men.	Calibration plot depicted in Figure 2; however, no statistical tests conducted.[148] Calibration plots presented in Figures 1 and 2 of the manuscript, overall good agreement between predicted and expected[167] Women MOF minus humerus and Hip: across deciles of risk, observed risks were very close to predicted except for the highest, which was not as close Men MOF minus humerus and hip: across deciles of risk, observed risks were very close to predicted except for the highest, which was not as close	NR

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
QFracture (2012 version)[146, 147, 169, 171, 173, 177, 311]	***Hippisley-Cox et al, 2014***[171] Observed vs. expected by decile of predicted risk. Hip 10 year: Observed and expected were similar in both men and women except for the highest decile of risk for which the model overestimated predicted risk. MOF 10 year: Observed risks generally agreed with predicted risk but were overestimated at the highest decile of risk. ***Hippisley-Cox et al, 2012***[147] O/E ratios for deciles of predicted risk very close to 1.0 except for the highest decile of risk for which predicted risks were overestimated for both men and women and for both MOF and hip fractures ***Dagan et al, 2017***[169] O/E ratio hip fracture across ages (note these were adjusted for 5-year risk predictions) Women: Range: 1.1 (highest age group) to 3.7 (lowest age group) Men: Range 0.9 (highest age group) to 3.5 (lowest age group) O/E ratio hip fracture across deciles of risk (note these were adjusted for 5-year risk predictions) Women: Range 1.0 (highest decile) to 3.5 (second lowest decile); all deciles except the highest were 2.3 or higher. Men: Range 0.9 (highest decile) to 5.6 (lowest decile); all deciles except the highest were 2.5 or higher. ***Davis et al, 2019***[173] O/E ratio: Hip 48 (3.94%) vs. 49.5 (4.06%)	***Hippisley-Cox et al, 2012*** [147] Calibration plots depicted in Figures 2 (MOF) and 3 (hip) of the article; men and women depicted separately. ***Dagan et al, 2017***[169] Calibration plots were presented in Figure 3 Slope, calibration in the large Women: 0.68; -0.49 Men: 0.60; -0.99 FRAX Women: 0.94; 0.39 Men: 0.94; 0.39 Garvan Women: 0.64; 0.18 Men: 0.68; -0.95 ***Gourlay et al, 2017***[146] Calibration plots were depicted in Figure 1 by deciles of risk. Risks at 5 lowest deciles of risk were overestimated, and risks at 5 highest deciles of risk were underestimated (0.0096 and 0.0001 in the 2 cohorts with and without BMD). The text reports the opposite findings; author query sent. No calibration plots for MOF. ***Davis et al, 2019***[173] Calibration plot presented as Figure 1 by deciles of predicted risk. No clear pattern; both under- and overprediction observed.	***Gourlay et al, 2017***[146] Hosmer-Lemeshow goodness of fit p-value, small p-value represents poor fit 0.0006 with BMD cohort; <0.0001 without BMD cohort

Appendix D Table 11. Calibration Outcomes for Predictive Accuracy of Risk Assessment Instruments (Key Question 2a)

Risk Assessment Instrument	Observed/Expected Ratio	Calibration Plots	Hosmer-Lemeshow Goodness of Fit
QFracture (2012 version)[146, 147, 169, 171, 173, 177, 311] (continued)	***Jain et al, 2023***[311] O/E Ratio (calibration in the large) FRAX: 0.97 (Overprediction in the highest quantile risk group for MOF and hip; otherwise, reasonable predictions across quantiles; underestimation in youngest age group 50-59) QFracture: 2.02 (underprediction in all quantiles for both MOF and Hip EXCEPT for the highest quantile group for hip which was overpredicted; also underprediction in youngest age group 50-59) By race: adjustment factors to convert MOF risk for White men and women were accurate for Black women, but underestimated risk in Black men, Hispanic women and men Risk was underestimated for Black women, but accurate for other groups.	***Desbiens et al, 2020***[177] Calibration plots are depicted in Figure 4 of the manuscript. QFracture globally underestimated risk across all risk levels. ***Hippisley-Cox et al, 2014***[171] Calibration plots presented in Figure 1 of manuscript	
QFracture[131] (Systematic review summary)	NR	NR	Hip: 1 study; P<0.0001 indicating poor calibration GRADE: very low certainty for poor performance MOF: 1 study; P<0.0001 indicating poor calibration GRADE: very low certainty for poor performance

Note: The Sun et al SR[137] did not synthesize calibration outcomes by instrument and are not included in this table. Authors of this SR summarized calibration findings as follows: **"**Calibration measurements were reported for 33 (24%) models, with 31 (22%) models showing good fitness. Calibration was assessed with calibration slope (n=18, 13%), the Hosmer-Lemeshow test (n=11, 8%), and the calibration intercept (n=4, 3%). Only 22 (16%) models used suitable methods (calibration slope or calibration intercept) for calibration calculation (Table 2)." (pg. 1229, Sun et al[137]).

Abbreviations: BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; HR=hazard ratio; MOF=major osteoporotic fracture; NR=not reported; O/E=observed/expected; OF=osteoporotic fracture; OST=Osteoporosis Self-Assessment Tool; PLE=product limit estimate; SD=standard deviation; vs.=versus; U.S.=United States.

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Adami et al, 2023[138] Fair	Hip	FRAX 10 year with BMD	Men and women	2 Participants: 46,300	AUC: NR Sn: 63% to 77% (95% CI, 55 to 81) Sp: 72% to 80% (95% CI, 72 to 81) 3% threshold
				2 Participants: 46,300	AUC: NR Sn: 43% to 66% (95% CI, 35 to 70) Sp: 83% to 89% (95% CI, 83 to 90) 5% threshold
Beaudoin et al, 2019[135] Good	Hip	FRAX 10 year with BMD	Men and women	17 (19) Participants: NR	AUC: 0.78 (0.75 to 0.81) Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	FRAX 10 year with BMD	Men and women	3 (3) Participants: 276,786	AUC: 0.77 (0.73 to 0.81) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	FRAX 10 year with BMD	Men and women	3 (3) Participants: NR	AUC: 0.80 to 0.83 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	MOF	FRAX 10 year with BMD	Men and women	2 Participants: 46,300	10% threshold Sn: 53% to 68% (95% CI, 49 to 70) Sp: 60% to 72% (95% CI, 60 to 73)
				2 Participants: 46,300	20% threshold Sn: 18% to 28% (95% CI, 15 to 30) Sp: 91% to 94% (95% CI, 90 to 94)
				2 Participants: 46,300	30% threshold Sn: 6% to 9% (95% CI, 4 to 11) Sp: 98% (95% CI, 98 to 99)
Beaudoin et al, 2019[135] Good	MOF	FRAX 10 year with BMD	Men and women	20 (25) Participants: NR	AUC: 0.67 (0.65 to 0.69) Sn: NR Sp: NR

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Marques et al, 2015[134] Good	MOF	FRAX 10 year with BMD	Men and women	3 (3) Participants: 276,786	AUC: 0.63 (0.60 to 0.66) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year with BMD	Men and women	3 (3) Participants: NR	AUC: 0.69 to 0.71 Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Other	FRAX 10 year with BMD	Men and women	6 (10) Participants: NR	AUC: 0.63 (0.62 to 0.65) Sn: NR Sp: NR Garvan-defined OP fractures
Adami et al, 2023[138] Fair	Hip	FRAX 10 year without BMD	Men and women	3 Participants: 43,494	3% threshold Sn: 26% to 78% (95% CI, 18 to 81) Sp: 64% to 90% (95% CI, 64 to 96)
				3 Participants: 43,494	5% threshold Sn: 22% to 65% (95% CI, 14 to 69) Sp: 87% to 97% (95% CI, 87 to 99)
Beaudoin et al, 2019[135] Good	Hip	FRAX 10 year without BMD	Men and women	23 (27) Participants: NR	AUC: 0.77 (0.73 to 0.80) Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	FRAX 10 year without BMD	Men and women	3 (3) Participants: 276,786	AUC: 0.67 (0.61 to 0.73) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	FRAX 10 year without BMD	Men and women	3 (3) Participants: NR	AUC: 0.77 to 0.83 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	MOF	FRAX 10 year Without BMD	Men and women	1 Participants: 194	3% threshold Sn: 52% (95% CI, 42 to 61) Sp: 69% (95% CI, 58 to 79)
				1 Participants: 194	5% threshold Sn: 35% (95% CI, 26 to 44) Sp: 81% (95% CI, 71 to 89)

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
				3 Participants: 46,494	10% threshold Sn: 24% to 65% (95% CI, 16 to 67) Sp: 59% to 93% (95% CI, 58 to 97)
				2 Participants: 46,300	20% threshold Sn: 16% to 29% (95% CI, 13 to 31) Sp: 88% to 93% (95% CI, 87 to 94)
				2 Participants: 46,300	30% threshold Sn: 4% to 10% (95% CI, 3 to 11) Sp: 97% to 99% (95% CI, 97 to 99)
Beaudoin et al, 2019[135] Good	MOF	FRAX 10 year without BMD	Men and women	22 (28) Participants: NR	AUC: 0.65 (0.63 to 0.67) Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	FRAX 10 year without BMD	Men and women	3 (3) Participants: 276,786	AUC: 0.61 (0.57 to 0.64) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year without BMD	Men and women	4 (4) Participants: NR	AUC: 0.66 to 0.71 Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Other	FRAX 10 year without BMD	Men and women	6 (11) Participants: NR	AUC: 0.60 (0.57 to 0.63) Sn: NR Sp: NR Garvan-defined OP fracture
Jiang et al, 2017[136] Good	Hip	FRAX 10 year NR	Men and women	6 (6) Participants: 50,944	AUC: NR Sn: 45.7% (95% CI, 24.9 to 68.1) Sp: 84.7% (95% CI, 76.4 to 90.4) Threshold of ≥3% hip Fx risk
Jiang et al, 2017[136] Good	MOF	FRAX 10 year NR	Men and women	7 (7) Participants: 57,027	AUC: NR Sn: 10.3% (95% CI, 3.8 to 25.1) Sp: 97.0% (95% CI, 91.2 to 99.0) Threshold of ≥20% MOF risk

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Adami et al, 2023[138] Fair	Hip	FRAX 10 year with BMD	Women	12 Participants: NR	AUC: 0.79 (95% CI, 0.74 to 0.83)
				4 Participants: 57,176	3% threshold Sn: 43% to 62% (95% CI, 28 to 64) Sp: 72% to 87% (95% CI, 69 to 89)
				3 Participants: 8,161	5% threshold Sn: 29% to 76% (95% CI, 19 to 72) Sp: 63% to 91% (95% CI, 61 to 94)
				1 Participants: 5,166	10% threshold Sn: 33% (95% CI, 28 to 39) Sp: 86% (95% CI, 85 to 87)
Marques et al, 2015[134] Good	Hip	FRAX 10 year with BMD	Women	5 (5) Participants: 115,611	AUC: 0.79 (0.73 to 0.85) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	FRAX 10 year with BMD	Women	12 (12) Participants:	AUC: 0.64 to 0.88 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	MOF	FRAX 10 year with BMD	Women	23 Participants: NR	AUC: 0.71 (95% CI, 0.68 to 0.74)
				1 Participants: 76	3% threshold Sn: 67% (95% CI, 30 to 93) Sp: 75% (95% CI, 63 to 84)
				1 Participants: 790	5% threshold Sn: 66% (95% CI, 57 to 73) Sp: 71% (95% CI, 67 to 74)
				5 Participants: 10,610	10% threshold Sn: 42% to 97% (95% CI, 28 to 98) Sp: 15% to 84% (95% CI, 14 to 88)

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
				4 Participants: 62,421	20% threshold Sn: 8% to 41% (95% CI, 2 to 44) Sp: 81% to 97% (95% CI, 80 to 98)
				1 Participants: 506	30% threshold Sn: Not estimable Sp: 99% (95% CI, 97 to 100)
Marques et al, 2015[134] Good	MOF	FRAX 10 year with BMD	Women	5 (5) Participants: 14,224	AUC: 0.67 (0.64 to 0.71) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year with BMD	Women	18 (18) Participants: NR	AUC: 0.61 to 0.78 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	Hip	FRAX 10 year without BMD	Women	17 Participants: NR	AUC: 0.74 (95% CI, 0.70 to 0.78)
				3 Participants: 881,295	3% threshold Sn: 8% to 77% (95% CI, 0 to 82) Sp: 39% to 100% (95% CI, 36 to 100)
				3 Participants: 886,490	5% threshold Sn: 42% to 78% (95% CI, 41 to 82) Sp: 50% to 92% (95% CI, 49 to 92)
Marques et al, 2015[134] Good	Hip	FRAX 10 year without BMD	Women	9 (9) Participants: 131,244	AUC: 0.74 (0.68 to 0.80) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	FRAX 10 year without BMD	Women	10 (10) Participants: NR	AUC: 0.64 to 0.90 Sn: NR Sp: NR
	MOF	FRAX 10 year without BMD	Women	19 Participants: NR	AUC: 0.67 (95% CI, 0.65 to 0.70)

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Adami et al, 2023[138] Fair				2 Participants: 1,191	3% threshold Sn: 57% to 85% (95% CI, 49 to 90) Sp: 34% to 79% (95% CI, 23 to 82)
				1 Participants: 908	5% threshold Sn: 34% (95% CI, 27 to 42) Sp: 89% (95% CI, 86 to 91)
				3 Participants: 7,963	10% threshold Sn: 46% to 100% (95% CI, 31 to 100) Sp: 0% to 77% (95% CI, 0 to 81)
				1 Participants: 506	20% threshold Sn: 8% (95% CI, 2 to 20) Sp: 95% (95% CI, 93 to 97)
				1 Participants: 506	30% threshold Sn: 4% (95% CI, 0 to 14) Sp: 99% (95% CI, 98 to 100)
Marques et al, 2015[134] Good	MOF	FRAX 10 year without BMD	Women	7 (7) Participants: 24,726	AUC: 0.65 (0.63 to 0.68) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year without BMD	Women	13 (13) Participants: NR	AUC: 0.58 to 0.75 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	Hip	FRAX 10 year with BMD	Men	3 Participants: NR	AUC: 0.80 (95% CI, 0.71 to 0.90) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	FRAX 10 year with BMD	Men	4 (4) Participants: NR	AUC: 0.72 to 0.77 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	Hip	FRAX 10 year without BMD	Men	6 Participants: NR	AUC: 0.76 (95% CI, 0.66 to 0.86) Sn: NR Sp: NR

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Sun et al, 2022[137] Good	Hip	FRAX 10 year without BMD	Men	1 (1) Participants: NR	AUC: 0.69 (NR) Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	FRAX 10 year without BMD	Men	2 (2) Participants: 11,199	AUC: 0.71 (0.65 to 0.77) Sn: NR Sp: NR
Adami et al, 2023[138] Fair	MOF	FRAX 10 year without BMD	Men	5 Participants: NR	AUC: 0.66 (95% CI, 0.57 to 0.76) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year without BMD	Men	2 (2) Participants:	AUC: 0.69 to 0.70 Sn: NR Sp: NR
Adami et al, 2023[138] Fair	MOF	FRAX 10 year with BMD	Men	5 Participants: NR	AUC: 0.73 (95% CI, 0.62 to 0.83) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	FRAX 10 year with BMD	Men	8 (8) Participants:	AUC: 0.64 to 0.85 Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	FRAX 10 year without BMD	Men	2 (2) Participants: 11,199	AUC: 0.63 (0.60 to 0.66) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Hip	FRC 10 year with BMD	Men and women	2 (2) Participants: NR	AUC: 0.82 (0.77 to 0.88) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	FRC 10 year with BMD	Men and women	1 (1) Participants: NR	AUC: 0.70 (0.67 to 0.73) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Hip	FRC 10 year without BMD	Men and women	2 (2) Participants: NR	AUC: 0.77 (0.65 to 0.89) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	FRC 10 year without BMD	Men and women	1 (1) Participants: NR	AUC: 0.66 (0.63 to 0.69) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Hip	Garvan 10 year with BMD	Men and women	5 (7) Participants: NR	AUC: 0.76 (0.71 to 0.80) Sn: NR Sp: NR

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Beaudoin et al, 2019[135] Good	Other	Garvan 10 year with BMD	Men and women	6 (8) Participants: NR	AUC: 0.72 (0.66 to 0.79) Sn: NR Sp: NR Garvan-defined OP fractures
Beaudoin et al, 2019[135] Good	Hip	Garvan 10 year without BMD	Men and women	2 (3) Participants: NR	AUC: 0.70 (0.64 to 0.76) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	Garvan 10 year with BMD	Women	1 (1) Participants: NR	AUC: 0.70 (0.65 to 0.75) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good[135]	Other	Garvan 10 year without BMD	Men and women	3 (4) Participants: NR	AUC: 0.67 (0.59 to 0.74) Sn: NR Sp: NR Garvan-defined OP fracture
Sun et al, 2022[137] Good[135]	Hip	Garvan 10 year NR	Men and women	1 (1) Participants: NR	AUC: 0.78 Sn: NR Sp: MR
Beaudoin et al, 2019[135] Good	MOF	Garvan 10 year without BMD	Women	1 (1) Participants: NR	AUC: 0.66 (0.61 to 0.72) Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	Garvan 10 year with BMD	Women	2 (2) Participants: 5,574	AUC: 0.74 (0.61 to 0.87) Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	Garvan 10 year with BMD	Women	3 (3) Participants: 6,932	AUC: 0.70 (0.64 to 0.75) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	Garvan 10 year NR	Women	7 (7) Participants: NR	AUC: 0.57 to 0.70 Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	Garvan 10 year NR	Women	3 (3) Participants: NR	AUC: 0.57 to 0.80 Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	Garvan 10 year with BMD	Men	2 (2) Participants: 5,010	AUC: 0.73 (0.68 to 0.78) Sn: NR Sp: NR
Sun et al, 2022[137] Good Sun et al, 2022[137]	MOF	Garvan 10 year NR	Men	3 (3) Participants: NR	AUC: 0.57 to 0.69 Sn: NR Sp: NR

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Sun et al, 2022[137] Good	Hip	Garvan 10 year NR	Men	1 (1) Participants: NR	AUC: 0.85 Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	QFracture 10 year	Men and women	1 (1) Participants: NR	AUC: 0.71 Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	QFracture 10 year	Men and women	1 (1) Participants: NR	AUC: 0.88 Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	QFracture 10 year	Women	3 (3) Participants: 1,779,154	AUC: 0.89 (0.88 to 0.89) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	QFracture 10 year	Women	3 (3) Participants: NR	AUC: 0.89 Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	QFracture 10 year	Women	3 (3) Participants: 1,778,570	AUC: 0.81 (0.78 to 0.834) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	QFracture 10 year	Women	3 (3) Participants: NR	AUC: 0.79 to 0.82 Sn: NR Sp: NR
Marques et al, 2015[134] Good	Hip	QFracture 10 year	Men	2 (2) Participants: 1,741,983	AUC: 0.87 (0.86 to 0.88) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	QFracture 10 year	Men	3 Participants: NR	AUC: 0.86 to 0.88 Sn: NR Sp: NR
Marques et al, 2015[134] Good	MOF	QFracture 10 year	Men	2 (2) Participants: 1,741,983	AUC: 0.72 (0.67 to 0.76) Sn: NR Sp: NR
Sun et al, 2022[137] Good	MOF	QFracture 10 year	Men	3 (3) Participants: NR	AUC: 0.69 to 0.74 Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	ORAI without BMD	Women	1 (1) Participants: NR	AUC: 0.71 (0.68 to 0.75) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	OSIRIS without BMD	Women	1 (1) Participants: NR	AUC: 0.70 (0.66 to 0.74) Sn: NR Sp: NR

Appendix D Table 12. Discrimination Outcomes from Included Systematic Reviews for Predictive Accuracy of Risk Assessment Instruments for Fracture (Key Question 2a)

Author, Year Study Quality	Fracture Type	Instrument	Sex	Number of Studies (Number of Comparisons) Number of Participants	Results (95% CI)
Beaudoin et al, 2019[135] Good	MOF	OST without BMD	Women	2 (2) Participants: NR	AUC: 0.63 (0.49 to 0.77) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Hip	QFracture 2009 10 year without BMD	Men and women	2 (4) Participants: NR	AUC: 0.88 (0.86 to 0.89) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	QFracture 2009 10 year without BMD	Men and women	2 (4) Participants: NR	AUC: 0.79 (0.75 to 0.82) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	MOF	SCORE without BMD	Women	1 (1) Participants: NR	AUC: 0.71 (0.68 to 0.75) Sn: NR Sp: NR
Beaudoin et al, 2019[135] Good	Hip	WHI 5 year without BMD	Women	2 (2) Participants: NR	AUC: 0.81 (0.78 to 0.84) Sn: NR Sp: NR
Sun et al, 2022[137] Good	Hip	WHI	Women	1 (1) Participants: NR	AUC: 0.82 Sn: 0.69 Sp: 0.80

Abbreviations: AUC=area under the curve; BMD=body mass index; CI=confidence interval; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; Fx=fracture; MOF=major osteoporotic fracture; NR=not reported; OP=osteoporosis; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=Osteoporosis Index of Risk; OST=Osteoporosis Self-Assessment Tool; SCORE=Simple Calculated Osteoporosis Risk Estimation; Sn=sensitivity; Sp=specificity; WHI=Women's Health Initiative.

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Baleanu et al, 2021[179] Fracture Risk Brussels Epidemiological Enquiry (FRISBEE)	**Hip:** Nontraumatic hip fracture validated by written medical reports N (%): 47 (1.5) Length of followup: 5 years **MOF:** Nontraumatic fractures of hip, clinical spine, forearm, shoulder validated with written medical reports **N (%):** 281 (9.3) Length of followup: 5 years **Osteoporotic Fractures:** Nontraumatic fractures at any location excluding digits verified by written medical reports. **N (%):** 356 (11.7) Length of followup: 5 years	BMD site/fracture/length of followup NR/Hip/5 years AUC: 0.81 (95% CI, 0.76 to 0.86) BMD site/fracture/length of followup NR/MOF/5 years AUC: 0.69 (95% CI, 0.65 to 0.72) BMD site/fracture/length of followup NR/Garvan-defined OF/5 years AUC: 0.68 (95% CI, 0.65 to 0.71)	NR
Black et al, 2018[192] Study of Osteoporotic Fractures (SOF)	**Hip:** Hip fractures excluding traumatic fractures **N (%):** 1,290 (15.9%) Length of followup: 25 years **Nonvertebral:** Nonvertebral fractures excluding traumatic fractures **N (%):** 3,267 (43.7) Length of followup: 20 years	NR	Hip Fx incidence over 25 years' followup Lowest BMD quartile: 29.6% Highest BMD quartile: 7.6% Nonvertebral incidence over 20 years followup Lowest BMD quartile: 59.7% Highest BMD quartile: 32.9%

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Bolland et al, 2011[166]	**MOF:** Shoulder, hip, forearm, clinical vertebral resulting from minimal trauma **N (%):** 279 (16.1%) Length of followup: 8.8 years (mean) **Garvan OF:** Hip, clinical vertebral, forearm, metacarpal, humerus, scapula, clavicle, distal femur, proximal tibia, patella, pelvis, sternum resulting from minimal trauma (Garvan definition) **N (%):** 229 (19.6%) Length of followup: 8.8 years (mean) **Hip:** Hip fractures N (%): 57 (4%) Length of followup: 8.8 years (mean) **Fragility:** Fracture from a fall at a standing height or less N (%): NR Length of followup: 8.8 years (mean)	BMD site/fracture/length of followup FN/Hip Fx/8.8 years AUC: 0.64 (95% CI, 0.57 to 0.72) BMD site/fracture/length of followup FN/Fragility/8.8 years AUC: 0.59 (95% CI, 0.56 to 0.62) BMD site/fracture/length of followup FN/MOF/8.8 years AUC: 0.60 (95% CI, 0.56 to 0.64) BMD site/fracture/length of followup FN/Garvan OF/8.8 years AUC: 0.60 (95% CI, 0.56 to 0.64)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Chapurlat et al, 2020[175] OFELY and QUALYOR Cohorts (2 population based cohorts in France)	**Vertebral and nonvertebral:** Clinical fractures excluding head, toes, and fingers **N (%):** 126 (Cannot determine) Length of followup: 4 years Subgroup: QUALYOR **MOF:** Hip, clinical vertebral, humerus, forearm **N (%):** 61 (Cannot determine) Length of followup: 4 years Subgroup: QUALYOR **Vertebral and Nonvertebral:** Clinical fractures excluding head, toes, and fingers **N (%):** 106 (Cannot determine) Length of followup: 8 years Subgroup: OFELY **MOF:** Hip, clinical vertebral, humerus, forearm **N (%):** 65 Cannot determine Length of followup: 8 years Subgroup: OFELY	BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Vertebral and NV/4 years AUC: 0.581 (95% CI, 0.54 to 0.62) Sensitivity: 14.2% (95% CI, NR) Specificity: 95.4% (95% CI, NR) BMD site/fracture/length of followup FN/MOF/4 years AUC: 0.617 (95% CI, 0.56 to 0.68) Sensitivity: 22.4% (95% CI, NR) Specificity: 95.4% (95% CI, NR) BMD site/fracture/length of followup FN/Vertebral and NV/4 years AUC: NR Sensitivity: 21.9% (95% CI, NR) Specificity: 94.3% (95% CI, NR) Age 70 or older subgroup BMD site/fracture/length of followup FN/MOF/4 years AUC: NR Sensitivity: 25.0% (95% CI, NR) Specificity: 93.8% (95% CI, NR) Age 70 or older subgroup	NR
Cheung et al, 2012[150] Hong Kong Osteoporosis Study	**MOF:** Wrist, clinical spine, hip, or humerus **N (%):** 106 (4.7) Length of followup: 4.5 (2.8) years Hip fracture: NR **N (%):** 21 (0.9) Length of followup: 4.5 (2.8) years Vertebral fracture: Clinical **N (%):** 43 (1.9) Length of followup: 4.5 (2.8) years	BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/MOF/4.5 (2.8) years AUC: 0.711 (95% CI, 0.66 to 0.76) Sensitivity: 45.3% Specificity: NR Sn calculated, unable to calculate Sp based on data provided in study. BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Hip fracture/4.5 (2.8) years AUC: 0.855 (95% CI, 0.791 to 0.919) Sensitivity: 66.7% (95% CI, NR) Specificity: NR Sn calculated, Sp could not be calculated from data provided in study	Figure 1 and Figure 2 of study report depicted proportion of participants who sustained MOF and hip fractures, respectively, by quartile of predicted risk. A dose-response effect is observed with participants in the 4th quartile having the highest observed risk and participants in the first quartile having the lowest predicted risk.

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Fraser et al, 2011[154] Canadian Multicentre Osteoporosis Study (CaMos)	**MOF:** Hip, clinical spine, humerus, forearm/wrist **N (%):** 573 (12.0) Length of followup: 10 years **Subgroup:** Women **MOF:** Hip, clinical spine, humerus, forearm/wrist **N (%):** 122 (6.4) Length of followup: 10 years **Subgroup:** Men **Hip:** Hip fracture **N (%):** 129 (2.7) Length of followup: 10 years **Subgroup:** Women **Hip:** Hip fracture **N (%):** 46 (2.4) Length of followup: 10 years Subgroup: Men	BMD Site/fracture/length of followup FN/MOF/10 years AUC: 0.66 (95% CI, 0.64 to 0.69) BMD site/fracture/length of followup FN/Hip fracture/10 years AUC: 0.76 (95% CI, 0.72 to 0.79) BMD site/fracture/length of followup FN or LS/MOF/10 years AUC: 0.67 (95% CI, 0.65 to 0.70) BMD site/fracture/length of followup FN or LS/Hip/10 years AUC: 0.75 (95% CI, 0.71 to 0.78)	NR
Goldshtein et al, 2018[172] Maccabi Healthcare Services	**MOF:** Hip, clinical vertebral, proximal humerus, distal forearm **N (%):** 2,263 (13.7%) Length of followup: 10 years Hip: Hip fracture **N (%):** 481 (2.9%) Length of followup: 10 years	BMD site/fracture/length of followup FN/MOF/10 years AUC: 0.62 (95% CI, 0.59 to 0.64) BMD site/fracture/length of followup FN/Hip/10 years AUC: 0.78 (95% CI, 0.74 to 0.83)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Gourlay et al, 2017[146] MrOs	**Hip:** Incident hip fracture **N (%):** 175 (3.5%) Length of followup: 15.8 years **Subgroup:** Among those in the BMD analysis **MOF:** Incident MOF (clinical spine, hip, forearm, shoulder) **N (%):** 326 (6.6%) Length of followup: 15.8 years **Subgroup:** Among those in the BMD analysis **Hip:** Incident hip fracture **N (%):** 218 (4.2%) Length of followup: 15.8 years **Subgroup:** Among those in the without BMD analysis **MOF:** Incident MOF (clinical spine, hip, forearm, shoulder) **N (%):** 387 (7.4%) Length of followup: 15.8 years **Subgroup:** Among those in the without BMD analysis	BMD site/fracture/length of followup FN/-0.36/Hip/15.8 years AUC: 0.76 (95% CI, 0.72 to 0.81) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 43% (95% CI, 41% to 44%) Threshold chosen to be equivalent to 90% sensitivity BMD site/fracture/length of followup FN/-0.21/MOF/15.8 years AUC: 0.76 (95% CI, 0.71 to 0.80) Sensitivity: 90% (95% CI, 85% to 95%) Specificity: 38% (95% CI, 37% to 39%) Threshold chosen to be equivalent to 90% sensitivity	Authors reported that model with continuous FN BMD T-score showed good calibration Hosmer-Lemeshow Goodness of Fit Test Hip Fracture: p=0.2655 MOF: p=0.1672
Iki et al, 2021[194] Japanese Population-based Osteoporosis Study	**Hip:** Hip fractures **N (%):** 68 (5.1%) **Length of followup:** Median 19.8 years	BMD site/fracture/length of followup FN/Hip/19.8 years AUC: 0.858 (95% CI, NR) BMD site/fracture/length of followup TH/Hip/10 years AUC: 0.869 (95% CI, NR)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Kwok et al, 2012[182] MrOs (Hong Kong)	**Nonvertebral:** Fragility fracture at site other than spine confirmed by X-ray or medical record reports **N (%):** 107 (5.6%) **Length of followup:** Mean 6.5 (1.7) years **Hip:** Fragility hip fracture confirmed by X-ray or medical record reports **N (%):** 28 (1.5%) **Length of followup:** Mean 6.5 (1.7) years **MOF:** Major fragility fractures **N (%):** 713.7% **Length of followup:** Mean 6.5 (1.7) years	NR	NR
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry	From Leslie et al, 2010[157] **Hip:** Hip fracture **N (%):** 506 (2.7%) (95% CI, 2.1% to 3.4%) [Kaplan-Meier estimate]) Length of followup: 10 years Subgroup: Female **Hip:** Hip fracture **N (%):** 43 (3.5%) (95% CI, 0.8% to 6.2%) [Kaplan-Meier estimate]) Length of followup: 10 years Subgroup: Male	From Leslie et al, 2010[157] BMD site/fracture/length of followup FN/Hip/10 years AUC: 0.801 (95% CI, 0.783 to 0.819) BMD site/fracture/length of followup FN/MOF/10 years AUC: 0.679 (95% CI, 0.668 to 0.690) BMD site/fracture/length of followup FN or TH or LS/Hip/10 years AUC: 0.770 (95% CI, 0.753 to 0.787)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry (continued)	**MOF:** Hip, clinical vertebral, forearm, or humerus **N (%):** 2,380 (12.1%) (95% CI, 10.8% to 13.4%) [Kaplan-Meier Estimate] Length of followup: 10 years Subgroup: Female **MOF:** Hip, clinical vertebral, forearm, or humerus **N (%):** 163 (10.7%) (95% CI, 6.6% to 14.9%) [Kaplan-Meier Estimate] Length of followup: 10 years Subgroup: Male From Hans et al, 2011[183] and Leslie et al, 2013[184] **Vertebral:** Clinical vertebral fracture **N (%):** 439 (1.5%) Length of followup: 4.7 (2.2) years **Hip:** Hip fracture **N (%):** 293 (1.0%) Length of followup: 4.7 (2.2) years **MOF:** Any MOF **N (%):** 1,668 (5.7%) Length of followup: 4.7 (2.2) years	BMD site/fracture/length of followup FN or TH or LS/MOF/10 years AUC: 0.675 (95% CI, 0.665 to 0.686) From Hans et al, 2011[183] and Leslie et al, 2013[184] BMD site/fracture/length of followup TH/Vertebral/4.7 years AUC: 0.71 (95% CI, 0.68 to 0.73) BMD site/fracture/length of followup FN/Vertebral/4.7 years AUC: 0.71 (95% CI, 0.68 to 0.73) BMD site/fracture/length of followup LS/Vertebral/4.7 years AUC: 0.69 (95% CI, 0.67 to 0.72) BMD site/fracture/length of followup TH/Hip/4.7 years AUC: 0.81 (95% CI, 0.79 to 0.83) BMD site/fracture/length of followup FN/Hip/4.7 years AUC: 0.80 (95% CI, 0.77 to 0.82) BMD site/fracture/length of followup LS/Hip/4.7 years AUC: 0.62 (95% CI, 0.62 to 0.69) BMD site/fracture/length of followup TH/MOF/4.7 years AUC: 0.68 (95% CI, 0.66 to 0.69)	

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry (continued)	From Leslie et al, 2016[161] **MOF:** Nontraumatic hip, clinical vertebral, forearm, humerus fracture **N (%):** 3,905 (11.5%) Length of followup: Mean 9.8 years From Leslie et al, 2018[162] **MOF:** Nontraumatic hip, clinical vertebral, forearm, humerus **N (%):** 5,345 (8.6%) **Length of followup:** Mean 7.2 (SD 4.2) years Subgroup: Women **MOF:** Nontraumatic hip, clinical vertebral, forearm, humerus **N (%):** 405 (6.3%) **Length of followup:** Mean 5.4 (3.9) years Subgroup: Men **Hip:** Hip fracture **N (%):** 1,471 (2.4%) **Length of followup:** Mean 7.2 (SD 4.2) years Subgroup: Women **Hip:** Hip fracture **N (%):** 108 (1.7%) **Length of followup:** Mean 5.4 (3.9) years Subgroup: Men From Crandall et al, 2019[160] **MOF:** Based on claims data; humerus, hip, clinical vertebral, forearm **N (%):** 6,208 (11.4%) Length of followup: 10.5 years	BMD site/fracture/length of followup FN/MOF/4.7 years AUC: 0.68 (95% CI, 0.66 to 0.69) BMD site/fracture/length of followup LS/MOF/4.7 years AUC: 0.64 (95% CI, 0.63 to 0.66) From Leslie et al, 2016[161] BMD site/cutoff/fracture/length of followup FN or TH or LS/T-score <-2.5/MOF/9.8 years AUC: NR Sensitivity: 51.3% (95% CI, NR) Specificity: 70.9% (95% CI, NR) From Leslie et al, 2018[162] BMD site/cutoff/fracture/length of followup FN or TH or LS/T-score <-2.5/MOF/7.2 years AUC: NR Sensitivity: 28.0% (95% CI, NR) Specificity: 89.3% (95% CI, NR) Women BMD site/cutoff/fracture/length of followup FN or TH or LS <-2.5/Hip/7.2 years AUC: NR Sensitivity: 43.0% (95% CI, NR) Specificity: 88.6% (95% CI, NR) Women BMD site/fracture/length of followup FN or TH or LS/MOF/5.4 years AUC: NR Sensitivity: 17.5% (95% CI, NR) Specificity: 92.2% (95% CI, NR) Men	

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry (continued)	**Hip:** Based on claims data **N (%):** 1,906 (3.5%) Length of followup: 10.5 years From Agarwal et al, 2022[193] **OF:** Based on claims data, any nontraumatic fracture excluding craniofacial, hand, foot, and ankle **N (%):** Women 681 (4.1); men 140 (0.9) **Length of followup:** mean 2.6 years (SD 1.6) **Hip:** Nontraumatic hip fractures based on claims data **N (%):** Women 119 (0.7); men 22 (0.8) **Length of followup:** Mean 2.6 years (SD 1.6)	BMD site/fracture/length of followup FN or TH or LS/Hip/5.4 years AUC: NR Sensitivity: 30.6% (95% CI, NR) Specificity: 92.0% (95% CI, NR) Men From Crandall et al, 2019[160] BMD Site/cutoff/fracture/length of followup FN/T-score <-2.5/MOF/10.5 years AUC: NR At T-score Sensitivity: 25.7% (95% CI, NR) Specificity: 89.5% (95% CI, NR) Ages 40–49 (n=5,324): Sn 6.7%, Sp 98.0% Ages 50–59 (n=15,466): Sn 9.7%, Sp 96.2% Ages 60–69 (n=16,026): Sn 18.5%, Sp 91.6% Ages 70–79 (n=12,492): Sn 30.1%, Sp 82.0% Ages 80+ (n=5,151): Sn 49.0%, Sp 67.5% BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Hip/10.5 years AUC: NR Sensitivity: 38.1% (95% CI, NR) Specificity: 88.8% (95% CI, NR) Ages 40–49 (n=5,324): Sn 19.4%, Sp 97.8% Ages 50–59 (n=15,466): Sn 20.0%, Sp 95.9% Ages 60–69 (n=16,026): Sn 28.9%, Sp 91.0% Ages 70–79 (n=12,492): Sn 36.0%, Sp 81.1% Ages 80+ (n=5,151): Sn 53.6%, Sp 66.1%	
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193] Manitoba BMD Registry (continued)		From Agarwal et al, 2022[193] BMD site/cutoff/fracture/length of followup FN/NA/OF/2.6 years AUC: 0.61 (95% CI, 0.56 to 0.66) Sensitivity: NR Specificity: NR BMD site/cutoff/fracture/length of followup FN/NA/Hip/2.6 years AUC: 0.79 (95% CI, 0.71 to 0.88) Sensitivity: NR Specificity: NR	

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Marques et al, 2017[176] SAOL, IPR, and EPIPorto (3 Portuguese Cohorts)	**MOF:** Hip, wrist, shoulder, clinical vertebral (regardless of degree of trauma) **N (%):** 145 (7.5%) **Length of followup:** Mean (SD) 9.12 (1.5) years Subgroup: Women **MOF:** Hip, wrist, shoulder, clinical vertebral (regardless of degree of trauma) **N (%):** 33 (4.8%) **Length of followup:** Mean (SD) 9.12 (1.5) years Subgroup: Men **Hip:** Hip fracture **N (%):** 20 (1.0%) **Length of followup:** Mean (SD) 9.12 (1.5) years Subgroup: Women **Hip:** Hip fracture **N (%):** 8 (1.2%) **Length of followup:** Mean (SD) 9.12 (1.5) years Subgroup: Men	BMD site/fracture/length of followup FN/Hip/9.8 years AUC: 0.68 (95% CI, 0.66 to 0.71) Women BMD site/fracture/length of followup FN/Hip/9.8 years AUC: 0.82 (95% CI, 0.78 to 0.86) Men BMD site/fracture/length of followup FN/MOF/9.8 years AUC: 0.66 (95% CI, 0.63 to 0.68) Women BMD site/fracture/length of followup FN/MOF/9.8 years AUC: 0.80 (95% CI, 0.76 to 0.84) Men	NR
Nguyen et al, 2004[186] Dubbo Osteoporosis Epidemiology Study (DOES)	**Fragility:** Any symptomatic fractures resulting from minimal or no trauma **N (%):** 77 (14%) Length of followup: NR	BMD site/fracture/length of followup LS/Fragility/NR AUC: 0.77 (95% CI, NR) BMD site/fracture/length of followup FN/Fragility/NR AUC: 0.76 (95% CI, NR)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Prince et al, 2019[191] Perth Longitudinal Study of Aging in Women (PLSAW)	**Vertebral:** Clinical vertebral fracture **N (%):** 73 (6.7%) Length of followup: 14.5 years **Hip:** Hip fracture hospitalization **N (%):** 121 (11.2%) Length of followup: 14.5 years **Serious fragility fracture:** Low-trauma fracture hospitalization **N (%):** 305 (28.1%) Length of followup: 14.5 years	BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Vertebral/14.5 years AUC: NR Sensitivity: 19.7% (95% CI, 11.2% to 30.9%) Specificity: 92.4% (95% CI, 90.6% to 94.0%) BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Low-trauma fracture hospitalization/14.5 years AUC: NR Sensitivity: 13.2% (95% CI, 9.6% to 17.6%) Specificity: 93.4% (95% CI, 91.4% to 95.1%)	NR
Robbins et al, 2007[187] Women's Health Initiative	**Hip fracture:** Incident hip fracture confirmed with records **N (%):** 80 (0.7%) Length of followup: 8.7 years	BMD site/cutoff/fracture/length of followup NR/T-score <-2.5/Hip/8.7 years AUC: 0.79 (95% CI, 0.73 to 0,85) Sensitivity: 25.0% (95% CI, NR) Specificity: 95.3% (95% CI, NR)	NR
Sornay-Rendu et al, 2010[188] Os des Femmes de Lyon (OFELY) cohort	**Fragility:** Fractures at any site resulting from minimal trauma excluding fingers, toes, skull, and face **N (%):** 116 (13.4%) Length of followup: 10 years **Hip:** Hip fracture **N (%):** 17 (2.0%) Length of followup: 10 years **Vertebral:** Clinical vertebral fracture **N (%):** 25 (2.9%) Length of followup: 10 years **Forearm:** Forearm fracture **N (%):** 44 (5.1%) Length of followup: 10 years	BMD site/fracture/length of followup FN/Fragility fractures/10 years AUC: 0.74 (95% CI, 0.71 to 0.77)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Stewart et al, 2006[190] Aberdeen Prospective Osteoporosis Screening Study (APOSS)	**Confirmed fractures:** Fracture at any site confirmed by X-ray or primary care physician **N (%):** 325 (8.4%) Length of followup: 3 to 12 years **MOF:** Hip, vertebral, wrist, and humerus fractures **N (%):** 128 (3.3%) Length of followup: 3 to 12 years	BMD site/fracture/length of followup LS/MOF/3 to 12 years AUC: 0.66 (95% CI, 0.64 to 0.68) BMD site/fracture/length of followup FN/MOF/3 to 12 years AUC: 0.64 (95% CI, 0.63 to 0.66)	NR
Sund et al, 2014[185] Kuopio Osteoporosis Risk Factor and Prevention (OSTPRE)	**Hip:** Hip fractures validated with medical records **N (%):** 21 (0.76) Length of followup: 10 years	BMD site/fracture/length of followup FN/Hip Fx/10 years AUC: 0.739 (95% CI, 0.644 to 0.834)	O/E ratios across risk quintiles for hip fracture Quintile 1: 3/6.3 Quintile 2: 10/9.4 Quintile 3: 18/12.8 Quintile 4: 17/18.2 Quintile 5: 26/42.3 All: 74/88.9 O/E ratio: 0.83, 95% CI, 0.65 to 1.04 Hosmer-Lemeshow goodness of fit p=0.015

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Tamaki et al, 2011[151] Japanese Population-Based Osteoporosis Study (JPOS)	**Hip:** Hip fracture **N (%):** 4 (0.5%) Length of followup: 10 years **Vertebral:** Clinical vertebral fracture **N (%):** 13 (1.6%) Length of followup: 10 years **Distal forearm:** Distal forearm fracture **N (%):** 25 (3.1%) Length of followup: 10 years **Proximal humerus:** Proximal humerus fracture **N (%):** 1 (0.1%) Length of followup: 10 years **MOF**: Major osteoporotic fractures **N (%):** 43 (5.3%) Length of followup: 10 years	BMD site/fracture/length of followup FN/MOF/10 years AUC: 0.64 (95% CI, 0.57 to 0.72) BMD site/fracture/length of followup FN/Hip fracture/10 years AUC: 0.82 (95% CI, 0.67 to 0.98)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Tanaka et al, 2010[156] Miyama and Taiji Cohorts	**MOF:** Clinical vertebral, proximal humerus, distal forearm **N (%):** 60 (15%) Length of followup: 10 years **Subgroup:** Study reported number of fractures, not number of persons with a fracture. **Vertebral:** Clinical vertebral fractures **N (%):** 44 (12%) Length of followup: 10 years **Subgroup:** Study reported number of fractures, not number of persons with a fracture. **Hip:** Hip fractures **N (%):** 8 (2.0%) Length of followup: 10 years **Subgroup:** Study reported number of fractures, not number of persons with a fracture.	BMD site/fracture/length of followup FN/MOF/10 years AUC: 0.651 (95% CI, 0.575 to 0.728)	NR

Appendix D Table 13. Outcomes from Included Studies for Predictive Accuracy of Bone Mineral Density Alone for Fracture (Key Question 2b)

Author, Year Study Cohort Name	Fracture Type and Definition Frequency (%) Length of Followup Subgroup (If Applicable)	Accuracy Results	Calibration Outcomes
Trajanoska et al, 2018[15] Rotterdam Study	Hip: **N (%):** 133 (2.8) **Length of followup:** mean 10.7 (6.2) years Subgroup: Men Nonvertebral: **N (%):** 586 (12.3) **Length of followup:** mean 10.7 (6.2) years Subgroup: Men Hip: **N (%):** 431 (6.9) **Length of followup:** mean 10.7 (6.2) years Subgroup: Women Nonvertebral: **N (%):** 1,647 (26.2) **Length of followup:** mean 10.7 (6.2) years Subgroup: Women	BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Hip Fx/10.7 years AUC: NR Sensitivity: 29% (95% CI, NR) Specificity: 94% (95% CI, NR) Men BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Hip Fx/10.7 years AUC: NR Sensitivity: 38% (95% CI, NR) Specificity: 91% (95% CI, NR) Women BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Nonvertebral Fx/10.7 years AUC: NR Sensitivity: 12% (95% CI, NR) Specificity: 94% (95% CI, NR) Men BMD site/cutoff/fracture/length of followup FN/T-score <-2.5/Nonvertebral/10.7 years AUC: NR Sensitivity: 38% (95% CI, NR) Specificity: 91% (95% CI, NR) Women	NR
Tremollieres et al, 2010[189] Menopause et Os (MENOS) Study	**MOF:** spine, vertebral, hip, distal forearm, and humerus **N (%):** 145 (6.6%) Length of followup: 13.4 years	BMD site/fracture/length of followup Hip/MOF/13.4 years AUC: 0.66 (95% CI, 0.60 to 0.73)	NR

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CaMos=Canadian Multicentre Osteoporosis Study; CI=confidence interval; FN=femoral neck; Fx=fracture; HR=hazard ratio; LS=lumbar spine; MOF=major osteoporotic fracture; N=number; NR=not reported; NV=nonvertebral; OF=osteoporotic fracture; OR=odds ratio; QUALYOR=QUalité Osseuse LYon Orléans; SD=standard deviation; Sn=sensitivity; Sp=specificity; TH=total hip.

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Adler et al, 2003[221] U.S. Fair	Index Test/Cutoff/BMD Site OST/<2/FN or TH or LS AUC: 0.836 (95% CI, 0.747 to 0.924) Sn: 82% (95% CI, NR) Sp: 74% (95% CI, NR) AUC by race, by age White: 0.848 (95% CI, NR) Black: 0.800 (95% CI, NR) 50–59 y: 0.938 (95% CI, NR) 60–69 y: 0.894 (95% CI, NR) 70–79 y: 0.696 (95% CI, NR) ≥80 y: 0.993 (95% CI, NR) **Index Test/Cutoff/BMD Site** OST/<1/FN or TH or LS AUC: NR Sn: 93% (95% CI, NR) Sp: 66% (95% CI, NR) Index Test/Cutoff/BMD Site OST/<3/FN or TH or LS AUC: NR Sn: 75% (95% CI, NR) Sp: 80% (95% CI, NR) Index Test/Cutoff/BMD Site OST/NR/FN AUC: 0.814 (95% CI, 0.717 to 0.910) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OST/NR/LS AUC: 0.845 (95% CI, 0.731 to 0.960) Sn: NR Sp: NR

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Bansal et al, 2015[220] Pecina et al, 2016[234] U.S. Fair	From Bansal et al, 2015[220] Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥9.3%/LS or FN AUC: 0.58 Sn: 37% (95% CI, NR) Sp: 74% (95% CI, NR) (Sp and Sn also reported for FRAX MOF risk ≥5.5%) From Pecina et al, 2016[234] Index Test/Cutoff/BMD Site ORAI/≥9/FN or LS AUC: 0.60 (95% CI, NR) Sn: 52% (95% CI, 37% to 66%) Sp: 67% (95% CI, 61% to 73%) Index Test/Cutoff/BMD Site OST/<2/FN or LS AUC: 0.63 (95% CI, NR) Sn: 56% (95% CI, 41% to 69%) Sp: 69% (95% CI, 63% to 75%) Index Test/Cutoff/BMD Site SCORE/≥6/FN or LS AUC: 0.58 (95% CI, NR) Sn: 74% (95% CI, 59% to 84%) Sp: 42% (95% CI, 36% to 49%) Index Test/Cutoff/BMD Site FRAX without BMD/≥9.3%/FN or LS AUC: 0.55 (95% CI, NR) Sn: 36% (95% CI, 23% to 50%) Sp: 73% (95% CI, 67% to 79%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Brenneman et al, 2003[222] OPRA U.S. Fair	Index Test/Cutoff/BMD Site SCORE/≥7/FN or TH or LS AUC: 0.73 (95% CI, NR) Sn: 93.7% (95% CI, 88.3% to 99.1%) Sp: 23.8% (95% CI, 9.6% to 38.0%) Index Test/Cutoff/BMD Site SOF/≥5/FN or TH or LS AUC: 0.54 (95% CI, NR) Sn: 32.6% (95% CI, 26.6% to 38.6%) Sp: 76.0% (95% CI, 63.5% to 88.6%)
Cadarette et al, 2004[223] Canada Fair	Index Test/Cutoff/BMD Site ORAI/>8/FN or LS AUC: 0.802 (95% CI, NR) Sn: 92.5% (95% CI, 85.6% to 96.7%) Sp: 38.7% (95% CI, 34.5% to 42.9%) (AUC SE=0.02) Index Test/Cutoff/BMD Site OST/<2/FN or LS AUC: 0.733 (95% CI, NR) Sn: 95.3% (95% CI, 89.3% to 98.5%) Sp: 39.6% (95% CI, 35.4% to 43.9%) (AUC SE=0.02)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Cadarette et al, 2001[206] Canadian Multicentre Osteoporosis Study (CaMOS) Canada Fair	Index Test/Cutoff/BMD Site ABONE/≥2/FN AUC: 0.72 (95% CI, NR) Sn: 83.3% (95% CI, 78.5% to 88.0%) Sp: 47.7% (95% CI, 45.6% to 49.8%) **Index Test/Cutoff/BMD Site** NOF ≥1/FN AUC: 0.70 (95% CI, NR) Sn: 96.2% (95% CI, 93.8% to 98.6%) Sp: 17.8% (95% CI, 16.2% to 19.4%) **Index Test/Cutoff/BMD Site** ORAI/≥9/FN AUC: 0.79 (95% CI, NR) Sn: 97.5% (95% CI, 95.5% to 99.5%) Sp: 27.8% (95% CI, 25.9% to 29.7%) **Index Test/Cutoff/BMD Site** SCORE/≥6/FN AUC: 0.80 (95% CI, NR) Sn: 99.6% (95% CI, 98.8% to 100.0%) Sp: 17.9% (95% CI, 16.2% to 19.5%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Cass et al, 2016[232] Shepherd et al, 2010[202] NHANES U.S. Fair	From Cass et al[232] Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥9.3%/TH or FN AUC: 0.79 (95% CI, 0.74 to 0.84) Sn: 39% (95% CI, 27% to 51%) Sp: 89% (95% CI, 87% to 91%) **Index Test/Cutoff/BMD Site** MORES/≥6/TH AUC: 0.87 (95% CI, 0.84 to 0.91) Sn: 96% (95% CI, 87% to 99%) Sp: 61% (95% CI, 58% to 63%) From Shepherd et al[202] Index Test/Cutoff/BMD Site MORES/≥6/any site (thoracic vertebra, LS, arms, ribs, pelvis, legs) AUC: 0.73 (95% CI, NR) Sn: 66% (95% CI, 58% to 72%) Sp: 68% (95% CI, 65% to 70%) Sn [95% CI] by race/ethnicity White: 59.9% [51.8 to 67.5] African American: 78.7% [48.6 to 93.5] Mexican American: 71.3% [57.8 to 81.9] Other: 95.1% [82.5 to 98.7] Sp [95% CI] by race/ethnicity White: 69.4% [66.6 to 72.1] African American: 62.9% [58.2 to 67.3] Mexican American: 58.8% [52.8 to 64.5] Other: 55.1% [44.9 to 65.0])

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Cass et al, 2016[232] Shepherd et al, 2010[202] NHANES U.S. Fair (continued)	Index Test/Cutoff/BMD Site MORES/≥6/LS AUC: 0.66 (95% CI, NR) Sn: 58% (95% CI, 46% to 69%) Sp: 65% (95% CI, 63% to 68%) Sn [95% CI] by race/ethnicity White: 51.1% [38.1 to 63.9] African American: 76.3% [25.3 to 96.9] Mexican American: 59.6% [39.5 to 76.8] Other: 90.4% [66.2 to 97.8] Sp [95% CI] by race/ethnicity White: 67.2% [64.6 to 69.8) African American: 61.6% [56.6 to 66.4] Mexican American: 55.5% [49.9 to 61.0] Other: 49.9% [40.2 to 59.6])
Cass et al, 2006[224] U.S. Fair	Index Test/Cutoff/BMD Site ORAI/≥9/TH or LS AUC: 0.74 (95% CI, 0.63 to 0.84) Sn: 68% (95% CI, 49% to 88%) Sp: 66% (95% CI, 59% to 73%) Hispanic, Estimate [95% CI] Sn: 0.86 [0.47 to 0.99], Sp: 0.59 [0.44 to 0.72], AUC: 0.75 [0.59 to 0.91] African American, Estimate [95% CI] Sn: 0.60 [0.34 to 0.91], Sp: 0.67 [0.55 to 0.76], AUC: 0.69 [0.52 to 0.87] **Index Test/Cutoff/BMD Site** SCORE/≥6/TH or LS AUC: 0.67 (95% CI, 0.54 to 0.79) Sn: 54% (95% CI, 34% to 75%) Sp: 72% (95% CI, 65% to 78%) Hispanic, Estimate [95% CI] Sn 0.71 [0.29 to 0.96], Sp 0.49 [0.35 to 0.63], AUC 0.69 [0.48 to 0.90] African American, Estimate [95% CI] Sn 0.30 [0.00 to 0.56], Sp 0.92 [0.86 to 0.98], AUC 0.70 [0.51 to 0.89]

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Cass et al, 2013[197] U.S. Fair	Index Test/Cutoff/BMD Site MORES/≥6/FN or TH AUC: 0.82 (95% CI, 0.71 to 0.92) Sn: 80% (95% CI, 52% to 96%) Sp: 70% (95% CI, 64% to 74%) (Data reported on includes information for validation study. Article also reports information for development study.)
Chan et al, 2006[207] Singapore Fair	Index Test/Cutoff/BMD Site ABONE/≥3/FN AUC: 0.70 (95% CI, 0.63 to 0.78) Sn: 81.8% (95% CI, NR) Sp: 55.9% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥9/FN AUC: NR Sn: 100% (95% CI, NR) Sp: 9.8% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥20/FN AUC: 0.76 (95% CI, 0.68 to 0.84) Sn: 75.8% (95% CI, NR) Sp: 66.7% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-2/FN AUC: 0.82 (95% CI, 0.75 to 0.90) Sn: 90.9% (95% CI, NR) Sp: 58.8% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: NR Sn: 97.0% (95% CI, NR) Sp: 43.1% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Chan et al, 2006[207] (continued)	Index Test/Cutoff/BMD Site SCORE/≥8/FN AUC: 0.80 (95% CI, 0.72 to 0.87) Sn: 93.9% (95% CI, NR) Sp: 60.8% (95% CI, NR) Index Test/Cutoff/BMD Site ABONE/≥3/LS AUC: 0.66 (95% CI, 0.58 to 0.74) Sn: 73.0% (95% CI, NR) Sp: 54.1% (95% CI, NR) Index Test/Cutoff/BMD Site ABONE/≥2/FN AUC: 0.70 (95% CI, 0.63 to 0.78) Sn: 100% (95% CI, NR) Sp: 16.7% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥16/LS AUC: 0.68 (95% CI, 0.59 to 0.77) Sn: 62.0% (95% CI, NR) Sp: 62.0% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/LS AUC: 0.73 (95% CI, 0.64 to 0.82) Sn: 91.9% (95% CI, NR) Sp: 42.9% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥8/LS AUC: 0.72 (95% CI, 0.63 to 0.80) Sn: 86.5% (95% CI, NR) Sp: 60.2% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Chang et al, 2016[240] Taiwan Fair	Index Test/Cutoff/BMD Site OST/empirically derived threshold (-1.86)/FN AUC: NR Sn: 69.2% (95% CI, NR) Sp: 63.0% (95% CI, NR) (AUC was calculated with OST as a continuous variable rather than categorical using a threshold; therefore, it is reported separately.) Index Test/Cutoff/BMD Site OST/NR/FN AUC: 0.70 (95% CI, 0.66 to 0.74) Sn: NR Sp: NR
Chen et al, 2016[233] Taiwan Fair	Index Test/Cutoff/BMD Site ABONE/>2/FN AUC: 0.78 (95% CI, 0.64 to 0.93) Sn: 100% (95% CI, NR) Sp: 28% (95% CI, NR) Men Index Test/Cutoff/BMD Site ABONE/>2/FN AUC: 0.70 (95% CI, 0.61 to 0.77) Sn: 100% (95% CI, NR) Sp: 10% (95% CI, NR) Women Index Test/Cutoff/BMD Site FRAX Hip without BMD/≥3%/FN AUC: 0.86 (95% CI, 0.73 to 0.98) Sn: 80% (95% CI, NR) Sp: 71% (95% CI, NR) Men

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Chen et al, 2016[233] (continued)	Index Test/Cutoff/BMD Site FRAX Hip without BMD/≥3%/FN AUC: 0.75 (95% CI, 0.65 to 0.86) Sn: 83% (95% CI, NR) Sp: 54% (95% CI, NR) Women Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥20%/FN AUC: 0.77 (95% CI, 0.61 to 0.94) Sn: 0% (95% CI, NR) Sp: 99% (95% CI, NR) Men Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥20%/FN AUC: 0.71 (95% CI, 0.60 to 0.82) Sn: 17% (95% CI, NR) Sp: 96% (95% CI, NR) Women Index Test/Cutoff/BMD Site Garvan Hip without BMD/≥3%/FN AUC: 0.72 (95% CI, 0.44 to 1.00) Sn: 60% (95% CI, NR) Sp: 79% (95% CI, NR) Men Index Test/Cutoff/BMD Site Garvan Hip without BMD/≥3%/FN AUC: 0.80 (95% CI, 0.73 to 0.88) Sn: 28% (95% CI, NR) Sp: 95% (95% CI, NR) Women Index Test/Cutoff/BMD Site Garvan any osteoporotic fx without BMD/≥20%/FN AUC: 0.72 (95% CI, 0.46 to 0.98) Sn: 20% (95% CI, NR) Sp: 96% (95% CI, NR) Men

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Chen et al, 2016[233] (continued)	Index Test/Cutoff/BMD Site Garvan any osteoporotic fx without BMD/≥20%/FN AUC: 0.75 (95% CI, 0.66 to 0.85) Sn: 55% (95% CI, NR) Sp: 73% (95% CI, NR) Women Index Test/Cutoff/BMD Site ORAI/≥9/FN AUC: 0.87 (95% CI, 0.72 to 1.00) Sn: 100% (95% CI, NR) Sp: 19% (95% CI, NR) Men Index Test/Cutoff/BMD Site ORAI/≥9/FN AUC: 0.77 (95% CI, 0.69 to 0.85) Sn: 100% (95% CI, NR) Sp: 5% (95% CI, NR) Women Index Test/Cutoff/BMD Site OSIRIS/≤1/FN AUC: 0.94 (95% CI, 0.88 to 1.00) Sn: 100% (95% CI, NR) Sn: 29% (95% CI, NR) Men Index Test/Cutoff/BMD Site OSIRIS/≤1/FN AUC: 0.83 (95% CI, 0.75 to 0.90) Sn: 100% (95% CI, NR) Sp: 6% (95% CI, NR) Women Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: 0.94 (95% CI, 0.87 to 1.00) Sn: 100% (95% CI, NR) Sp: 58% (95% CI, NR) Men

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Chen et al, 2016[233] (continued)	Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: 0.83 (95% CI, 0.91 to 0.91) Sn: 100% (95% CI, NR) Sp: 27% (95% CI, NR) Women Index Test/Cutoff/BMD Site SCORE/≥6/FN AUC: 0.91 (95% CI, 0.83 to 0.99) Sn: 100% (95% CI, NR) Sp: 45% (95% CI, NR) Men Index Test/Cutoff/BMD Site SCORE/≥6/FN AUC: 0.80 (95% CI, 0.71 to 0.89) Sn: 100% (95% CI, NR) Sp: 15% (95% CI, NR) Women

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Christodoulou et al, 2016[242] Greece Fair	Index Test/Cutoff/BMD Site SCORE/>20.75/site NR AUC: 0.678 (95% CI, 0.640 to 0.717) Sn: 72% (95% CI, NR) Sp: 60% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/>10.5/site NR AUC: 0.632 (95% CI, 0.591 to 0.673) Sn: 65% (95% CI, NR) Sp: 60%(95% CI, NR) Index Test/Cutoff/BMD Site ABONE/>1.5/site NR AUC: 0.618 (95% CI, 0.576 to 0.659) Sn: 66% (95% CI, NR) Sp: 60% (95% CI, NR) Index Test/Cutoff/BMD Site OST/>-2.9/site NR AUC: 0.644 (95% CI, 0.604 to 0.684) Sn: 80% (95% CI, NR) Sp: 43% (95% CI, NR) Index Test/Cutoff/BMD Site OSIRIS/<0.5/site NR AUC: 0.641 (95% CI, 0.601 to 0.681) Sn: 63% (95% CI, NR) Sp: 57% (95% CI, NR) Index Test/Cutoff/BMD Site OSIRIS/<1.5/site NR AUC: 0.641 (95% CI, 0.601 to 0.681) Sn: 76% (95% CI, NR) Sp: 44% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Cook et al, 2005[208] U.K. Fair	Index Test/Cutoff/BMD Site ORAI/≥14/LS or TH AUC: 0.664 (95% CI, 0.595 to 0.793) Sn: 43% (95% CI, NR) Sp: 86% (95% CI, NR) Index Test/Cutoff/BMD Site OSIRIS/≤0/LS or TH AUC: 0.747 (95% CI, 0.702 to 0.805) Sn: 70% (95% CI, NR) Sp: 73% (95% CI, NR) Index Test/Cutoff/BMD Site OST/≤-1/LS or TH AUC: 0.716 (95% CI, 0.775 to 0.669) Sn: 52% (95% CI, NR) Sp: 82% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥12/LS or TH AUC: 0.720 (95% CI, 0.674 to 0.779) Sn: 50% (95% CI, NR) Sp: 83% (95% CI, NR) Index Test/Cutoff/BMD Site SOFSURF/≥1/LS or TH AUC: 0.717 (95% CI, 0.670 to 0.777) Sn: 72% (95% CI, NR) Sp: 67% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Crandall et al, 2014[195] Crandall et al, 2019[141] Crandall et al, 2023[143] Women's Health Initiative U.S. Fair	**Index Test/Cutoff/BMD Site**[195] (Among women ages 50 to 64) FRAX MOF without BMD/≥9.3%/FN (among non-users of hormone therapy) AUC: 0.60 (95% CI, 0.56 to 0.63) Sn: 33.3% (95% CI, 26.3% to 40.4%) Sp: 86.4% (95% CI, 85.1% to 87.7%) Additional score thresholds of >2.24, 3.51, 4.11, 4.59, and 5.04 also reported. **Index Test/Cutoff/BMD Site**[143] (Among women ages 50 to 64) FRAX MOF without BMD/None/FN AUC (95% CI) All: 0.72 (95% CI, 0.68 to 0.75) Black: 0.74 (0.61 to 0.87) Hispanic: 0.74 (0.60 to 0.88) White: 0.72 (0.68 to 0.75) Asian: NR **Index Test/Cutoff/BMD Site**[143] (Among women ages 50 to 64) FRAX MOF without BMD/None/Any site AUC (95% CI) All: 0.64 (0.62 to 0.66) Black: 0.68 (0.63 to 0.73) Hispanic: 0.68 (0.59 to 0.76) White: 0.68 (0.65 to 0.70) Asian: NR **Index Test/Cutoff/BMD Site**[141] FRAX MOF without BMD/≥8.4%/FN or TH or LS AUC: NR Sn: 48.5% (95% CI, 43.4% to 53.6%) Sp: 63.4% (95% CI, 60.9% to 65.9%) Ages 60–64 years

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Crandall et al, 2014[195] Crandall et al, 2019[141] Crandall et al, 2023[143] (continued)	**Index Test/Cutoff/BMD Site**[141] FRAX MOF without BMD/≥8.4%/FN or TH or LS AUC: NR Sn: 5.2% (95% CI, 0.7% to 9.7%) Sp: 95.8% (95% CI, 94.7% to 96.9%) Ages 50–54 years **Index Test/Cutoff/BMD Site**[141] FRAX MOF without BMD/≥8.4%/FN or TH or LS AUC: NR Sn: 16.9% (95% CI, 11.9% to 21.9%) Sp: 87.1% (95% CI, 85.4% to 88.8%) Ages 55–59 years **Index Test/Cutoff/BMD Site**[141] FRAX MOF without BMD/≥8.4%/FN or TH or LS AUC: NR Sn: 5.2% (95% CI, 0.7% to 9.7%) Sp: 95.8% (95% CI, 94.7% to 96.9%) Ages 50–54 years **Index Test/Cutoff/BMD Site**[195] (Among women ages 50 to 64) OST/<2/FN AUC: 0.75 (95% CI, 0.72 to 0.78) Sn: 79.3% (95% CI, 73.2% to 85.4%) Sp: 70.1% (95% CI, 68.4% to 71.8%) Additional score thresholds of <3, 4, 8 also reported. **Index Test/Cutoff/BMD Site**[143] (Among women ages 50 to 64) OST/None/FN AUC (95% CI) All: 0.83 (0.80 to 0.85) Black: 0.85 (0.74 to 0.96) Hispanic: 0.79 (0.65 to 0.93) White: 0.82 (0.80 to 0.85) Asian: NR

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Crandall et al, 2014[195] Crandall et al, 2019[141] Crandall et al, 2023[143] (continued)	**Index Test/Cutoff/BMD Site**[143] (Among women ages 50 to 64) OST/None/Any site AUC (95% CI) All: 0.74 (0.72 to 0.76) Black: 0.76 (0.71 to 0.81) Hispanic: 0.76 (0.68 to 0.84) White: 0.75 (0.73 to 0.78) Asian: NR **Index Test/Cutoff/BMD Site**[195] (Among women ages 50 to 64) SCORE/>7/FN AUC: 0.72 (95% CI, 0.69 to 0.76) Sn: 74.1% (95% CI, 67.6% to 80.7%) Sp: 70.8% (95% CI, 69.1% to 72.5%) Additional score thresholds of >5, 6, and >-6 also reported.
D'Amelio et al, 2005[225] Italy Fair	Index Test/Cutoff/BMD Site AMMEB/≥10/FN or LS AUC: 0.71 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site NOF/≥1/FN or LS AUC: 0.60 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site ORAI/>8/FN or LS AUC: 0.32 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OST/<2/FN or LS AUC: 0.33 (95% CI, NR) Sn: NR Sp: NR

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
D'Amelio et al, 2013[199] Italy Fair	Index Test/Cutoff/BMD Site AMMEB/≥10/FN or LS AUC: 0.63 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site NOF/≥1/FN or LS AUC: 0.60 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site ORAI/>8/FN or LS AUC: 0.68 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OST/<2/FN or LS AUC: 0.32 (95% CI, NR) Sn: NR Sp: NR

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Diem et al, 2017[235] Lynn et al, 2008[218] MrOs Multicountry (incl. U.S.) Fair	From Diem et al, 2017[235] Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥7%/TH or LS or FN AUC: 0.62 (95% CI, NR) Sn: 81% (95% CI, 75% to 86%) Sp: 33% (95% CI, 32% to 35%) Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥8%/TH or LS or FN AUC: 0.62 (95% CI, NR) Sn: 71% (95% CI, 65% to 77%) Sp: 46% (95% CI, 45% to 48%) Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥9%/TH or LS or FN AUC: 0.62 (95% CI, NR) Sn: 62% (95% CI, 55% to 69%) Sp: 56% (95% CI, 54% to 58%) Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥9.3%/TH or LS or FN AUC: 0.62 (95% CI, NR) Sn: 59% (95% CI, 52% to 66%) Sp: 59% (95% CI, 57% to 60%) Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥10%/TH or LS or FN AUC: 0.62 (95% CI, NR) Sn: 53% (95% CI, 46% to 60%) Sp: 65% (95% CI, 63% to 66%) Index Test/Cutoff/BMD Site OST/<-1/TH or LS or FN AUC: 0.68 (95% CI, NR) Sn: 47% (95% CI, 40% to 54%) Sp: 78% (95% CI, 77% to 79%) Index Test/Cutoff /BMD Site OST/<0/TH or LS or FN AUC: 0.68 (95% CI, NR) Sn: 63% (95% CI, 56% to 69%) Sp: 78% (95% CI, 77% to 79%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Diem et al, 2017[235] Lynn et al, 2008[218] MrOs (continued)	Index Test/Cutoff/BMD Site OST/<1/TH or LS or FN AUC: 0.68 (95% CI, NR) Sn: 77% (95% CI, 71% to 82%) Sp: 51% (95% CI, 50% to 53%) Index Test/Cutoff/BMD Site OST/<2/TH or LS or FN AUC: 0.68 (95% CI, NR) Sn: 83% (95% CI, 77% to 87%) Sp: 36% (95% CI, 35% to 38%) Index Test/Cutoff/BMD Site OST/<3/TH or LS or FN AUC: 0.68 (95% CI, NR) Sn: 89% (95% CI, 84% to 93%) Sp: 25% (95% CI, 24% to 26%) From Lynn et al[218] Index Test/Cutoff/BMD Site MOST/≤26/FN or LS or TH AUC: 0.799 Sn: 88.5% (95% CI, 84.3% to 92.5%) Sp: 50.0% (95% CI, 48.5% to 51.5%) U.S. participants only; at a threshold of ≤27, Sn was 94.7% and Sp was 37.8%. CIs were calculated. Index Test/Cutoff/BMD Site OST/<2/FN or LS or TH AUC: 0.714 (95% CI, NR) Sn: 87.6% (95% CI, NR) Sp: 36.1% (95% CI, NR) U.S. participants only

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Diem et al, 2017[235] Lynn et al, 2008[218] MrOs (continued)	Index Test/Cutoff/BMD Site MOST/≤21/FN or LS or TH AUC: 0.831 (95% CI, NR) Sn: 86.8% (95% CI, 79.6 to 93.3) Sp: 59.3% (95% CI, 57.0 to 61.6) Hong Kong participants only; at a threshold of ≤22, the Sn was 94.2% and the Sp was 42.3%. CIs were calculated. Index Test/Cutoff/BMD Site OST/≤-2/FN or LS or TH AUC: 0.759 (95% CI, NR) Sn: 81.8% (95% CI, NR) Sp: 56.2% (95% CI, NR) Hong Kong participants only; at a threshold of ≤-1 Sn was 91.1% and Sp was 36.4%. Index Test/Cutoff/BMD Site MOST/NR/FN AUC: 0.808 (95% CI, NR) Sn: NR Sp: NR U.S. participants only Index Test/Cutoff/BMD Site OST/<2/FN AUC: 0.740 (95% CI, NR) Sn: NR Sp: NR U.S. participants only Index Test/Cutoff /BMD Site MOST/NR/FN AUC: 0.876 (95% CI, NR) Sn: NR Sp: NR Hong Kong participants only
Diem et al, 2017[235] Lynn et al, 2008[218] MrOs (continued)	Index Test/Cutoff/BMD Site OST/NR/FN AUC: 0.849 (95% CI, NR) Sn: NR Sp: NR Hong Kong participants only

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Erjiang et al, 2021[243] Ireland Fair	Index Test/Cutoff/BMD Site OSTi/<2/FN or TH or LS AUC: NR Sn: 89.9% (95% CI, NR) Sp: 46.2% (95% CI, NR) Women Index Test/Cutoff/BMD Site OSTi/<2/FN or TH or LS AUC: 0.739 Sn: 71.19% Sp: 63.73% Men
Geusens et al, 2002[231] U.S. Fair	Index Test/Cutoff/BMD Site ORAI/≥9/FN AUC: NR Sn: 90.1% (95% CI, NR) Sp: 52.0% (95% CI, NR) U.S. Clinic Sample Index Test/Cutoff/BMD Site OST/≤1/FN AUC: NR Sn: 87.5% (95% CI, NR) Sp: 51.7% (95% CI, NR) U.S. clinic sample only Index Test/Cutoff/BMD Site SOFSURF/≥0/FN AUC: NR Sn: 92.1% (95% CI, NR) Sp: 37.3% (95% CI, NR) Results also reported for the U.S. clinic sample Index Test/Cutoff/BMD Site SCORE/≥7/FN AUC: NR Sn: 94.1% (95% CI, NR) Sp: 48.9% (95% CI, NR) Results also reported for the U.S. clinic sample

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Gourlay et al, 2008[230] Study of Osteoporotic Fractures U.S. Fair	Index Test/Cutoff/BMD Site OST/≤-1/FN or LS AUC: 0.72 (95% CI, 0.71 to 0.73) Sn: 85% (95% CI, 83% to 87%) Sp: 52% (95% CI, 51% to 54%) Index Test/Cutoff/BMD Site SCORE/≥6/FN or LS AUC: 0.71 (95% CI, 0.70 to 0.72) Sn: 99% (95% CI, 99% to 99%) Sp: 93% (95% CI, 93% to 94%) Index Test/Cutoff/BMD Site ORAI/≥9/FN or LS AUC: 0.70 (95% CI, 0.69 to 0.71) Sn: 100% Sp: 100% Index Test/Cutoff/BMD Site OST/≤-1/FN AUC: 0.76 (95% CI, 0.74 to 0.77) Sn: NR Sp: NR
Gourlay et al, 2005[209] Richy et al, 2004[210] Ben Sedrine et al, 2001[211] Belgium Fair	Index Test/Cutoff/BMD Site ORAI/≥8/FN AUC: NR Sn: 82% (95% CI, NR) Sp: 45% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥8/FN AUC: 0.75 (95% CI, 0.71 to 0.79) Sn: 88.5% (95% CI, 82.0% to 93.3%) Sp: 46.2% (95% CI, 44.2% to 48.2%) For ages 45–64 years

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Gourlay et al, 2005[209] Richy et al, 2004[210] Ben Sedrine et al, 2001[211] (continued)	Index Test/Cutoff/BMD Site ORAI/≥8/FN or TH or LS AUC: 0.67 (95% CI, NR) Sn: 76% (95% CI, NR) Sp: 48% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥13/FN AUC: 0.75 (95% CI, 0.71 to 0.78) Sn: 89.2% (95% CI, 84.6% to 92.8%) Sp: 44.7% (95% CI, 42.0% to 47.5%) For ages 65 years or older Index Test/Cutoff/BMD Site OST/≤1/FN AUC: 0.77 (95% CI, 0.73 to 0.81) Sn: 89.2% (95% CI, 82.8% to 93.8%) Sp: 45.0% (95% CI, 43.0% to 47.0%) For ages 45–64 years Index Test/Cutoff/BMD Site OST/≤-1/FN AUC: 0.76 (95% CI, 0.73 to 0.79) Sn: 84.6% (95% CI, 79.5% to 89.0%) Sp: 47.5% (95% CI, 44.7% to 50.3%) For ages 65 years or older Index Test/Cutoff/BMD Site OST/<2/FN or TH or LS AUC: 0.726 (95% CI, NR) Sn: 86% (95% CI, NR) Sp: 40% (95% CI, NR) Index Test/Cutoff /BMD Site OST/<2/FN AUC: NR Sn: 92% (95% CI, NR) Sp: 37% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year **Study Name** **Country** **Study Quality**	**Index Test/Score Threshold/Site of Reference BMD Measurement** **AUC (95% CI)** **Sensitivity (95% CI)** **Specificity (95% CI)**
Gourlay et al, 2005[209] Richy et al, 2004[210] Ben Sedrine et al, 2001[211] (continued)	Index Test/Cutoff/BMD Site SCORE/≥6/FN or TH or LS AUC: 0.71 (95% CI, NR) Sn: 93.9% (95% CI, NR) Sp: 23.7% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥6/FN AUC: NR Sn: 96.9% (95% CI, NR) Sp: 21.4% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥6/LS AUC: NR Sn: 93.5% (95% CI, NR) Sp: 21.7% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥7/FN AUC: NR Sn: 88% (95% CI, NR) Sp: 40% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥7/FN AUC: 0.76 (95% CI, 0.72 to 0.80) Sn: 88.5% (95% CI, 82.0% to 93.3%) Sp: 39.8% (95% CI, 37.8% to 41.7%) For ages 45 to 64 years Index Test/Cutoff/BMD Site SCORE/≥7/FN or TH or LS AUC: 0.708 (95% CI, NR) Sn: 86% (95% CI, NR) Sp: 40% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Gourlay et al, 2005[209] Richy et al, 2004[210] Ben Sedrine et al, 2001[211] (continued)	Index Test/Cutoff/BMD Site SCORE/≥8/FN or TH or LS AUC: NR Sn: 82.4% (95% CI, NR) Sp: 42.4% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥11/FN AUC: 0.75 (95% CI, 0.71 to 0.78) Sn: 88.8% (95% CI, 84.1% to 92.5%) Sp: 42.3% (95% CI, 39.6% to 45.1%) For ages 65 years or older Index Test/Cutoff/BMD Site OSIRIS/<1/FN or TH or LS AUC: 0.73 (95% CI, NR) Sn: 64% (95% CI, NR) Sp: 69% (95% CI, NR) Index Test/Cutoff/BMD Site OSIRIS/<1/FN AUC: NR Sn: 75% (95% CI, NR) Sp: 66% (95% CI, NR)
Hamdy et al, 2018[238] U.S. Fair	Index Test/Cutoff/BMD Site FRAX without BMD/MOF≥20% or hip ≥3%/FN AUC: NR Sn: 26.7% (95% CI, NR) Sp: 88.0% (95% CI, NR) Index Test/Cutoff/BMD Site FRAX without BMD/risk ≥hypothetical man of same age, weigh, height with no risk factors/FN AUC: NR Sn: 79.1% (95% CI, NR) Sp: 31.9% (95% CI, NR) Index Test/Cutoff/BMD Site FRAX without BMD/hip ≥1% or MOF >5%/FN AUC: NR Sn: 91.9% (95% CI, NR) Sp: 18.8% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Harrison et al, 2006[217] United Kingdom Fair	Index Test/Cutoff/BMD Site ORAI/NR/FN or TH or LS AUC: 0.67 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OSIRIS/NR/FN or TH or LS AUC: 0.70 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OST/NR/FN or TH or LS AUC: 0.69 (95% CI, NR) Sn: NR Sp: NR Index Test/Cutoff/BMD Site SCORE/NR/FN or TH or LS AUC: 0.67 (95% CI, NR) Sn: NR Sp: NR

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Inderjeeth et al, 2020[239] Australia Fair	Index Test/Cutoff/BMD Site Garvan hip without BMD/empirically derived, age-stratified risk thresholds/FN or TH or LS or forearm AUC: 0.721 (95% CI, 0.674 to 0.768) Sn: 71.5% (95% CI, NR) Sp: 90.0% (95% CI, NR) AUC values abstracted from Figure 2. Sn and Sp calculated from “total” data provided in Table 2. Index Test/Cutoff/BMD Site Garvan MOF without BMD/empirically derived, age-stratified risk thresholds/FN or TH or LS or forearm AUC: 0.706 (95% CI, 0.658 to 0.753) Sn: 68.5% (95% CI, NR) Sp: 94.8% (95% CI, NR) AUC values abstracted from Figure 2. Sn and Sp calculated from “total” data provided in Table 2. Index Test/Cutoff/BMD Site FRAX Hip without BMD/≥3%/FN or TH or LS or forearm AUC: 0.75 (95% CI, 0.70 to 0.80) Sn: NR Sp: NR AUC values abstracted from Figure 2. Sn and Sp could not be calculated. Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥20%/FN or TH or LS or forearm AUC: 0.76 (95% CI, 0.71 to 0.81) Sn: NR Sp: NR AUC values abstracted from Figure 2. Sn and Sp could not be calculated.

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Jiang et al, 2016[236] U.S. Fair	Index Test/Cutoff/BMD Site FRAX MOF without BMD/≥9.3%/site NR AUC: 0.62 (95% CI, 0.52 to 0.72) Sn: 24% (95% CI, 11% to 40%) Sp: 83% (95% CI, 79% to 87%) (An additional threshold of FRAX ≥4.7% was reported in the study but not included in the index test results because it was not prespecified.) Index Test/Cutoff/BMD Site OST/<2/site NR AUC: 0.73 (95% CI, 0.65 to 0.81) Sn: 79% (95% CI, 63% to 91%) Sp: 56% (95% CI, 51% to 61%) (A threshold of OST <3 was also reported in this study but not included in the index test results because it was not prespecified.) Index Test/Cutoff/BMD Site ORAI/≥9/site NR AUC: 0.69 (95% CI, 0.60 to 0.78) Sn: 74% (95% CI, 57% to 87%) Sp: 62% (95% CI, 57% to 67%) Index Test/Cutoff/BMD Site SCORE/≥6/site NR AUC: 0.75 (95% CI, 0.67 to 0.83) Sn: 92% (95% CI, 79% to 98%) Sp: 34% (95% CI, 29% to 39%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Jimenez-Nunez et al, 2013[203] Spain Fair	Index Test/Cutoff/BMD Site ORAI/≥9/FN or LS AUC: 0.68 (95% CI, NR) Sn: 78% (95% CI, NR) Sp: 52% (95% CI, NR) Index Test/Cutoff/BMD Site OSIRIS/≤-3/FN or LS AUC: 0.71 (95% CI, NR) Sn: 81% (95% CI, NR) Sp: 54% (95% CI, NR) Index Test/Cutoff/BMD Site OST/≤-1/FN or LS AUC: 0.71 (95% CI, NR) Sn: 83% (95% CI, NR) Sp: 52% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥6/FN or LS AUC: 0.67 (95% CI, NR) Sn: 68% (95% CI, NR) Sp: 60% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Kung et al, 2005[212] Kung et al, 2003[213] Hong Kong Fair	From Kung et al, 2005[212] Men Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: 0.85 (95% CI, 0.80 to 0.89) Sn: 83% (95% CI, NR) Sp: 67% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/LS AUC: 0.79 (95% CI, 0.74 to 0.83) Sn: 72% (95% CI, NR) Sp: 65% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/LS or FN AUC: 0.78 (95% CI, 0.73 to 0.82) Sn: 71% (95% CI, NR) Sp: 68% (95% CI, NR) From Kung et al, 2003[213] Women Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: 0.80 (95% CI, 0.78 to 0.84) Sn: 88% (95% CI, NR) Sp: 54% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/FN or LS AUC: 0.75 (95% CI, 0.72 to 0.79) Sn: 79% (95% CI, NR) Sp: 60% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Leslie et al, 2013[200] Manitoba BMD Registry Canada Fair	Index Test/Cutoff/BMD Site FRAX MOF without BMD/NR/FN or TH or LS AUC: 0.67 (95% CI, 0.66 to 0.68) Sn: NR Sp: NR (Sn and Sp cannot be calculated based on data provided in the study.) Index Test/Cutoff/BMD Site OST/NR/FN or TH or LS AUC: 0.72 (95% CI, 0.71 to 0.73) Sn: NR Sp: NR (Sn and Sp cannot be calculated based on data provided in the study.)
Machado et al, 2010[201] Portugal Fair	Index Test/Cutoff/BMD Site OST/<2/FH or TH or LS AUC: 0.63 (95% CI, 0.52 to 0.73) Sn: 61.8% (95% CI, NR) Sp: 63.7% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/<2/FH or TH or LS AUC: 0.62 (95% CI, 0.51 to 0.72) Sn: 55.9% (95% CI, NR) Sp: 67.9% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Martinez-Aguila et al, 2007[214] Spain Fair	Index Test/Cutoff/BMD Site ORAI/≥9/FN or LS AUC: 0.62 (95% CI, 0.56 to 0.67) Sn: 64.1% (95% CI, 54.7% to 72.7%) Sp: 58.9% (95% CI, 54.7% to 63.1%) Index Test/Cutoff/BMD Site OSIRIS/≤1/FN or LS AUC: 0.63 (95% CI, 0.57 to 0.69) Sn: 58.1% (95% CI, 48.6% to 67.2%) Sp: 67.9% (95% CI, 63.8% to 71.8%) Index Test/Cutoff/BMD Site OST/<2/FN or LS AUC: 0.64 (95% CI, 0.59 to 0.69) Sn: 69.2% (95% CI, 60.0% to 77.4%) Sp: 58.8% (95% CI, 54.5% to 62.9%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Mauck et al, 2005[226] U.S. Fair	Index Test/Cutoff/BMD Site NOF/≥1/FN AUC: 0.70 (95% CI, 0.63 to 0.77) Sn: 100% (95% CI, 95% to 100%) Sp: 10% (95% CI, 5% to 16%) For the 45- to 64-year-old group (n=79): AUC 0.69 [95% CI, 0.51 to 0.70]; Sn 100% [95% CI, 72% to 100%], Sp 19% [95% CI, 11% to 31%]) For the ≥65 years group (n=123): Unadjusted AUC 0.60 (95% CI, 0.51 to 0.70); Sn 100% (95% CI, 94 to 100); Sp 0% (95% CI, 0 to 6) Index Test/Cutoff/BMD Site ORAI/≥9/FN AUC: 0.84 (95% CI, 0.78 to 0.89) Sn: 99% (95% CI, 92% to 100%) Sp: 36% (95% CI, 28% to 44%) For the 45- to 64-year-old group (n=79): AUC 0.82 [95% CI, 0.71 to 0.94]; Sn 91% [95% CI, 59% to 100%], Sp 69% [95% CI, 57% to 80%]) For the ≥65-year-old group (n=123): Unadjusted AUC 0.79 (95% CI, 0.71 to 0.87); Sn 100% (95% CI, 94 to 100); Sp 0% (95% CI, 0 to 6) Index Test/Cutoff/BMD Site SCORE/≥6/FN AUC: 0.87 (95% CI, 0.81 to 0.92) Sn: 100% (95% CI, 95% to 100%) Sp: 25% (95% CI, 18% to 33%) For the 45- to 64-year-old group (n=79): AUC 0.85 [95% CI, 0.0.72 to 0.99]; Sn 100% [95% CI, 72% to 100%], Sp 41% [95% CI, 29% to 54%]) For the ≥65-year-old group (n=123): Unadjusted AUC 0.80 (95% CI, 0.72 to 0.88); Sn 100% (95% CI, 94 to 100); Sp 8% (95% CI, 3 to 17)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
McLeod et al, 2015[205] Canada Good	Index Test/Cutoff/BMD Site OST/<2/FN AUC: 0.81 (95% CI, 0.692 to 0.985) Sn: 87.5% (95% CI, NR) Sp: 62.7% (95% CI, NR) Index Test/Cutoff/BMD Site OST/<2/LS AUC: 0.71 (95% CI, 0.517 to 0.846) Sn: 78.6% (95% CI, NR) Sp: 63.7% (95% CI, NR) Index Test/Cutoff/BMD Site OST/<0/FN AUC: 0.81 (95% CI, 0.69 to 0.99) Sn: 50.0% (95% CI, NR) Sp: 91.6% (95% CI, NR) Index Test/Cutoff/BMD Site OST/<0/LS AUC: 0.71 (95% CI, 0.56 to 0.85) Sn: 28.6% (95% CI, NR) Sp: 91.2% (95% CI, NR)
Moon et al, 2016[244] KNHANES Republic of Korea Fair	Index Test/Cutoff/BMD Site OSTA/<0.5/Mean T-score from FN, TH and LS AUC: 0.737 (95% CI, 0.69 to 0.78) Sn: 71.9% (95% CI, NR) Sp: 64.0% (95% CI, NR)
Morin et al, 2009[159] Manitoba BMD Registry Canada Fair	Index Test/Cutoff/BMD Site OST/≤1/FN or TH or LS AUC: 0.71 (95% CI, 0.69 to 0.72) Sn: 46.8% (95% CI, 45.7% to 47.9%) Sp: 81.1% (95% CI, 80.3% to 82.0%) Index Test/Cutoff/BMD Site OST/≤1/FN AUC: 0.77 (95% CI, 0.75 to 0.79) Sn: 60.2% (95% CI, 59.2% to 61.3%) Sp: 78.8% (95% CI, 77.9% to 79.6%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year **Study Name** **Country** **Study Quality**	**Index Test/Score Threshold/Site of Reference BMD Measurement** **AUC (95% CI)** **Sensitivity (95% CI)** **Specificity (95% CI)**
Nguyen et al, 2004[227] Dubbo Osteoporosis Epidemiology Study Australia Fair	Index Test/Cutoff/BMD Site ORAI/>15/FN or LS AUC: NR Sn: 61% (95% CI, NR) Sp: 68% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/<-1/FN or LS AUC: NR Sn: 41% (95% CI, NR) Sp: 24% (95% CI, NR) Index Test/Cutoff/BMD Site SOFSURF/>1.7/FN or LS AUC: NR Sn: 78% (95% CI, NR) Sp: 36% (95% CI, NR)
Oh et al, 2013[204] KNHANES Republic of South Korea Fair	Index Test/Cutoff/BMD Site OSTA/<0/FN or LS AUC: 0.62 (95% CI, NR) Sn: 94.2% (95% CI, 91.0% to 96.5%) Sp: 29.2% (95% CI, 26.0% to 32.6%) (SE AUC 0.011) Index Test/Cutoff/BMD Site OSTA/≤-1/FN or LS AUC: NR Sn: 76.1% (95% CI, 71.0% to 80.8%) Sp: 67.1% (95% CI, 63.6% to 70.5%)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	**Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)**
Oh et al, 2016[229] KNHANES Republic of Korea Fair	Index Test/Cutoff/BMD Site OSTA/≤1/FN or LS AUC: 0.627 (95% CI, NR) Sn: 92.3% (95% CI, 84.8% to 96.9%) Sp: 33.2% (95% CI, 30.3% to 36.2%) (AUC SE=0.016) Index Test/Cutoff/BMD Site OSTA/≤0/FN or LS AUC: 0.627 (95% CI, NR) Sn: 84.6% (95% CI, 75.5% to 91.3%) Sp: 48.4% (95% CI, 45.3% to 51.5%) (AUC SE=0.016)
Pang et al, 2014[196] Australia Fair	Index Test/Cutoff/BMD Site FRAX Hip without BMD/>3%/FN or TH or LS AUC: 0.70 (95% CI, 0.64 to 0.75) Sn: 92.2% (95% CI, NR) Sp: 37.1% (95% CI, NR) Index Test/Cutoff/BMD Site FRAX MOF without BMD/>6.5%/FN or TH or LS AUC: 0.68 (95% CI, 0.63 to 0.74) Sn: 89.6% (95% CI, NR) Sp: 35.0% (95% CI, NR) Index Test/Cutoff/BMD Site OST/<0/FN or TH or LS AUC: 0.76 (95% CI, 0.71 to 0.82) Sn: 90.9% (95% CI, NR) Sp: 39.9% (95% CI, NR)
Park et al, 2003[215] Republic of Korea Fair	Index Test/Cutoff/BMD Site OSTA/≤-1/FN AUC: 0.87 (95% CI, NR) Sn: 87% (95% CI, NR) Sp: 67% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Richards et al, 2014[198] U.S. Fair	Index Test/Cutoff/BMD Site OST/≤6/FN or TH AUC: 0.67 (95% CI, NR) Sn: 82.6% (95% CI, NR) Sp: 33.6% (95% CI, NR) (Also reported by race and age: Caucasian: Sn 85.5%, Sp 32.2% [n=373] African American: Sn 70%, Sp 36.4% [n=130] Age ≤65 years: Sn 69%, Sp 50.5% [n=270] Age >65 years: Sn 94%, Sp 17.1% [n=250]) Index Test/Cutoff/BMD Site OST/≤0/FN or TH AUC: NR Sn: 40.2% (95% CI, NR) Sp: 85.4% (95% CI, NR) (Also reported by race and age: Caucasian: Sn 42.0%, Sp 84.9% [n=373] African American: Sn 25%, Sp 87.3% [n=130] Age ≤65 years: Sn 14.3%, Sp 99% [n=270] Age >65 years: Sn 62%, Sp 72.2% [n=250])
Rud et al, 2005[228] Danish Osteoporosis Prevention Study Denmark Fair	Index Test/Cutoff/BMD Site ORAI/>10/FN or TH or LS AUC: 0.64 (95% CI, 0.58 to 0.70) Sn: 44% (95% CI, NR) Sp: 77% (95% CI, NR) (also reported for cutoff >2, 5, 7, and 11) Index Test/Cutoff/BMD Site OST/<2/FN or TH or LS AUC: 0.68 (95% CI, 0.63 to 0.74) Sn: 53% (95% CI, NR) Sp: 72% (95% CI, NR) (also reported for cutoff <6, <5, <4, <3) Index Test/Cutoff/BMD Site SCORE/>6/FN or TH or LS AUC: 0.68 (95% CI, 0.63 to 0.73) Sn: 62% (95% CI, NR) Sp: 65% (95% CI, NR) (Also reported for thresholds >3, 4, 5, 6, 7)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Shepherd et al, 2010[202] NHANES U.S. Fair	Index Test/Cutoff/BMD Site MORES/≥6/any site (thoracic vertebra, LS, arms, ribs, pelvis, legs) AUC: 0.73 (95% CI NR) Sn: 66% (95% CI, 58% to 72%) Sp: 68% (95% CI, 65% to 70%) By race/ethnicity Sn (95% CI) White: 59.9% (51.8 to 67.5) African American: 78.7% (48.6 to 93.5) Mexican American: 71.3% (57.8 to 81.9) Other: 95.1% (82.5 to 98.7) Sp (95% CI) White: 69.4% (66.6 to 72.1) African American: 62.9% (58.2 to 67.3) Mexican American: 58.8% (52.8 to 64.5) Other: 55.1% (44.9 to 65.0) By Age Age Group 50-54 Sn: 28.8%, Sp: 89.8% Age Group 55-59 Sn: 46.7%, Sp: 63.7% Age Group 60-64 Sn: 87.8%, Sp: 67.1% Age Group 65-69 Sn: 82.3%, Sp: 61.8% Age Group 70-74 Sn: 71.0%, Sp: 49.7% Age Group 75-79 Sn: 83.9%, Sp: 47.2% Age Group 80-84 Sn: 85.1%, Sp: 31.0% Age Group 85-89 Sn: 75.1%, Sp: 22.7% Index Test/Cutoff/BMD Site MORES/≥6/LS AUC: 0.66 (95% CI, NR) Sn: 58% (95% CI, 46% to 69%) Sp: 65% (95% CI, 63% to 68%) Sn (95% CI) White: 51.1% (38.1 to 63.9) African American: 76.3% (25.3 to 96.9) Mexican American: 59.6% (39.5 to 76.8) Other: 90.4% (66.2 to 97.8) Sp (95% CI) White: 67.2% (64.6 to 69.8) African American: 61.6% (56.6 to 66.4) Mexican American: 55.5% (49.9 to 61.0) Other: 49.9% (40.2 to 59.6)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Shuler et al, 2016[241] U.S. Fair	Index Test/Cutoff/BMD Site FRAX without BMD/≥3% for Hip or ≥20% for MOF/Site NR AUC: NR Sn: 100% (95% CI, NR) Sp: 91% (95% CI, NR)
Sinnott et al, 2006[219] U.S. Fair	Index Test/Cutoff/BMD Site OST/<2/FN or TH or LS AUC: 0.89 (95% CI, 0.75 to 1.03) Sn: 89% (95% CI, NR) Sp: 71% (95% CI, NR) (Values reported are for <2 threshold; threshold <4 Sn 89%, Sp 54%.)
Toh et al, 2019[245] Malaysia Fair	Index Test/Cutoff/BMD Site SCORE/≥6/FN or LS AUC: 0.161 (95% CI, NR) Sn: 100% (95% CI, NR) Sp: 8.2% (95% CI, NR) Index Test/Cutoff/BMD Site ORAI/≥9/FN or LS AUC: 0.129 (95% CI, NR) Sn: 93.8% (95% CI, NR) Sp: 20.0% (95% CI, NR) Index Test/Cutoff/BMD Site ABONE/≥2/FN or LS AUC: 0.088 (95% CI, NR) Sn: 87.5% (95% CI, NR) Sp: 28.4% (95% CI, NR) Index Test/Cutoff/BMD Site ABONE/≥2/FN AUC: 0.034 (95% CI, NR) Sn: 83.3% (95% CI, NR) Sp: 27.1% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Toh et al, 2019[245] Malaysia Fair (continued)	Index Test/Cutoff/BMD Site Malaysian Osteoporotic Screening Tool/≥4/FN or LS AUC: 0.105 (95% CI, NR) Sn: 100% (95% CI, NR) Sp: 2.2% (95% CI, NR) Index Test/Cutoff/BMD Site OSTA/≤-1/FN or LS AUC: 0.078 (95% CI, NR) Sn: 68.8%(95% CI, NR) Sp: 51.5% (95% CI, NR) Index Test/Cutoff /BMD Site OSTA/≤-1/FN AUC: 0.03 (95% CI, NR) Sn: 50.0% (95% CI, NR) Sp: 49.3% (95% CI, NR) Index Test/Cutoff/BMD Site SCORE/≥6/FN AUC: 0.072 (95% CI, NR) Sn: 100% (95% CI, NR) Sp: 7.6% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Williams et al, 2017[237] U.S. Fair	Index Test/Cutoff/BMD Site FRAX without BMD/Hip ≥3% or MOF≥20%/TH or LS or FN as captured from data in existing electronic health record AUC: 0.65 (95% CI, 0.59 to 0.71) Sn: 68.8% (95% CI, NR) Sp: 54.4% (95% CI, NR) (Sn, Sp, and AUC were not reported separately for hip and MOF.) **Index Test/Cutoff/BMD Site** as captured from direct patient questionnaire FRAX without BMD Hip/≥3% or MOF≥20%/TH or LS or FN AUC: 0.72 (95% CI, 0.67 to 0.78) Sn: NR Sp: NR Index Test/Cutoff/BMD Site OST/<0.99/TH or LS or FN AUC: 0.71 (95% CI, 0.65 to 0.76) Sn: 68.8% (95% CI, NR) Sp: 59.8% (95% CI, NR) (Based on the scoring methodology for OST, assumed that the test indicated risk for osteoporosis if a participant scored below 0.99.) Index Test/Cutoff/BMD Site VA-FARA/Hip≥3% or MOF≥20%/TH or LS or FN AUC: 0.64 (95% CI, 0.58 to 0.70) Sn: 64.3% (95% CI, NR) Sp: 58.4% (95% CI, NR)

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

Author, Year Study Name Country Study Quality	Index Test/Score Threshold/Site of Reference BMD Measurement AUC (95% CI) Sensitivity (95% CI) Specificity (95% CI)
Zimering et al, 2007[216] U.S. Fair	Index Test/Cutoff/BMD Site MSCORE/>9/FN AUC: 0.84 (95% CI, 0.74 to 0.95) Sn: 88% (95% CI, NR) Sp: 57% (95% CI, NR) African American Cohort Using a Caucasian T-score reference range Sn 100% (95% CI, NR) Sp 73% (95% CI, NR) Using an African American T-score reference range Sn 93% (95% CI, NR) Sp 79% (95% CI, NR) Index Test/Cutoff/BMD Site OST<2/FN AUC: 0.81 (95% CI, 0.70 to 0.92) Sn: 75% (95% CI, NR) Sp: 68% (95% CI, NR) (Cutoff established in elderly male population. The study also reported data for an African American validation cohort but combined data from 95 new subjects and 39 subjects from development cohort, so it was not pure external validation cohort. Caucasian reference range, Sn 100%, Sp 83%; African American reference range, Sn 71%, Sp 86%.) Index Test/Cutoff/BMD Site OST/<3/FN AUC: 0.81 (95% CI, 0.70 to 0.92) Sn: 75% (95% CI, NR) Sp: 59% (95% CI, NR) (Cutoff established in male veteran population. The study also reported data for an African American validation cohort but combined data from 95 new subjects and 39 subjects from development cohort, so it was not pure external validation cohort. Caucasian reference range, Sn 100%, Sp 76%; African American reference range, Sn 79%, Sp 80%.) Index Test/Cutoff/BMD Site Reduced MScore >9/FN AUC: 0.81 (95% CI, 0.69 to 0.92) Sn: 85% (95% CI, NR) Sp: 58% (95% CI, NR)

Abbreviations: ABONE=Age, Bone, No Estrogen; AMMEB=Age, years after Menopause, age at Menarche; AUC=area under the curve; BMD=body mass index; CI=confidence interval; FH= femoral head; FN=femoral neck; FRAX=Fracture Risk Assessment Tool; fx=fracture; KNHANES=Korean National Health and Nutrition Examination Survey; LS=lumbar spine; MOF=major osteoporotic fracture; MORES=Male Osteoporosis Risk Estimation Score; MOST=Male Osteoporosis Screening Tool; MSCORE=Male Simple Calculated Osteoporosis Risk Estimation; NA=not applicable; NHANES=National Health and Nutrition Examination Survey; NOF=National Osteoporosis Foundation Score; NR=not reported; ORAI=Osteoporosis Risk Assessment Instrument; OSIRIS=OSteoporosis Index of RISk; OST=Osteoporosis Self-Assessment Tool; OSTA=OST for Asians;

Appendix D Table 14. Outcomes From Included Studies for Diagnostic Accuracy (Key Question 2c)

SE=standard error; Sn=sensitivity; SOF=Study of Osteoporotic Fractures Score; SOFSURF=Study of Osteoporotic Fractures Research Group Study Utilizing Risk Factors; Sp=specificity; U.K.=United Kingdom; U.S.=United States; VA-FARA=electronic record adaptation of FRAX.

Appendix D Table 15. Outcomes From Included Studies for Key Question 2d

Study	Mean Length of Followup, Years	N	Participant Characteristics	Fracture Site	Results
Berry et al, 2013[246]	9.6 after repeat test	802	Mean age: 74.8 (SD 4.5) % women: 61	Unclear*	AUC* (95% CI) for baseline BMD 0.71 (0.65 to 0.78) AUC* (95% CI) for BMD % change 0.68 (0.62 to 0.75) AUC* (95% CI) for BMD baseline and % change 0.72 (0.66 to 0.79)
Crandall et al, 2020[249]	9.0 after repeat test	7,419	Mean age: 66.1 (SD 7.2) % women: 100	Hip fracture†	BMD at FN AUC (95% CI) for baseline BMD 0.71 (0.67 to 0.75) <65: 0.69 (0.57, 0.81) 65–74: 0.66 (0.60, 0.73) ≥75: 0.63 (0.56, 0.70) AUC (95% CI) for BMD % change 0.61 (0.56 to 0.65) <65: 0.63 (0.49, 0.77) 65–74: 0.54 (0.47, 0.61) ≥75: 0.60 (0.52, 0.68) AUC (95% CI) for BMD baseline and % change 0.73 (0.69 to 0.77) <65: 0.74 (0.61, 0.86) 65–74: 0.67 (0.61, 0.73) ≥75: 0.68 (0.62, 0.75)
				MOF†	BMD at FN AUC (95% CI) for baseline BMD 0.61 (0.59 to 0.63) <65: 0.59 (0.55, 0.62) 65–74: 0.61 (0.58, 0.64) ≥75: 0.60 (0.55, 0.65) AUC (95% CI) for BMD % change 0.53 (0.51 to 0.56) <65: 0.53 (0.49, 0.57) 65–74: 0.54 (0.51, 0.57) ≥75: 0.58 (0.53, 0.63) AUC (95% CI) for BMD baseline and % change 0.62 (0.60 to 0.64) <65: 0.59 (0.56, 0.63) 65–74: 0.61 (0.58, 0.64) ≥75: 0.60 (0.55, 0.65)

Appendix D Table 15. Outcomes From Included Studies for Key Question 2d

Study	Mean Length of Followup, Years	N	Participant Characteristics	Fracture Site	Results
Ensrud et al, 2022[250]	8.2 years after repeat test	3,651	Mean age: 72.3 (SD 5.1) at time of initial BMD	Hip fracture‡	BMD at TH AUC (95% CI) for baseline BMD 0.73 (0.69 to 0.76) AUC (95% CI) for BMD % change 0.68 (0.64 to 0.72) AUC (95% CI) for BMD baseline and % change 0.74 (0.71 to 0.77)
				Any clinical fracture‡	BMD at TH AUC (95% CI) for baseline BMD 0.64 (0.62 to 0.66) AUC (95% CI) for BMD % change 0.60 (0.58 to 0.62) AUC (95% CI) for BMD baseline and % change 0.64 (0.62 to 0.67)
				MOF‡	BMD at TH AUC (95% CI) for baseline BMD 0.68 (0.66 to 0.71) AUC (95% CI) for BMD % change 0.63 (0.61 to 0.66) AUC (95% CI) for BMD baseline and % change 0.69 (0.66 to 0.71)
Hillier et al, 2007[247]	11.4 total (5 years after repeat test)	4,124	Mean age: 74 (SD 4) % women: 100	Hip fracture§	AUC for baseline BMD (95% CI) 0.73 (CI, NR) AUC for BMD % change 0.68 (CI, NR) AUC for BMD baseline and % change 0.74 (CI, NR)
				Nonspine fracture§	AUC for baseline BMD (95% CI) 0.65 (CI ,NR) AUC for BMD % change 0.61 (CI, NR) AUC for BMD baseline and % change 0.65 (CI, NR)
				Spine fracture	AUC for baseline BMD (95% CI) 0.67 (CI, NR) AUC for BMD % change 0.62 (CI, NR) AUC for BMD baseline and % change 0.68 (CI, NR)
Leslie et al, 2017[248]	7.7	3,961	Mean age 60.4 (SD 9.6) % women: 100	MOF‖	Change in BMD; Gradient of risk (HR per SD increase (95% CI) over a mean of 4.1 years Total hip: 1.02 (0.89 to 1.17) Femoral neck: 0.93 (0.81 to 1.06) Spine:1.02 (0.89 to 1.18)

* Authors depicted two separate receiver operating characteristics curves: one for hip fracture and one for MOF, but only one set of AUC values were reported. AUC adjusted for age, sex, BMI, weight loss, and history of fracture measured at the time of the second BMD.

† Adjusted for current hormone use (yes/no), and Women's Health Initiative Study component (clinical trial/observational study). Major osteoporotic fractures included hip, spine, lower arm/wrist, and upper arm/shoulder.

‡ Adjusted for age, race/ethnicity, study enrollment site, prior fracture between baseline and 7-year BMD measurements, fall in past year, multimorbidity, score, physical activity, BMI, and percentage weight change between baseline and 7-year BMD measurements.

§ Adjusted for age and weight change.

‖ Major osteoporotic fracture defined as nontraumatic hip, clinical vertebral, forearm, and humerus fracture. HR adjusted for baseline fracture probability.

Abbreviations: AUC=area under the curve; BMD=bone mineral density; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; HR=hazard ratio; MOF=major osteoporotic fracture; N=number; NR=not reported; SD=standard deviation; U.S.=United States.

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Adachi et al, 2009[296] RCT Some concerns/Some concerns/Fair Alendronate 10 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 39/291 Placebo: 14/147 RR: NR ARD: NR Serious Adverse Events: Alendronate: 4/291 Placebo: 1/147 RR: 2.02 (95% CI, 0.23 to 17.91) ARD: NR Gastrointestinal Adverse Events: Alendronate Upper GI AE 66/291 Serious upper GI AE 59/291 Placebo Upper GI AE 30/147 Serious upper GI AE 19/147 Upper GI AE RR 1.11 (95% CI, 0.76 to 1.63) Serious upper GI AE RR 1.57 (95% CI, 0.97 to 2.53) Other Adverse Events: Any adverse event Alendronate: 166/291 Placebo: 76/147 RR: 1.10 (95% CI, 0.92 to 1.33) Dyspepsia Alendronate: 23/291 Placebo: 0/147 Esophageal spasm Alendronate: 1/291 Placebo: 0/147 Nonserious upper GI bleed Alendronate: 1/291 Placebo: 0/147

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Ascott-Evans et al, 2003[252] RCT Some concerns/Fair Alendronate 10 mg/day	Vertebral Fracture: (clinical fractures) Alendronate: 0/95 Placebo: 0/47 RR: NR ARD: NR Nonvertebral Fracture: Alendronate: 0/95 Placebo: 0/47 RR: RR not estimable ARD: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 10/95 Placebo: 10/49 RR: Calculated RR, 0.49 (95% CI, 0.22 to 1.11) ARD: NR Serious Adverse Events: Alendronate: NR Placebo: NR RR: NR ARD: NR Gastrointestinal Adverse Events: Alendronate: 15/95 Placebo: 6/49 Calculated RR, 1.24 (95% CI, 0.51 to 2.98) Other Adverse Events: Any clinical adverse event Alendronate: 60/95 Placebo: 30/49 Calculated RR: 0.99 (95% CI, 0.76 to 1.29)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Bell et al, 2002[276] RCT Some concerns/Fair Alendronate 10 mg/day	Any clinical fracture Alendronate: 1/33 Placebo: 3/32 RR: NR ARD: NR	NR	Discontinuation due to Adverse Events: Alendronate: 2/33 Placebo: 1/32 RR: NR ARD: NR Serious Adverse Events: Alendronate: 3/33 Placebo: 5/32 RR: NR ARD: NR Gastrointestinal Adverse Events: Alendronate: 14 Placebo: 11 RR: NR ARD: NR Other Adverse Events: Any adverse experience Alendronate: 30/33 Placebo: 30/32

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Bone et al, 1997[277] RCT Some concerns/Fair Alendronate, 1 mg/day Alendronate, 2.5 mg/day Alendronate, 5 mg/day	Vertebral Fracture: Alendronate 1 mg: 4/86 Alendronate 2.5 mg: 3/89 Alendronate 5 mg: 4/93 Placebo: 6/91 RR: NR ARD: NR Nonvertebral Fracture: Alendronate 1 mg: 15/86 Alendronate 2.5 mg: 9/89 Alendronate 5 mg: 9/93 Placebo: 16/91 RR: NR ARD: NR	NR	Discontinuation due to Adverse Events: Alendronate 1 mg: 8/86 Alendronate 2.5 mg: 8/89 Alendronate 5 mg: 13/93 Placebo: 9/91 RR: NR ARD: NR Gastrointestinal Adverse Events: No significant trend toward increasing frequency of upper gastrointestinal adverse experiences across treatment groups (data NR) Other Adverse Events: Number with adverse events suspected to be drug-related Alendronate 1 mg: 17/86 Alendronate 2.5 mg: 23/89 Alendronate 5 mg: 16/93 Placebo: 21/91 "No significant differences between groups" per study authors.

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Bone et al, 2008[285] RCT Some concerns/Fair Denosumab 60 mg every 6 months at baseline, 6, 12, and 18 months	Vertebral Fracture: Denosumab: 0/166 Placebo: 1/166 RR: NR ARD: NR Nonvertebral Fracture: Denosumab: 2/166 Placebo: 7/166 RR: NR ARD: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality Denosumab: 0/164 Placebo: 0/165 RR: NR	Discontinuation due to Adverse Events: Denosumab: 1/164 Placebo: 2/165 Calculated RR: 0.50 (95% CI, 0.05 to 5.49) ARD: NR Serious Adverse Events: Denosumab: 18/164 Placebo: 9/165 Calculated RR: 2.01 (95% CI, 0.93 to 4.35) ARD: NR Gastrointestinal Adverse Events: Denosumab: 2/164 Placebo: 0/165 RR: NR ARD: NR Other Adverse Events: Rash Denosumab: 14/164 Placebo: 5/165 Calculated RR: 2.82 (95% CI, 1.04 to 7.64) Serious infections Denosumab: 8/164 Placebo: 1/165 Calculated RR: 8.1 (95% CI, 1.02 to 63.6)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Boonen et al, 2012[251] RCT Low/Good Zoledronic acid 5-mg IV at baseline and 1 y	Vertebral Fracture: Zoledronic acid: 9/588 Placebo: 28/611 RR: 0.33 (95% CI, 0.16 to 0.70), 2 y (based on 24 months of n=553 for zoledronic acid and n=574 for placebo) ARD: NR Nonvertebral Fracture: Zoledronic acid: 5/588 Placebo: 8/611 RR: 0.65 (95% CI, 0.21 to 1.97), 2 y ARD: NR Hip Fracture: NR Other Fractures: Clinical fractures (vertebral and nonvertebral), 2 y Zoledronic acid: 6/588 Placebo: 11/611 RR: 0.57 (95% CI, 0.21 to 1.52)	All-Cause Mortality Zoledronic acid: 15/588 Placebo: 18/611 RR: 0.87 (95% CI, 0.44 to 1.70)	Discontinuation due to Adverse Events: Zoledronic acid: NR Placebo: NR RR: NR ARD: NR Serious Adverse Events: Zoledronic acid: 149/588 Placebo: 154/611 RR: 1.01 (95% CI, 0.83 to 1.22) ARD: NR Gastrointestinal Adverse Events: NR Other Adverse Events: Any Adverse Event: Zoledronic acid: 534/588 Placebo: 466/611 RR: 1.19 (95% CI, 1.13 to 1.25) Atrial fibrillation: Zoledronic acid: 7/588 Placebo: 5/611 RR: 1.45 (95% CI, 0.46 to 4.56) Myocardial infarction: Zoledronic acid: 9/588 Placebo: 2/611 RR: 4.68 (95% CI, 1.015 to 21.55) Osteonecrosis of the jaw: Zoledronic acid: 0/588 Placebo: 0/611 RR: NR Arthralgia: Zoledronic acid: 123/588 Placebo: 68/611 RR: 1.88 (95% CI, 1.43 to 2.47) Myalgia: Zoledronic acid: 129/588 Placebo: 25/611 RR: 5.20 (95% CI, 3.44 to 7.86)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Chapurlat et al, 2013[288] RCT Some concerns/Fair Ibandronate 150 mg/month	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Ibandronate: 4/71 Placebo: 6/76 RR: 0.71 (95% CI, 0.21 to 2.42) Serious Adverse Events: Ibandronate: 15/71 Placebo: 13/76 RR: 1.23 (95% CI, 0.63 to 2.41) Gastrointestinal Adverse Events: NR Other Adverse Events: NR
Chesnut et al, 1995[253] RCT Some concerns/Fair Alendronate 5 mg/day; alendronate 10 mg/day; alendronate 40 mg/day (for 3 months then 2.5 mg/day for 21 months); alendronate 20 mg/day (for 1 y then placebo for 1 y); alendronate 40 mg/day (for 1 y then placebo for 1 y)	Vertebral Fracture: Alendronate 5 mg/day: 0/32 Alendronate 10 mg/day: 0/30 Alendronate 40 mg/day (for 3 months then 2.5 mg/day for 21 months): 0/32 Alendronate 20 mg/day (for 1 y then placebo for 1 y): 0/32 Alendronate 40 mg/day (for 1 y then placebo for 1 y): 0/31 Placebo: 0/31 RR: NR ARD: NR Nonvertebral Fracture: Alendronate: Unclear Placebo: Unclear RR: NR ARD: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality	Discontinuation due to Adverse Events: Alendronate: Unclear Placebo: NR RR: NR ARD: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Alendronate: 9 withdrew due to adverse GI events Placebo: NR RR: NR ARD: NR Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Cryer et al, 2005[297] RCT Some concerns/Fair Alendronate 70 mg weekly	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 10/224 Placebo: 18/230 RR: 0.57 (95% CI, 0.27 to 1.21) ARD: NR Serious Adverse Events: Alendronate: 9/224 Placebo: 8/230 RR: 1.16 (95% CI, 0.45 to 2.94) ARD: NR Gastrointestinal Adverse Events: Any upper GI event Alendronate: 79/224 Placebo: 86/230 Dyspepsia Alendronate: 11/224 Placebo: 9/230 Abdominal pain Alendronate: 6/224 Placebo: 3/230 GERD Alendronate: 3/224 Placebo: 3/230 Any upper GI event: Calculated RR 0.94 (95% CI, 0.74 to 1.20) Dyspepsia: Calculated RR 1.26 (95% CI, 0.53 to 2.97) Abdominal pain: Calculated RR 2.05 (95% CI, 0.52 to 8.11) GERD: Calculated RR 1.03 (95% CI, 0.21 to 5.03) Other Adverse Events: Any AE Alendronate: 141/224 Placebo: 120/230 Calculated RR, 1.21 (95% CI, 1.03 to 1.42)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283] FREEDOM (Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months) Trial RCT Some concerns/Fair Denosumab 60 mg/6 months subcutaneously	Followup time frame 3 years Vertebral Fracture: Denosumab: 86/3,702 (2.3%) Placebo: 264/3,691 (7.2%) ARD per 1,000: 48 fewer (95% CI, from 58 fewer to 39 fewer) RR: 0.32 (95% CI, 0.26 to 0.41) Clinical Vertebral Fractures: Denosumab: 29 (0.8) Placebo: 92 (2.6) HR: 0.31 (95% CI, 0.20 to 0.47) ARD per 1,000 participants: 17 fewer (from 23 fewer to 11 fewer) Nonvertebral Fracture: Denosumab: 238/3,902 (6.5%) Placebo: 293/3,906 (8.0%) HR: 0.80 (95% CI, 0.67 to 0.95) ARD per 1,000 participants: 15 fewer (from 27 fewer to 3 fewer) Hip Fracture: Denosumab: 26/3,902 (0.7%) Placebo: 43/3,906 (1.2%) HR: 0.60 (95% CI, 0.37 to 0.97) Calculated ARD per 1,000 participants: 3 fewer (from 7 fewer to 1 more)	All-Cause Mortality Denosumab: 70/3,886 (1.8%) Placebo: 90/ 3,876 (2.3%) Calculated RR, 0.78 (95% CI, 0.57 to 1.06) Calculated ARD per 1,000 participants: 5 fewer (from 10 fewer to 1 more)	Discontinuation due to Adverse Events: Denosumab: 93/3,886 Placebo: 81/3,876 Calculated RR, 1.15 (95% CI, 0.85 to 1.54) ARD: NR Serious Adverse Events: Denosumab: 1,004/3,886 Placebo: 972/3,876 Calculated RR, 1.03 (95% CI, 0.95 to 1.11) ARD: NR Gastrointestinal Adverse Events: NR Other Adverse Events: Osteonecrosis of the jaw Denosumab: 0/3,886 Placebo: 0/3,876 RR not calculable Cardiovascular events Denosumab: 186/3,886 Placebo: 178/3,876 Calculated RR, 1.04 (95% CI, 0.85 to 1.27) Eczema Denosumab: 118/3,886 Placebo: 65/3,876 Calculated RR, 1.81 (95% CI, 1.34 to 2.44) Serious infections Denosumab: 159/3,886 Placebo: 133/3,876 Calculated RR 1.19 (95% CI, 0.95 to 1.49) Serious skin infections (cellulitis and erysipelas) Denosumab: 15/3,886 Placebo: 1/3,876 Calculated RR, 14.96 (95% CI, 1.98 to 113.21)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283] (continued)	Other Fractures: Multiple (≥2) new vertebral fractures: Denosumab: 23/3,702 (0.6%) Placebo: 59/3,691(1.6%) RR, 0.39 (95% CI, 0.24 to 0.63) ARD per 1,000 participants: 1 fewer (95% CI, from 15 fewer to 5 fewer) Wrist fractures: Denosumab: 90/3,902 Placebo: 107/3,906 RR:16% (95% CI, -11% to 37%) Subgroup Analyses: No significant interaction was observed between treatment effect and baseline fracture probability (p=0.72), However, a cubic spline function was found to give a significantly (p<0.001) better fit for the relation between treatment effect baseline fracture probability In subgroup analyses based on history of prior fracture at baseline compared with placebo, denosumab had similar effects in women without a prior fragility fracture (RRR 40%, p<0.0001) as women with a history of a prior fragility fracture (RRR 39%, p<0.0001) **Comments:** The subgroup analyses are from McCloskey et al[282] and Palacios et al[425]		

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255] Fracture Intervention Trial (FIT) RCT Low/Good Alendronate 5 mg/day for 2 y then 10 mg/day for 1 y for those without existing vertebral fractures, and 2 to 2.6 y for those with vertebral fractures at baseline	Vertebral Fracture: (radiographic) Alendronate: 43/2,214 (2.1%) Placebo: 78/2,218 (3.8%) HR: 0.56 (95% CI, 0.39 to 0.80) ARD: NR Nonvertebral Fracture: Alendronate: 261/2,214 (11.8%) Placebo: 294/2,218 (13.3%) HR: 0.88 (95% CI, 0.74 to 1.04) ARD: NR Hip Fracture: Alendronate: 19/2,214 (0.9%) Placebo: 24/2,218 (1.1%) HR: 0.79 (95% CI, 0.43 to 1.44) ARD: NR Other Fractures: Wrist fractures Alendronate: 83/2,214 (3.7%) Placebo: 70/2,218 (3.2%) HR, 1.19 (95% CI, 0.87 to 1.62) ARD: NR Clinical fractures (primary endpoint) defined as clinical vertebral, hip, wrist, and other sites excluding face and skull Alendronate: 272/2,214 (12.3%) Placebo: 312/2,218 (14.1%) HR: 0.86 (95% CI, 0.73 to 1.01) ARD: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 221/2,214 Placebo: 227/2,218 HR: 0.98 (95% CI, 0.80 to 1.16) ARD: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Bauer 2000[289] (all FIT participants) Any upper GI AE Alendronate: 1,536/3,226 Placebo: 1,490/3,223 Calculated RR (95% CI): 1.03 (0.98 to 1.08) Any gastric or duodenal AE Alendronate: 130/3,226 Placebo: 129/3,223 Calculated RR (95% CI): 1.01 (0.79 to 1.28) Gastritis Alendronate: 82/3,226 Placebo: 75/3,223 Calculated RR (95% CI): 1.05 (0.90 to 1.22) Any gastric or duodenal perforations, ulcers, bleeding Alendronate: 53/3,226 Placebo: 61/3,223 Calculated RR (95% CI): 0.87 (0.60 to 1.25) Any esophageal AE Alendronate: 322/3,226

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255] (continued)	Subgroup Analyses: Quandt, 2005[255] (FIT participants with osteopenia) Clinical vertebral fracture: Alendronate: 12/1,878 Placebo: 29/1,859 RR, 0.40 (95% CI, 0.19 to 0.76) Radiographic vertebral fractures Alendronate: 48/1,775 Placebo: 81/1,757 RR, 0.57 (95% CI, 0.41 to 0.81)		Placebo: 303/3,223 Calculated RR (95% CI): 1.59 (1.34 to 1.89) Acid regurgitation/reflux Alendronate: 279/3,226 Placebo: 269/3,223 Calculated RR (95% CI): 1.04 (0.88 to 1.22) Other Adverse Events: Cummings, 2007[290] (all participants): Serious atrial fibrillation: Alendronate: 47/3,236 Placebo: 31/3,223 HR, 1.51 (95% CI, 0.96 to 2.37) Any atrial fibrillation: Alendronate: 81/3,236 Placebo: 71/3,226 HR, 1.14 (95% CI, 0.83 to 1.56)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Devogelaer et al, 1996[302] RCT Some concerns/Fair Alendronate 5 mg/d Alendronate 10 mg/d Alendronate 20 mg/d for 2 y, then 5 mg/d for 1 y All groups received 500 mg calcium carbonate qd	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate 5 mg: 8/104 (7.7) Alendronate 10 mg: 3/102 2.9) Alendronate 20 mg/5 mg: 9/105 (8.6) Placebo: 11/205 (5.4) Calculated RR: 0.55 (95% CI, 0.16 to 1.92) ARD: NR Serious Adverse Events: Alendronate 5 mg: 14/104 (13.5) Alendronate 10 mg: 7/102 (6.9) Alendronate 20 mg/5 mg: 18/105 (17.1) Placebo: 34/205 (16.6) RR: NR ARD: NR Gastrointestinal Adverse Events: GI AEs considered to be drug-related Alendronate 5 mg: 18/104 (17.3) Alendronate 10 mg: 15/102 (14.7) Alendronate 20 mg/5 mg: 19/105 (18.1) Placebo: 35/205 (17.1) RR: NR ARD: NR Other Adverse Events: Overall clinical AE Alendronate 5 mg: 89/104 (85.6) Alendronate 10 mg: 84/102 (82.4) Alendronate 20mg/5mg: 89/105 (84.8) Placebo: 177/205 (86.3)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Eisman et al, 2004[298] RCT Low/Good Alendronate 70 mg weekly	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 11/225 Placebo: 6/224 RR: NR ARD: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Any upper GI event Alendronate: 22/225 Placebo: 21/224 RR (95% CI): Any upper GI event: 1.04 (0.59 to 1.84) Abdominal pain Alendronate: 2/225 Placebo: 2/224 RR (95% CI): 1.00 (0.14 to 7.01) Dyspepsia Alendronate: 2/225 Placebo: 1/224 RR (95% CI): 1.99 (0.18 to 21.80) Gastritis Alendronate: 0/225 Placebo: 2/224 Esophogeal ulcer Alendronate: 0/225 Placebo: 1/224 GERD Alendronate: 0/225 Placebo: 1/224

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Eisman et al, 2004[298] (continued)			Other Adverse Events: Any AE Alendronate: 91/225 Placebo: 86/224 RR 1.05 (95% CI, 0.84 to 1.33) ARD: 2.1% (95% CI, -6.9% to 11.0%) Discontinuations due to drug-related upper GI adverse events Alendronate: 6/225 Placebo: 5/224 ARD 0.4% (95% CI, -5.1% to 5.9%)
Greenspan et al, 2002[299] RCT Some concerns/Fair Alendronate 70 mg weekly	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 10/224 Placebo: 11/226 Calculated RR, 0.92 (95% CI, 0.40 to 2.12) ARD: NR Serious Adverse Events: Alendronate: 28/224 Placebo: 34/226 Calculated RR, 0.83 (95% CI, 0.52 to 1.32) ARD: NR Gastrointestinal Adverse Events: Total upper GI events Alendronate: 25/224 Placebo: 30/226 RR (95% CI): 0.84 (0.51 to 1.38)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Greenspan et al, 2002[299] (continued)			Abdominal pain Alendronate: 7/224 Placebo: 8/226 RR (95% CI): 0.88 (0.33 to 2.39) Dyspepsia Alendronate: 4/224 Placebo: 6/226 RR (95% CI): 0.67 (0.19 to 2.35) GERD Alendronate: 3/224 Placebo: 1/226 RR (95% CI): 3.03 (0.32 to 28.88) Duodenal ulcer Alendronate: 1/224 Placebo: 0/226 RR NR Gastritis Alendronate: 1/224 Placebo: 0/226 RR NR Other Adverse Event: Any adverse event Alendronate: 104/224 Placebo: 97/226 RR 1.08 (95% CI, 0.88 to 1.33)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Greenspan et al, 2003[300] RCT Low/Good Alendronate 10 mg/day Study included 2 other arms that included estrogen that are not relevant to this update	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Esophagitis Alendronate: 26/93 Placebo: 21/93 Calculated RR: 1.24 (95% CI, 0.75 to 2.04) ARD: NR Other Adverse Events: Myocardial infarction Alendronate: 2/93 Placebo: 1/93 Calculated RR: 2.0 (95% CI, 0.18 to 21.68)
Grey et al, 2010[262] Grey et al. 2009[263] RCT Some concerns/Fair Zoledronic acid 5-mg IV (onetime dose)	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: Zoledronate (finger, rib, forearm, and fibula): 4/25 Placebo (toe, forearm): 2/25 Calculated RR, 2.0 (95% CI, 0.40 to 9.95)	All-Cause Mortality NR	Discontinuation due to Adverse Events: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: NR Other Adverse Events: Atrial fibrillation Zoledronate: 0/25 Placebo: 0/25 RR NR Osteonecrosis of the jaw Zoledronate: 0/25 Placebo: 0/25 RR NR Symptomatic hypocalcemia Zoledronate: 0/25 Placebo: 0/25 RR NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274] RCT Some concerns/Fair Zoledronic acid 1 mg, 2.5 mg, and 5 mg (single-dose IV)	Vertebral Fracture: NR Nonvertebral Fracture: At 1 y Zoledronate 1 mg: 0/45 Zoledronate 2.5 mg: 2/45 Zoledronate 5 mg: 1/45 Placebo: 2/45 At 2 y Zoledronate 1 mg: 1/45 Zoledronate 2.5 mg: 2/45 Zoledronate 5 mg: 2/45 Placebo: 3/45 RR: NR ARD: NR Hip Fracture: NR Other Fractures: NR **Comments:** Did not include fracture data from 5-y extension study because participants were unblinded during that portion. Nonvertebral fracture sites included forearm, finger, metacarpal, metatarsal, hand, tibia. No vertebral fractures were reported.	All-Cause Mortality NR	Discontinuation due to Adverse Events: Zoledronate (all dose groups): 1 y: 0/45 2 y: 0/45 Placebo: 1 y: 0/45 2 y: 0/45 RR: NR ARD: NR Serious Adverse Events: Osteonecrosis of the jaw: 0 in all dose study arms at 1 and 2 years followup Gastrointestinal Adverse Events: GI acute phase reactions at 1 with postinfusion Zoledronate 1 mg: 9/45 Zoledronate 2.5 mg: 13/45 Zoledronate 5 mg: 13/45 Placebo: 5/45 OR 1 mg vs. placebo: 2.0 (95% CI, 0.6 to 6.6) OR 2.5 mg vs. placebo: 3.3 (95% CI, 1.1 to 10.3) OR 5 mg vs. placebo: 3.3 (95% CI, 1.1 to 10.3) Other Adverse Events: Atrial fibrillation: 0 in all active-dose study arms at 1 and 2 years followup

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Hosking et al, 2003[267] RCT Some concerns/Fair Alendronate 70 mg weekly Risedronate 5 mg daily	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: Clinically diagnosed vertebral or nonvertebral Alendronate: 6/172 Placebo: 2/89 RR, 1.55 (0.31 to 7.53)	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 31/219 Risedronate: 31/222 Placebo: 12/108 RR Alendronate vs. placebo: Calculated RR, 1.27 (95% CI, 0.68 to 2.38) Risedronate vs. placebo: Calculated RR, 1.26 (95% CI, 0.67 to 2.35) ARD: NR Serious Adverse Events: Alendronate: 17/219 Risedronate: 15/222 Placebo: 12/108 RR Alendronate vs. placebo: Calculated RR, 0.70 (95% CI, 0.35 to 1.41) Risdedroante vs. placebo: Calculated RR, 0.61 (95% CI, 0.29 to 1.25) ARD: NR Gastrointestinal Adverse Events: Alendronate: 62/219 Risedronate: 61/222 Placebo: 29/108 RR Alendronate vs. placebo: Calculated RR, 1.05 (95% CI, 0.72 to 1.54) Risedronate vs. placebo: Calculated RR, 1.02 (95% CI, 0.70 to 1.49) ARD: NR Other Adverse Events: Any AE Alendronate: 169/219 Risedronate: 169/222 Placebo: 76/108 Alendronate vs. placebo: Calculated RR, 1.10 (95% CI, 0.95 to 1.26)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Johnell et al, 2002[294] RCT Some concerns/Fair Alendronate, 10 mg/day Raloxifene, 60 mg/day (not included in this review)	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 8/83 Placebo: 4/82 RR: Alendronate: Calculated RR, 1.98 (95% CI, 0.62 to 6.30) ARD: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Alendronate: 9/83 Placebo: 5/82 Calculated RR, 1.78 (95% CI, 0.62 to 5.08) ARD: NR Other Adverse Events: Chest pain substernal Alendronate: 6/82 Placebo: 2/82 Calculated RR, 2.96 (95% CI, 0.62 to 14.26) Vasodilation Alendronate: 4/82 Placebo: 4/82 Sweating Alendronate 2/82 Placebo: 2/82

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Koh et al, 2016[286] RCT Some concerns/Fair Denosumab 60 mg (single dose IV)	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality Denosumab: 1/69 Placebo: 0/66 RR: NR	Discontinuation due to Adverse Events: Denosumab: 0/69 Placebo: 0/66 RR: NR ARD: NR Serious Adverse Events: Denosumab: 2/69 Placebo: 1/66 RR: NR ARD: NR Gastrointestinal Adverse Events: Constipation Denosumab: 5/69 Placebo: 2/66 RR: NR Gastritis Denosumab: 3/69 Placebo: 1/66 RR: NR Other Adverse Events: Any AE Denosumab: 38/69 Placebo: 32/66 RR: NR Treatment-related AEs Denosumab: 2/69 Placebo: 1/66 RR: NR Osteonecrosis of the jaw Denosumab: 0/69 Placebo: 0/66 RR: NR Atypical femur fracture Denosumab: 0/69 Placebo: 0/66 RR: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Lewiecki et al, 2007[284] McClung et al, 2006[304] RCT Some concerns/Fair Denosumab 6 mg, 14 mg, or 30 mg every 3 months or denosumab 14 mg, 60 mg, 100 mg, or 210 mg every 6 month, alternating with placebo to maintain blinding	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: Osteoporotic fractures Denosumab: 12/314 Placebo: 0/46 Calculated RR: 3.73 (95% CI, 0.22 to 61.96) Clinical fractures Denosumab: 21/314 Placebo: 1/46 Calculated RR: 1.58 (95% CI, 0.68 to 3.63)	All-Cause Mortality Denosumab: 1/314 Placebo: 0/46 RR: NR	Discontinuation due to Adverse Events: Denosumab: 11/314 Placebo: 1/46 Calculated RR, 1.61 (95% CI, 0.21 to 12.19) ARD: NR Serious Adverse Events: Denosumab: 42/314 Placebo: 4/46 Calculated RR, 1.54 (95% CI, 0.58 to 4.09) ARD: NR Gastrointestinal Adverse Events: Denosumab: 1/314 Placebo: 0/46 RR not calculated Other Adverse Events: Cardiac disorder Denosumab: 6/314 Placebo: 2/46 Calculated RR, 0.45 (95% CI, 0.02 to 10.83) Serious infections Denosumab: 6/314 Placebo: 0/46 Calculated RR: 3.5 (95% CI, 0.07 to 190.8)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	**Fracture Outcomes**	**Mortality**	**Harm Outcomes**
Liberman et al, 1995[256] RCT Some concerns/Fair Alendronate 5 or 10 mg/day for 3 years or 20 mg/day for 2 years followed by 5 mg/day for 1 year	Vertebral Fracture: (radiographic) Alendronate: 4/384 Placebo: 5/253 RR: 0.53 (0.14 to 1.94) ARD: NR Nonvertebral Fracture: Alendronate: 45/597 Placebo: 38/397 RR: 0.79 (0.52 to 1.22) ARD: NR Hip Fracture: Alendronate: 1/597 Placebo: 3/397 RR: NR ARD: NR Other Fractures: NR **Comments:** Results are for all doses of alendronate combined. The vertebral fractures were morphometric, not clinical fractures. The fractures reported here are only among the women without vertebral fractures at baseline.	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate: 35/597 Placebo: 24/397 RR: 0.97 (0.50 to 1.60) Serious Adverse Events: NR Gastrointestinal Adverse Events: Dyspepsia Alendronate: 7/196 Placebo: 14/397 RR: 1.01 (95% CI, 0.42 to 2.37) Other Adverse Events: Abdominal pain Alendronate: 13/196 Placebo: 19/397 RR: 1.32 (95% CI, 0.66 to 2.62)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2001[257] RCT Some concerns/Fair Risedronate 2.5 mg/d Risedronate 5 mg/d	Vertebral Fracture: NR Nonvertebral Fracture: Risedronate: 582/6,197 Placebo: 351/3,134 RR: 0.8 (95% CI, 0.7 to 1.0) ARD: NR Hip Fracture: Risedronate: 137/6,197 Placebo: 95/3,134 Calculated RR, 0.73 (95% CI, 0.56 to 0.94) ARD: NR Other Fractures: NR Subgroup Analyses: Subgroup ages 70 to 79 years with osteoporosis (n=5,445) Hip Fx RR: 0.6 (95% CI, 0.4 to 0.9) Nonvertebral fx RR: 0.8 (95% CI, 0.7 to 1.0) Subgroup ages 70 to 79 years without prevalent vertebral fracture (n=2,648) Risedronate: 14/1772 Placebo: 12/875 Hip Fx RR: 0.6 (95% CI, 0.3 to 1.2) Nonvertebral fx RR: NR Subgroup age ≥80 years with ≥1 clinical risk factor Hip Fx RR: 0.8 (95% CI, 0.6 to 1.2) Nonvertebral Fx RR: NR, p=0.43	All-Cause Mortality NR	Discontinuation due to Adverse Events: Risedronate: 550/3,104 Placebo: 564/3,134 Calculated RR, 0.98 (95% CI, 0.89 to 1.10) ARD: NR Serious Adverse Events: Risedronate: 943/3,104 Placebo: 973/3,134 Calculated RR, 0.98 (95% CI, 0.91 to 1.05) ARD: NR Gastrointestinal Adverse Events: Risedronate: 657/3,104 Placebo: 684/3,134 Calculated RR, 0.91 (95% CI, 0.88 to 1.07) RR: NR ARD: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2009[287] RCT Some concerns/Fair Zoledronic acid 5-mg IV, at baseline and at 1y Zoledronic acid 5-mg IV, at baseline only	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: NR Serious Adverse Events: Zoledronic acid 5-mg IV, at baseline and at 1 y: 19/198 Zoledronic acid 5-mg IV, at baseline only: 19/181 Placebo: 23/202 RR: Zoledronic acid 5-mg IV, at baseline and at 1 y: 0.75 (95% CI, 0.42 to 1.37) Zoledronic acid 5-mg IV, at baseline only: 1.0 (95% CI, 0.58 to 1.78) Gastrointestinal Adverse Events: NR Other Adverse Events: Total adverse events Zoledronic acid 5-mg IV, at baseline and at 1 y: 186/198 Zoledronic acid 5-mg IV, at baseline only: 173/181 Placebo: 186/202 RR (Zoledronic acid 5-mg IV, at baseline and at 1 y/placebo): 1.02 (95% CI, 0.96 to 1.07) RR (Zoledronic acid 5-mg IV, at baseline only/placebo): 1.04 (95% CI, 0.99 to 1.09) Myalgia Zoledronic acid 5-mg IV, at baseline and at 1 y: 38/198 Zoledronic acid 5-mg IV, at baseline only: 41/181 Placebo: 14/202 RR (Zoledronic acid 5-mg IV, at baseline and at 1 y/placebo) 2.77 (95% CI, 1.55 to 4.95) RR (Zoledronic acid 5-mg IV, at baseline only/placebo) 3.27 (95% CI, 1.84 to 5.79)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2009[287] (continued)			Arthralgia Zoledronic acid 5-mg IV, at baseline and at 1 y: 54/198; Zoledronic acid 5-mg IV, at baseline only: 34/181; Placebo: 39/202 RR (Zoledronic acid 5-mg IV, at baseline and at 1 y/Placebo): 1.41 (95% CI, 0.98 to 2.03) RR (Zoledronic acid 5-mg IV, at baseline only/placebo): 0.97 (95% CI, 0.64 to 1.47) Osteonecrosis of the jaw Zoledronic acid 5-mg IV, at baseline and at 1 y: 0/198 Zoledronic acid 5-mg IV, at baseline only: 0/181 Placebo: 0/202 Atrial fibrillation Zoledronic acid 5-mg IV, at baseline and at 1 y: 0/198 Zoledronic acid 5-mg IV, at baseline only: 0/181 Placebo: 0/202
McClung et al., 2009[275] RCT Fair Ibandronate 150 mg/ month Daily vitamin D (400 IU) and calcium (500 mg) supplements	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: Clinical fractures (all associated with trauma Ibandronate: 2/77 (2.6) (radius, upper limb) Placebo: 2/83 (2.4) (both in foot)	All-Cause Mortality Ibandronate: 0/7 (0) Placebo: 0/83 (0) Calculated RR: 1.00 (95% CI, 0.02 to 49.93)	Discontinuation due to Adverse Events: Ibandronate: 7/77 (9.1) Placebo: 3/83 (3.6) RR: NR ARD: NR Serious Adverse Events: Ibandronate: 3/77 (3.9) Placebo: 1/83 (1.2) RR: NR ARD: NR Gastrointestinal Adverse Events: Ibandronate: 24/77 (31.2) Placebo: 20/83 (24.1) RR: NR ARD: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2004[291] RCT Some concerns/Fair Ibandronate 0.5 mg/day Ibandronate 1.0 mg/day Ibandronate 2.5 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Any withdrawals because of AEs: Ibandronate 0.5 mg: 5/161 (3.1%) Ibandronate 1.0 mg: 5/165 (3.0%) Ibandronate 2.5 mg: 7/163 (4.3%) Placebo: 9/159 (5.7%) RR: Ibandronate 0.5 mg: Calculated RR, 0.55 (95% CI, 0.19 to 1.60) Ibandronate 1.0 mg: Calculated RR, 0.54 (95% CI, 0.18 to 1.56) Ibandronate 2.5 mg: Calculated RR, 0.76 (95% CI, 0.29 to 1.99) Serious Adverse Events: Ibandronate 0.5 mg: 6/161 Ibandronate 1.0 mg: 13/165 Ibandronate 2.5 mg: 5/163 Placebo: 8/159 Ibandronate 0.5 mg: Calculated RR, 0.74 (95% CI, 0.26 to 2.09) Ibandronate 1.0 mg: Calculated RR, 1.57 (95% CI, 0.67 to 3.68) Ibandronate 2.5 mg: Calculated RR, 0.61 (95% CI, 0.20 to 1.82) Any drug-related serious AEs: Ibandronate 0.5 mg: 0/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 0/163 Placebo: 0/159 RR not calculable

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2004[291] (continued)			Gastrointestinal Adverse Events: Dyspepsia Ibandronate 0.5 mg: 16/161 Ibandronate 1.0 mg: 14/165 Ibandronate 2.5 mg: 15/163 Placebo: 14/159 Ibandronate 0.5 mg: Calculated RR, 1.13 (95% CI, 0.57 to 2.23) Ibandronate 1.0 mg: Calculated RR, 0.96 (95% CI, 0.47 to 1.96) Ibandronate 2.5 mg: Calculated RR, 1.05 (95% CI, 0.52 to 2.09) Gastroenteritis Ibandronate 0.5 mg: 9/161 Ibandronate 1.0 mg: 4/165 Ibandronate 2.5 mg: 5/163 Placebo: 6/159 Ibandronate 0.5 mg: Calculated RR, 1.48 (95% CI, 0.54 to 4.07) Ibandronate 1.0 mg: Calculated RR, 0.64 (95% CI, 0.18 to 2.23) Ibandronate 2.5 mg: Calculated RR, 0.81 (95% CI, 0.25 to 2.61) Nausea Ibandronate 0.5 mg: 6/161 Ibandronate 1.0 mg: 1/165 Ibandronate 2.5 mg: 4/163 Placebo: 3/159 Ibandronate 0.5 mg: Calculated RR, 1.98 (95% CI, 0.50 to 7.76) Ibandronate 1.0 mg: Calculated RR, 0.32 (95% CI, 0.03 to 3.06) Ibandronate 2.5 mg: Calculated RR, 1.30 (95% CI, 0.30 to 5.72)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2004[291] (continued)			GI pain Ibandronate 0.5 mg: 2/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 4/163 Placebo: 4/159 Ibandronate 0.5 mg: Calculated RR, 0.49 (95% CI, 0.09 to 2.66) Ibandronate 1.0 mg: Calculated RR, 0.11 (95% CI, 0.01 to 1.98) Ibandronate 2.5 mg: Calculated RR, 0.98 (95% CI, 0.25 to 3.83) GI disorder Ibandronate 0.5 mg: 1/161 Ibandronate 1.0 mg: 2/165 Ibandronate 2.5 mg: 0/163 Placebo: 3/159 Ibandronate 0.5 mg: Calculated RR, 0.33 (95% CI, 0.03 to 3.13) Ibandronate 1.0 mg: Calculated RR, 0.64 (95% CI, 0.11 to 3.79) Ibandronate 2.5 mg: Calculated RR, 0.14 (95% CI, 0.01 to 2.68) Gastritis Ibandronate 0.5 mg: 0/161 Ibandronate 1.0 mg: 1/165 Ibandronate 2.5 mg: 2/163 Placebo: 1/159 Ibandronate 0.5 mg: Calculated RR, 0.33 (95% CI, 0.01 to 8.02) Ibandronate 1.0 mg: Calculated RR, 0.96 (95% CI, 0.06 to 15.28) Ibandronate 2.5 mg: Calculated RR, 1.95 (95% CI, 0.18 to 21.30)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2004[291] (continued)			Dysphagia Ibandronate 0.5 mg: 2/161 Ibandronate 1.0 mg: 1/165 Ibandronate 2.5 mg: 1/163 Placebo: 0/159 Ibandronate 0.5 mg: Calculated RR, 4.94 (95% CI, 0.24 to 102.06) Ibandronate 1.0 mg: Calculated RR, 2.89 (95% CI, 0.12 to 70.46) Ibandronate 2.5 mg: Calculated RR, 2.91 (95% CI, 0.12 to 71.32) Vomiting Ibandronate 0.5 mg: 2/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 1/163 Placebo: 0/159 Ibandronate 0.5 mg: Calculated RR, 4.94 (95% CI, 0.24 to 102.06) Ibandronate 1.0 mg: RR not calculable Ibandronate 2.5 mg: Calculated RR, 2.92 (95% CI, 0.12 to 71.32) Esophagitis Ibandronate 0.5 mg: 1/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 1/163 Placebo: 1/159 Ibandronate 0.5 mg: Calculated RR, 0.99 (95% CI, 0.06 to 15.65) Ibandronate 1.0 mg: Calculated RR, 0.32 (95% CI, 0.01 to 7.83) Ibandronate 2.5 mg: Calculated RR, 0.98 (95% CI, 0.06 to 15.46)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
McClung et al, 2004[291] (continued)			GI carcinoma Ibandronate 0.5 mg: 0/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 1/163 Placebo: 0/159 Ibandronate 0.5 mg: RR not calculable Ibandronate 1.0 mg: RR not calculable Ibandronate 2.5 mg: Calculated RR, 0.98 (95% CI, 0.02 to 49.17) GI hemorrhage Ibandronate 0.5 mg: 0/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 0/163 Placebo: 1/159 Ibandronate 0.5 mg: Calculated RR, 0.33 (95% CI, 0.01 to 8.02) Ibandronate 1.0 mg: Calculated RR, 0.32 (95% CI, 0.01 to 7.83) Ibandronate 2.5 mg: Calculated RR, 0.33 (95% CI, 0.01 to 7.92) Hemorrhage gastritis Ibandronate 0.5 mg: 1/161 Ibandronate 1.0 mg: 0/165 Ibandronate 2.5 mg: 0/163 Placebo: 0/159 Ibandronate 0.5 mg: Calculated RR, 2.96 (95% CI, 0.12 to 72.20) Ibandronate 1.0 mg: RR not calculable Ibandronate 2.5 mg: RR not calculable

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Mortensen et al, 1998[258] RCT Some concerns/Fair Risedronate 5 mg cyclic (daily for first 2 weeks of every month, then placebo daily for the rest of the month); Risedronate 5 mg/day	Vertebral Fracture: (radiographic) Cyclic risedronate: 1/38 Daily risedronate: 1/37 Placebo: 0/36 Calculated RR for daily risedronate, 2.97 (95% CI, 0.12 to 71.73), 1 y ARD: NR Nonvertebral Fracture: Cyclic risedronate: 3/38 Daily risedronate: 0/37 Placebo: 3/36 Calculated RR for daily risedronate, 0.14 (95% CI, 0.01 to 2.59), 3 y ARD: NR Hip Fracture: Cyclic risedronate: 0/38 Daily risedronate: 0/37 Placebo: 0/36 RR: RR not estimable ARD: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Cyclic risedronate: 3/38 Daily risedronate: 2/37 Placebo: 3/36 RR: NR ARD: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Dyspepsia Cyclic risedronate: 9/38 Daily risedronate: 6/37 Placebo:10/36 Calculated RR, 0.59 (95% CI, 0.24 to 1.44) Other Adverse Events: Abdominal pain Cyclic risedronate: 4/38 Daily risedronate: 3/37 Placebo: 4/36 Calculated RR, 0.73 (95% CI, 0.18 to 3.04)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Nakamura et al, 2012[279] RCT Some concerns/Fair Denosumab 14 mg subcutaneously every 6 months for 12 months Denosumab 60 mg subcutaneously every 6 months for 12 months Denosumab 100 mg subcutaneously every 6 months for 12 months or placebo every 6 months for 12 months	Vertebral Fracture: (radiographic or clinical) Denosumab: 0/157 Placebo: 0/55 RR: NR ARD: NR Nonvertebral Fracture: Denosumab: NR Placebo: NR RR: NR ARD: NR Hip Fracture: Denosumab: NR Placebo: NR RR: NR ARD: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: NR Serious Adverse Events: Denosumab 14 mg subcutaneously every 6 months for 12 months: 6/53 Denosumab 60 mg subcutaneously every 6 months for 12 months: 4/54 Denosumab 100 mg subcutaneously every 6 months for 12 months: 2/51 Placebo every 6 months for 12 months: 4/54 RR: NR ARD: NR Gastrointestinal Adverse Events: Serious GI disorders Denosumab 14 mg subcutaneously every 6 months for 12 months: 3/53 Denosumab 60 mg subcutaneously every 6 months for 12 months: 0/54 Denosumab 100 mg subcutaneously every 6 months for 12 months: 1/51 Placebo every 6 months for 12 months: 1/54 RR: NR Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Orwoll et al, 2012[278] ADAMO RCT Fair Denosumab 60 mg/6 months	Vertebral Fracture: Denosumab: 0/121 (0) Placebo: 1/121 (0.8) Calculated RR: 0.33 (95% CI, 0.01 to 8.10) ARD: NR Nonvertebral Fracture: Denosumab: 1/121 (0.8) Placebo: 2/121 (1.7) Calculated RR: 0.50 (95% CI, 0.05 to 5.44) ARD: NR Hip Fracture: Denosumab: 0/121 Placebo: 0/121 Calculated RR: 1.00 (95% CI, 0.02 to 49.99) ARD: NR Other Fractures: NR	All-Cause Mortality Denosumab: 1/121 (0.8) Unrelated to study treatment Placebo: 1/121 (0.8) Unrelated to study treatment Calculated RR: 1.00 (95% CI, 0.06 to 15.81)	Discontinuation due to Adverse Events: Densoumab: 3/120 (2.5) Placebo: 1/120 (0.8) Calculated RR: 3.0 (95% CI, 0.32 to 28.4) ARD: NR Serious Adverse Events: Denosumab: 11/120 (9.2) Placebo: 10/120 (8.3) Calculated RR: 1.10 (95% CI, 0.49 to 2.49) ARD: NR Gastrointestinal Adverse Events: Denosumab: NR Placebo: NR RR: NR ARD: NR Other Adverse Events: Osteonecrosis of the jaw Denosumab: 0/120 (0) Placebo: 0/120 (0) RR: NR Atypical femur fracture Denosumab: 0/120 (0) Placebo: 0/120 (0) RR: NR All adverse events Denosumab: 86/120 (71.7) Placebo: 84/120 (70.0) RR: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Pols et al, 1999[259] Fosamax International Trial (FOSIT) RCT Some concerns/Fair Alendronate 10 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: Alendronate: 19/950 Placebo: 37/958 RR 0.52 (95% CI, 0.30 to 0.89) ARD: NR Hip Fracture: Alendronate: 2/950 Placebo: 3/958 RR 0.67 (95% CI, 0.11 to 4.01) ARD: NR Other Fractures: Wrist fractures: Alendronate: 6/950 Placebo: 15/958 RR: 0.47 (95% CI, 0.19 to 1.15)	All-Cause Mortality Alendronate: NR Placebo: NR RR: NR	Discontinuation due to Adverse Events: Alendronate: 61/950 Placebo: 54/958 RR, 1.14 (95% CI, 0.80 to 1.62) ARD: NR Serious Adverse Events: Serious adverse events, specifically those resulting in hospitalization or permanent disability or cancers Alendronate: 61/950 Placebo: 60/958 RR: NR ARD: NR Gastrointestinal Adverse Events: Any upper gastrointestinal adverse event Alendronate: 185/950 Placebo: 202/958 RR:NR ARD: NR Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Ravn et al, 1996[264] RCT Some concerns/Fair Ibandronate 0.25 mg/day; 0.50 mg/day; 1.0 mg/day; 2.5 mg/day; 5.0 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality 0.25 mg/d ibandronate: 0/26 RR not calculable 0.50 mg/d ibandronate: 0/22 RR not calculable 1.0 mg/d ibandronate: 0/26 RR not calculable 2.5 mg/d ibandronate:1/24 Calculated RR, 0.32 (95% CI, 0.01 to 7.53)	Discontinuation due to Adverse Events: 0.25 mg/d ibandronate: 1/30 0.50 mg/d ibandronate: 4/30 1.0 mg/d ibandronate: 2/30 2.5 mg/d ibandronate: 0/30 5.0 mg/d ibandronate: 6/30 Placebo: 2/30 0.25 mg/d ibandronate: Calculated RR, 0.50 (95% CI, 0.05 to 5.22) 0.50 mg/d ibandronate: Calculated RR, 2.00 (95% CI, 0.40 to 10.11) 1.0 mg/d ibandronate: Calculated RR, 1.00 (95% CI, 0.15 to 6.64)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Ravn et al, 1996[264] (continued)		5.0 mg/d ibandronate: 0/18 RR not calculable Placebo: 1/25	2.5 mg/d ibandronate: Calculated RR, 0.20 (95% CI, 0.01 to 4.0) 5.0 mg/d ibandronate: Calculated RR, 3.00 (95% CI, 0.66 to 13.69) Serious Adverse Events: 0.25 mg/d ibandronate: 1/30 0.50 mg/d ibandronate: 1/30 1.0 mg/d ibandronate: 0/30 2.5 mg/d ibandronate: 2/30 5.0 mg/d ibandronate: 1/30 Placebo: 3/30 0.25 mg/d ibandronate: Calculated RR, 0.33 (95% CI, 0.04 to 3.03) 0.50 mg/d ibandronate: Calculated RR, 0.33 (95% CI, 0.04 to 3.03) 1.0 mg/d ibandronate: Calculated RR, 0.14 (95% CI, 0.01 to 2.65) 2.5 mg/d ibandronate: Calculated RR, 0.67 (95% CI, 0.12 to 3.71) 5.0 mg/d ibandronate: Calculated RR, 0.33 (95% CI, 0.04 to 3.03) Gastrointestinal Adverse Events: Any GI AE 0.25 mg/d ibandronate: 12/30 0.50 mg/d ibandronate: 17/30 1.0 mg/d ibandronate: 8/30 2.5 mg/d ibandronate: 5/30 5.0 mg/d ibandronate: 17/30 Placebo: 11/30

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Ravn et al, 1996[264] (continued)			0.25 mg/d ibandronate: Calculated RR, 1.09 (95% CI, 0.57 to 2.07) 0.50 mg/d ibandronate: Calculated RR, 1.55 (95% CI, 0.88 to 2.72) 1.0 mg/d ibandronate: Calculated RR, 0.73 (95% CI, 0.34 to 1.55) 2.5 mg/d ibandronate: Calculated RR, 0.45 (95% CI, 0.18 to 1.15) 5.0 mg/d ibandronate: Calculated RR, 1.55 (95% CI, 0.88 to 2.72) Diarrhea 0.25 mg/d ibandronate: 6/30 0.50 mg/d ibandronate: 5/30 1.0 mg/d ibandronate: 2/30 2.5 mg/d ibandronate: Placebo: 2/30 0.25 mg/d ibandronate: Calculated RR, 3.00 (95% CI, 0.66 to 13.69) 0.50 mg/d ibandronate: Calculated RR, 2.50 (95% CI, 0.53 to 11.89) 1.0 mg/d ibandronate: Calculated RR, 1.00 (95% CI, 0.15 to 6.64) 2.5 mg/d ibandronate: Calculated RR, 1.00 (95% CI, 0.15 to 6.64) 5.0 mg/d ibandronate: Calculated RR, 4.50 (95% CI, 1.06 to 19.11) RR: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Ravn et al, 1996[264] (continued)			Other Adverse Events: Infection 0.25 mg/d ibandronate: 1/26 Calculated RR, 2.89 (95% CI, 0.12 to 67.76) 0.50 mg/d ibandronate: 0/22 Calculated RR, 1.13 (95% CI, 0.02 to 54.72) 1.0 mg/d ibandronate: 0/26 Calculated RR, 0.96 (95% CI, 0.02 to 46.76) 2.5 mg/d ibandronate: 0/24 Calculated RR, 1.04 (95% CI, 0.02 to 50.43) 5.0 mg/d ibandronate: 0/18 Calculated RR, 1.37 (95% CI, 0.03 to 65.94) Placebo: 0/25

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reginster et al, 2005[266] Monthly Oral Pilot Study (MOPS) RCT Some concerns/Fair Ibandronate 50 mg/per month; ibandronate 50 mg for the first month/100 mg/for months 2–3; ibandronate 100 mg/per month; ibandronate 150 mg/per month	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality Ibandronate 50 mg: 0/10 Ibandronate 50/100 mg: 0/18 Ibandronate 100 mg: 0/36 Ibandronate 150 mg: 0/36 Placebo: 0/36 RR: NR	Discontinuation due to Adverse Events: Any AE leading to withdrawal Ibandronate 50 mg: 0/18 Ibandronate 50/100 mg: 0/18 Ibandronate 100 mg: 0/36 Ibandronate 150 mg: 1/36 Placebo: 2/36 RR: Ibandronate 50 mg: Calculated RR, 0.39 (95% CI, 0.02 to 7.71) Ibandronate 50/100 mg: Calculated RR, 0.39 (95% CI, 0.02 to 7.71) Ibandronate 100 mg: Calculated RR, 0.20 (95% CI, 0.01 to 4.03) Ibandronate 150 mg: Calculated RR, 0.50 (95% CI, 0.05 to 5.27) Any drug-related AE leading to withdrawal Ibandronate 50 mg: 0/18 Ibandronate 50/100 mg: 0/18 Ibandronate 100 mg: 0/36 Ibandronate 150 mg: 0/36 Placebo: 0/36 Serious Adverse Events: Ibandronate 50 mg: 0/18 Ibandronate 50/100 mg: 0/18 Ibandronate 100 mg: 0/36 Ibandronate 150 mg: 0/36 Placebo: 0/36 RR not calculable Gastrointestinal Adverse Events: Upper GI AEs within 3 days of treatment Ibandronate 50 mg: 0/18 Ibandronate 50/100 mg: 4/18 Ibandronate 100 mg: 8/36 Ibandronate 150 mg: 9/36 Placebo: 6/36 Ibandronate 50 mg: Calculated RR, 0.15 (95% CI, 0.01 to 2.52)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reginster et al, 2005[266] (continued)			Ibandronate 50/100 mg: Calculated RR, 1.33 (95% CI, 0.43 to 4.13) Ibandronate 100 mg: Calculated RR, 1.33 (95% CI, 0.51 to 3.46) Ibandronate 150 mg: Calculated RR, 1.50 (95% CI, 0.60 to 3.78) Upper GI AEs anytime during treatment: Ibandronate 50 mg: 3/18 Ibandronate 50/100 mg: 11/18 Ibandronate 100 mg: Placebo: 12/36 Ibandronate 50 mg: Calculated RR, 0.50 (95% CI, 0.16 to 1.55) Ibandronate 50/100 mg: Calculated RR, 1.83 (95% CI, 1.02 to 3.31) Ibandronate 100 mg: Calculated RR, 1.25 (95% CI, 0.68 to 2.28) Ibandronate 150 mg: Calculated RR, 1.25 (95% CI, 0.68 to 2.28)
Reid et al, 2002[260] RCT Some concerns/Fair Zoledronic acid IV 0.25 mg/3 mo 0.5 mg/3 mo 1 mg/3 mo 4 mg/1 y 2 mg/6 mo	Vertebral Fracture: (radiographic) Zoledronic acid: 0 0.25 mg/3 mo: 0/60 Zoledronic acid 0.5 mg/3 mo: 0/58 Zoledronic acid 1 mg/3 mo: 0/53 Zoledronic acid 4 mg/1 y: 0/61 Zoledronic acid 2 mg/6 mo: 0/60 Placebo: 0/56 RR: NR ARD: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Zoledronic acid: 13/292 Placebo: 1/59 Calculated RR 2.62, (95% CI, 0.35 to 19.70) ARD: NR Serious Adverse Events: Zoledronic acid: 26/292 Placebo: 3/59 Calculated RR 21.75 (95% CI, 0.55 to 5.60) ARD: NR Gastrointestinal Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2002[260] (continued)	Nonvertebral Fracture: Zoledronic acid: 0 0.25 mg/3 mo: 0/60 Zoledronic acid 0.5 mg/3 mo: 1/58 Zoledronic acid 1 mg/3 mo: 2/53 Zoledronic acid 4 mg/1 y: 1/61 Zoledronic acid 2 mg/6 mo: 1/60 Placebo: 1/59 Calculated RR for zoledronic acid of 4 mg delivered in 1 to 4 doses (4/174): 1.36 (0.15 to 11.89) ARD: NR Hip Fracture: NR Other Fractures: NR		Other Adverse Events: Any adverse event Zoledronic acid 1 to 4 mg over 1 y in 1 to 4 infusions: 262/292 Placebo: 45/59 Calculated RR: 1.18 (95% CI, 1.02 to 1.36) Myalgia Zoledronic acid 1 to 4 mg over 1 y in 1 to 4 infusions: 41/292 Placebo: 1/59 Calculated RR: 8.28 (95% CI, 1.16 to 59.04) Arthralgia Zoledronic acid 1 to 4 mg over 1 y in 1 to 4 infusions: 46/292 Placebo: 9/59 Calculated RR: 1.03 (95% CI, 0.54 to 1.99) Influenza-like illness Zoledronic acid 1 to 4 mg over 1 y in 1 to 4 infusions: 26/292 Placebo: 4/59 Nausea Zoledronic acid 1 to 4 mg over 1 y in 1 to 4 infusions: 26/292 Placebo: 3/59

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] RCT Low/Good Zoledronic acid 5-mg IV every 18 months	Vertebral Fracture: Total (clinical and radiographic) Zoledronic acid: 23/1,000 Placebo: 49/1,000 HR: 0.45 (95% CI, 0.27 to 0.73) Symptomatic (clinical) Zoledronic acid: 14/1,000 Placebo: 34/1,000 HR 0.41 (95% CI, 0.22 to 0.75) Nonvertebral Fracture: Excluded fractures of toes, metatarsal bones, fingers, metacarpal bones, skull, facial bones, and mandible. Zoledronic acid: 101/1,000 Placebo: 148/1,000 HR: 0.66 (95% CI, 0.51 to 0.85) Hip Fracture: Zoledronic acid: 8/1,000 Placebo: 12/1,000 HR: 0.66 (95% CI, 0.27 to 1.16) ARD: NR Other Fractures: Forearm/Wrist Zoledronic acid: 36/1,000 Placebo: 63/1,000 HR: 0.56 (95% CI, 0.37 to 0.85) ARD: NR All fragility fractures including nonvertebral fragility fractures (excluding fractures of the toes, metatarsals, fingers, metacarpals, skull, facial bones, and mandible) and morphometric vertebral fractures	All-Cause Mortality Zoledronic acid: 27/1,000 Placebo: 41/1,000 OR 0.65 (95% CI, 0.40 to 1.05) RR 0.66 (95% CI, 0.41 to 1.06)	Discontinuation due to Adverse Events: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: Zoledronic acid: 47/1,000 Placebo: 64/1,000 RR: 0.73 (95% CI, 0.51 to 1.06) ARD: NR Other Adverse Events: Myocardial infarction Zoledronic acid: 24/1,000 Placebo: 39/1,000 OR: 0.61 (95% CI, 0.36 to 1.02) Stroke Zoledronic acid: 17/1,000 Placebo: 20/1,000 OR: 0.85 (95% CI, 0.44 to 1.63) Composite of vascular events (sudden death, myocardial infarction, coronary artery revascularization, or stroke) Zoledronic acid: 53/1,000 Placebo: 69/1,000 OR: 0.76 (95% CI, 0.52 to 1.09) Transient ischemic attack Zoledronic acid: 23/1,000 Placebo: 14/1,000 OR: 1.66 (95% CI, 0.85 to 3.24) Osteonecrosis of the jaw Zoledronic acid: 0/1,000 Placebo: 0/1,000 OR: Not calcuable

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] (continued)	Zoledronic acid: 122/1,000 Placebo: 190/1,000 HR: 0.63 (95% CI, 0.50 to 0.79) Symptomatic fractures includes symptomatic vertebral and nonvertebral fractures Zoledronic acid: 163/1,000 Placebo: 214/1,000 HR: 0.73 (95% CI, 0.60 to 0.90) Subgroup Analyses: Fragility fractures ages 73 to 91 Zoledronic acid: 55/330 Placebo: 75/336 OR: 0.70 (95% CI, 0.47 to 1.03) Fragility fractures ages 68 to 73 Zoledronic acid: 32/339 Placebo: 54/329 OR: 0.53 (95% CI, 0.33 to 0.85) Fragility fractures ages 65 to 68 Zoledronic acid: 35/321 Placebo: 61/335 OR: 0.53 (95% CI, 0.34 to 0.83) Fragility fractures total hip BMD T-score ≥-1.5 Zoledronic acid: 68/652 Placebo: 115/670 OR: 0.56 (95% CI, 0.41 to 0.78) Fragility fractures total hip BMD T-score ≥-2 to -1.5 Zoledronic acid: 32/224 Placebo: 53/228 OR: 0.55 (95% CI, 0.34 to 0.89)		Atrial Fibrillation Zoledronic acid: 54/1,000 Placebo: 55/1,000 OR: 0.98 (95% CI, 0.67 to 1.44) GI cancer deaths Zoledronic acid: 20/1,000 Placebo: 28/1,000 RR: 0.71 (95% CI, 0.41 to 1.26) Cardiac deaths Zoledronic acid: 4/1,000 Placebo: 3/1,000 RR:1.33 (95% CI, 0.30 to 5.94) Sudden death Zoledronic acid: 3/1,000 Placebo: 1/1,000 OR: 3.01 (95% CI, 0.3 to 28.9)

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] (continued)	Fragility fractures Total hip BMD T-score <-2 Zoledronic acid: 22/124 Placebo: 22/101 OR: 0.78 (95% CI, 0.40 to 1.50) Fragility fractures femoral neck BMD T-score ≥-1.5 Zoledronic acid: 43/378 Placebo: 73/404 OR: 0.58 (95% CI, 0.39 to 0.87) Fragility fractures femoral neck BMD T-score ≥-2 to -1.5 Zoledronic acid: 54/398 Placebo: 69/368 OR: 0.68 (95% CI, 0.46 to 1.00) Fragility fractures femoral neck BMD T-score <-2 Zoledronic acid: 25/224 Placebo: 48/227 OR: 0.47 (95% CI, 0.28 to 0.79) Fragility fractures lumbar spine BMD T-score ≥-1.5 Zoledronic acid: 54/602 Placebo: 95/600 OR: 00.52 (95% CI, 0.37 to 0.75) Fragility fractures lumbar spine BMD T-score ≥-2 to -1.5 Zoledronic acid: 21/131 Placebo: 33/151 OR: 0.68 (95% CI, 0.37 to 1.25) Fragility fractures lumbar spine BMD T-score <-2 Zoledronic acid: 23/151 Placebo: 33/137 OR: 0.57 (95% CI, 0.31 to 1.02)		

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] (continued)	Fragility fractures FRAX 10-y hip fracture risk 1st tertile (<1.8) Zoledronic acid: NR Placebo: NR OR: 0.62 (95% CI, 0.39 to 0.97) Fragility fractures FRAX 10-y hip fracture risk 2nd tertile (1.8 to 3.2) Zoledronic acid: NR Placebo: NR OR: 0.50 (95% CI, 0.32 to 0.78) Fragility fractures FRAX 10-y hip fracture risk 3rd tertile (>3.2) Zoledronic acid: NR Placebo: NR OR: 0.65 (95% CI, 0.44 to 0.96) Fragility fractures FRAX 10-y MOF risk 1st tertile (<9.9) Zoledronic acid: NR Placebo: NR OR: 0.62 (95% CI, 0.39 to 0.99) Fragility fractures FRAX 10-y MOF risk 2nd tertile (9.9 to 15) Zoledronic acid: NR Placebo: NR OR: 0.50 (95% CI, 0.32 to 0.77) Fragility fractures FRAX 10-y MOF risk 3rd tertile (>15) Zoledronic acid: NR Placebo: NR OR: 0.63 (95% CI, 0.42 to 0.93) Fragility fractures Garvan 5-y hip fracture risk 1st tertile (<1.5) Zoledronic acid NR Placebo: NR OR: 0.60 (95% CI, 0.38 to 0.97)		

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271] (continued)	Fragility fractures Garvan 5-y hip fracture risk 2nd tertile (1.5 to 3) Zoledronic acid: NR Placebo: NR OR: 0.55 (95% CI, 0.35 to 0.86) Fragility fractures Garvan 5-y hip fracture risk 3rd tertile (>3) Zoledronic acid: NR Placebo: NR OR: 0.60 (95% CI, 0.41 to 0.88) Fragility fractures Garvan 5-y osteoporotic fracture risk 1st tertile (<7.7) Zoledronic acid: NR Placebo: NR OR: 0.50 (95% CI, 0.31 to 0.82) Fragility fractures Garvan 5-y osteoporotic fracture risk 1st tertile (7.7 to 12) Zoledronic acid: NR Placebo: NR OR: 0.62 (95% CI, 0.40 to 0.97) Fragility fractures Garvan 5-y osteoporotic fracture risk 1st tertile (>12) Zoledronic acid: NR Placebo: NR OR: 0.61 (95% CI, 0.42 to 0.59)		
Riis et al, 2001[265] RCT Some concerns/Fair Ibandronate 2.5 mg/d; Ibandronate 20 mg every other day for the first 24 days out of every 3 months, followed by a 9-week period without active drug (intermittent cyclical therapy)	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality Ibandronate 2.5 mg: 1/81 Ibandronate 20 mg: 0/78 Placebo: 1/81 Ibandronate 2.5 mg; calculated RR, 1.00 (95% CI, 0.06 to 15.72 Ibandronate 20 mg; calculated RR, 0.35 (95% CI, 0.01 to 8.37)	Discontinuation due to Adverse Events: NR Serious Adverse Events: NR Gastrointestinal Adverse Events: No differences between continuous treatment, intermittent treatment, and placebo. During the first 12 months, the ibandronate-treated groups showed a numerically higher incidence of diarrhea compared with the placebo groups. Incidence of diarrhea was lower during the second year. Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Shiraki et al, 2003[295] RCT Some concerns/Fair Risedronate 1 mg, 2.5 mg, 5 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: NR Serious Adverse Events: Risedronate 1 mg: 0/50 Risedronate 2.5 mg: 0/49 Risedronate 5 mg: 0/53 Placebo: 0/51 RR not calculable Gastrointestinal Adverse Events: Risedronate 1 mg: 4/50 Risedronate 2.5 mg: 10/49 Risedronate 5 mg: 13/53 Placebo: 7/51 Risedronate 5 mg vs. placebo: calculated RR 1.79 (95% CI, 0.78 to 4.11) Other Adverse Events: Cardiac disturbances Risedronate 1 mg: 0/50 Risedronate 2.5 mg: 0/49 Risedronate 5 mg: 02/53 Placebo: 0/51 RR not estimable Disturbances of skin and subcutaneous tissues Risedronate 1 mg: 0/50 Risedronate 2.5 mg: 0/49 Risedronate 5 mg: 0/53 Placebo: 2/51 RR not estimable Disturbances of musculoskeletal, bone, and connective tissues Risedronate 1 mg: 0/50 Risedronate 2.5 mg: 1/49 Risedronate 5 mg: 1/53 Placebo: 0/51 RR not estimable

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Tanko et al, 2003[292] RCT Some concerns/Fair Ibandronate, 5 mg, 10 mg or 20 mg weekly	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Ibandronate: 8/472 Placebo: NR Serious Adverse Events: Ibandronate: 12.5% experienced a serious AE, but none were assessed as related to study drug (6/472 withdrew as a result of serious AE) Placebo: NR Gastrointestinal Adverse Events: Ibandronate 5 mg: 9/155 Ibandroante 10 mg: 8/155 Ibandroante 20 mg: 5/158 Placebo: 5/156 Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Thiebaud et al, 1997[293] RCT Some concerns/Fair Ibandronate 0.25, 0.5 mg, 1.0, or 2.0 mg/3 months 1 g calcium/day	Vertebral Fracture: NR Nonvertebral Fracture: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Ibandronate: 7/126 Placebo: NR RR: NR Serious Adverse Events: Ibandronate: 3/126 Placebo: NR RR: NR Gastrointestinal Adverse Events: Ibandronate 0.25 mg: 6/24 Ibandronate 0.5 mg: 6/27 Ibandronate 1.0 mg: 7/26 Ibandronate 2.0 mg: 3/23 Placebo: 4/26 Ibandronate 0.2 5 mg: Calculated RR, 1.63 (95% CI, 0.52 to 5.07) Ibandronate 0.5 mg: Calculated RR, 1.44 (95% CI, 0.46 to 4.54] Ibandronate 1.0 mg: Calculated RR, 1.75 (95% CI, 0.58 to 5.27) Ibandronate 2.0 mg: Calculated RR, 0.85 (95% CI, 0.21 to 3.40) Other Adverse Events: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Tucci et al, 1996[268] RCT Some concerns/Fair Alendronate 5 mg/day; Alendronate 10 mg/day; Alendronate 20 mg/day for 2 years followed by 5 mg/day	Vertebral Fracture: NR Nonvertebral Fracture: Alendronate 5 mg/day: 9/98 (9.2%) Alendronate 10 mg/day: 7/94 (7.4%) Alendronate 20 mg/day for 2 years followed by 5 mg/day: 11/94 (12%) Placebo: 21/192 (11%) RR: NR ARD: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Alendronate 5 mg/day: 3/98 (3.1%) Alendronate 10 mg/day: 5/94 (5.3%) Alendronate 20 mg/day for 2 years followed by 5 mg/day: 7/94 (7.4%) Placebo: 13/192 (6.8%) RR: Alendronate 10 mg/day vs. placebo: NR Alendronate 10 mg/day vs. placebo: RR 0.79 (95% CI, 0.29 to 2.14) Alendronate 20 mg/day for 2 years followed by 5 mg/day: NR Serious Adverse Events: Alendronate 5 mg/day: 12/98 (12.2%) Alendronate 10 mg/day: 20/94 (21.3%) Alendronate 20 mg/day for 2 years followed by 5 mg/day: 14/94 (14.9%) Placebo: 35/192 (18.2%) RR: Alendronate 10 mg/day vs. placebo: NR Alendronate 10 mg/day vs. placebo: RR 1.17 (95% CI, 0.71 to 1.91) Alendronate 20 mg/day for 2 years followed by 5 mg/day: NR Gastrointestinal Adverse Events: Any upper GI AE Alendronate 5 mg/day: 35/98 (35.7%) Alendronate 10 mg/day: 49/94 (52.1%) Alendronate 20 mg/day for 2 years followed by 5 mg/day: 39/94 (41.5%) Placebo: 79/192 (41.4%) RR: Alendronate 10 mg/day vs. placebo: NR Alendronate 10 mg/day vs. placebo: RR 1.27 (95% CI, 0.98 to 1.64) Alendronate 20 mg/day for 2 years followed by 5 mg/day: NR

Appendix D Table 16. Outcomes from Included Studies for Benefits and Harms of Treatment (Key Questions 4 and 5)

Author, Year Trial Name Study Design ROB/Study Quality Drug Dose	Fracture Outcomes	Mortality	Harm Outcomes
Tucci et al, 1996[268] (continued			Other Adverse Events: Any AE Alendronate 5 mg/day: 92/98 (93.9%) Alendronate 10 mg/day: 89/94 (94.7%) Alendronate 20 mg/day for 2 years followed by 5 mg/day: 88/94 (93.6%) Placebo: 181/192 (94.3%) Alendronate 10 mg/day vs. placebo: RR NR Alendronate 10 mg/day vs. placebo: RR 1.00 (95% CI, 0.95 to 1.07) Alendronate 20 mg/day for 2 years followed by 5 mg/day: RR NR
Valimaki et al, 2007[261] RCT Some concerns/Fair Risedronate 5 mg/d	Vertebral Fracture: (clinical) Risedronate: 0/114 Placebo: 0/56 RR: NR ARD: NR Nonvertebral Fracture: Risedronate: 2/114 Placebo: 2/56 Calculated RR, 0.49 (95% CI, 0.07 to 3.40) ARD: NR Hip Fracture: NR Other Fractures: NR	All-Cause Mortality NR	Discontinuation due to Adverse Events: Risedronate: 10/115 Placebo: 9/55 Calculated RR, 0.53 (95% CI, 0.23 to 1.23) Serious Adverse Events: Risedronate: 12/114 Placebo: 3/56 Calculated RR, 1.96 (95% CI, 0.58 to 6.68) Gastrointestinal Adverse Events: Risedronate: 21/115 Placebo: 14/55 Calculated RR, 0.72 (95% CI, 0.40 to 1.30) Other Adverse Events: NR

Abbreviations: AE=adverse event; ARD=absolute risk difference; BMD=bone mineral density; CI=confidence interval; FRAX=Fracture Risk Assessment Tool; GERD=gastroesophageal reflux disease; GI=gastrointestinal; HR=hazard ratio; IV=intravenous; MOF=major osteoporotic fracture; NR=not reported; OR=odds ratio; RCT=randomized, controlled trial; RR=risk ratio; RRR=relative risk reduction; vs.=versus.

Appendix D Table 17. Outcomes from Included Cohort Studies for Harms of Treatment (Key Question 5)

First Author Year Cohort Title ROB/Study Quality	Gastrointestinal Adverse Events	Atypical Femur Fracture	Osteonecrosis of the Jaw	Other Adverse Events
Lee, 2019[307] Korean National Health Insurance Data Some concerns/Fair	NR	Exposed: Overall incidence: 682/348,311 Overall IR: 37.75/100,000 person-years (95% CI, 35.02 to 40.70) Females incidence: 633/316,472 Females IR: 38.20/100,000 person-years (95% CI, 35.34 to 41.30) Males incidence: 49/31,839 Males IR: 32.78/100,000 person-years (95% CI, 24.77 to 43.37) Comparator: Overall incidence: 475/348,548 Overall IR: 24.41/100,000 person-years (95% CI, 22.31 to 26.71), p<0.0001 vs. users Females incidence: 425/316,617 Females IR: 23.91/100,000 person-years (95% CI, 21.74 to 26.29) Males incidence: 50/31,877 Males IR: 29.79/100,000 person-years (95% CI, 22.58 to 39.30) RR: Adjusted HR 1.53 (95% CI, 1.36 to 1.73); adjusted for age, sex, use of systemic glucocorticoids, and comorbidity ARD: NR	NR	NR

Appendix D Table 17. Outcomes from Included Cohort Studies for Harms of Treatment (Key Question 5)

First Author Year Cohort Title ROB/Study Quality	Gastrointestinal Adverse Events	Atypical Femur Fracture	Osteonecrosis of the Jaw	Other Adverse Events
Pazianas, 2012[305] Danish National Prescription Database and Cause of Death registry Some concerns/Fair	Exposed: Any colon cancer diagnosis, mean followup time=3.4 years Alendronate incidence: 262/30,606 Death due to any colon cancer, mean followup time=4.9 years Alendronate incidence: 190/30,606 Comparator: Any colon cancer diagnosis Nonusers incidence: 1,421/122,424 Death due to any colon cancer Nonusers incidence: 1,083/122,424 RR: Any colon cancer diagnosis aHR: 0.69 (95% CI, 0.60 to 0.79) Death due to any colon cancer aHR: 0.62 (95% CI, 0.52 to 0.72) Any colon cancer diagnosis ≥12 months after alendronate start and use of >180 DDD aHR: 0.89 (95% CI, 0.66 to 1.22, p=0.48) HR adjusted for age, Charlson comorbidity index, known colon cancer risk factors (ulcerative colitis, Crohn's disease, celiac disease), hormone replacement therapy, and amount of prednisolone, nonsteroidal anti-inflammatory drugs and acetyl-salicylic acid used in the last 12 months ARD: NR	NR	NR	NR

Appendix D Table 17. Outcomes from Included Cohort Studies for Harms of Treatment (Key Question 5)

First Author Year Cohort Title ROB/Study Quality	Gastrointestinal Adverse Events	Atypical Femur Fracture	Osteonecrosis of the Jaw	Other Adverse Events
Rubin, 2020[306] Swedish and Danish National Health Registries Some concerns/Fair	NR	Exposed: NR Comparator: NR **Adjusted HR (95% CI):** 2.46 (1.17 to 5.15, proportional hazards assumption noted to be problematic); adjusted for age, previous fracture, comorbidities, and previous medication ARD: NR	Exposed: NR Comparator: NR Not enough data to determine adjusted HR ARD: NR	Atrial fibrillation Incidence: NR **Adjusted HR (95% CI):** 1.18 (0.99 to 1.40) adjusted for age, previous fracture, comorbidities, and previous medication ARD: NR Myocardial infarction Incidence: NR **Adjusted HR (95% CI):** 0.92 (0.64 to 1.31) adjusted for age, previous fracture, comorbidities, and previous medication ARD: NR Heart failure Incidence: NR **Adjusted HR (95% CI):** 1.32 (1.08 to 1.61) adjusted for age, previous fracture, comorbidities, and previous medication Cardiovascular mortality Incidence: NR **Adjusted HR (95% CI):** 0.97 (0.81 to 1.15) adjusted for age, previous fracture, comorbidities, and previous medication ARD: NR

Abbreviations: aHR=adjusted hazard ratio; ARD=absolute risk difference; CI=confidence interval; HR=hazard ratio; IR=incidence ratio; NR=not reported; ROB=risk of bias; RR=risk ratio.

Appendix D Table 18. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 1 Randomization and Allocation Concealment)

Author, Year Trial Name	Was method of randomization adequate?	Was allocation concealment adequate?	Were there baseline imbalances between groups that suggest a problem with randomization?	ROB: Randomization or Selection	Comments on Bias Arising From Randomization or Selection
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS	Yes	Yes	No	Low	None
Rubin, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE	Yes	No information	No	Low	No information about allocation concealment, but method used to invite this large number of participants (mailed letters) makes it unlikely.
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP	Yes	Yes	No	Low	None

Abbreviations: ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study.

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS	Screening: 59/5,575 (1.1%) at 18 months Usual care: 53/5,457 (1.0%) at 18 months No variation by outcome. Approximately 6% of data missing at 36 months; no breakdown by group. Author query sent.	No	Yes	No information	Low	None

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Rubin, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE	0% had missing data for the outcome (ITT analysis).	No	Yes	Other	Low	Conducted an ITT analysis for the fracture outcomes. This was a pragmatic trial, thus not entirely surprising that nearly 40% of participants did not return completed questionnaires and thus did not participate. Authors did identify differences between those who returned questionnaires and those who did not. They also conducted per-protocol analyses given a large proportion did not actually participate in the screening intervention.
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP	Pragmatic trial. 12 participants excluded post-randomization (0.09%), 6 in each group (0.045% in each group). No variation by different outcome.	No	Yes	Other	Low	None

Abbreviations: ITT=intention to treat; ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study.

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	Were the patients unaware of the assigned intervention status?	Were the trial personnel/ clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS	No	No	No	Yes	Some or unclear	Participants and clinicians not blinded; not cluster randomized so general practitioners who received education and training may have been more attuned to evaluation and treatment of osteoporosis in the usual care group; screening group: 1,347/5,575 randomized (24%) to screening did not receive screening 1,417/5,575 randomized (25%) to screening had an indication for treatment 1,154/5575 (21%) randomized to screening received treatment over the course of the study. 18% (982/5,575 randomized) reported starting treatment and 11.8% (657/5,575 randomized) reported still being on treatment at 36 months; of those without an indication, 1% (68/5575 randomized) reported treatment at 36-months. The discussion states that "31% of those with an indication did not start medication." 52/5,457 randomized (1%) to control were lost to followup and not included 291/5,457 randomized (5%) to control received treatment over the course of the study; 3% (167/5,457) by 18-months.
Rubin et al, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE	No	No	No	Yes	Some or unclear	Participants and clinicians not blinded. 7,793/17,072 randomized (45.6%) to screening did not receive screening with FRAX calculation (1,132 already on treatment, 2,894 returned questionnaire blank, 104 returned questionnaire with data missing to calculate FRAX, and the rest did not return the questionnaire).
Rubin et al, 2018[126] Rothman et al,						2,047/17,072 randomized (12%) were high-risk but did not have a DXA (830 weren't interested

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	Were the patients unaware of the assigned intervention status?	Were the trial personnel/ clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
2017[128] Hoiberg et al, 2019[129] ROSE (continued)						in a DXA and 1,217 dropped out); 5,009/17,072 randomized (29%) were high-risk and had a DXA. The authors report that 48% of those screened had a DXA which comes from the 10,411 with calculated FRAX scores and not the overall randomized intervention group of 17,072. 1,236/17,072 randomized (7%) had a DXA result with an indication for treatment. Eligibility for DXA required a completed questionnaire and high risk FRAX score (≥15%). 986/17,072 randomized (6%) received treatment; this number 986 appears to be based on only those who received DXA through the study and had an indication for treatment based on the study DXA who were then referred back to their GPs for further evaluation and management as part of the study. The authors state that 23% of the screening group received medication after the index date (mailing of questionnaire); which we assume includes the 1,132 women that indicated they were already receiving medication on the baseline questionnaire along with women who were randomized to screening but who did not return the questionnaire but who may have been prescribed medication by their GPs through the course of usual care outside of this study. 7831/17,157 randomized (45.6%) did not participate (1,168 were already on treatment, 3,143 returned a blank questionnaire, 111 returned a questionnaire with missing data to calculate FRAX, and the rest did not return the questionnaire)
Rubin et al, 2018[126] Rothman et al, 2017[128]						In the control group 7,026/17,157 randomized (41%) had FRAX ≥15% The number of participants in the control group

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	Were the patients unaware of the assigned intervention status?	Were the trial personnel/ clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Hoiberg et al, 2019[129] ROSE (continued)						that received a DXA was not reported but the authors report that 25% of women in the control group had a DXA vs. 48% in the screening group. Based on the information in the article, the denominator is likely "Calculated FRAX total" and this gives us a N/10,494= 25% such that likely N= 2,623.5 or 15% of total control group The authors note that 18% of the control group received medication after the index date (mailing of the questionnaire); it is unclear whether these were women with FRAX ≥15% and ≤15% or whether they received DXA prior to treatment, and whether this includes the 1,168 women who were excluded from FRAX calculation because they indicated they were taking treatment on the baseline questionnaire.
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP	No	No	No	Yes	Some or unclear	Participants and clinicians not blinded. Participants in control group may have been offered screening and/or treatment through usual care. This was a pragmatic trial carried out in general practice settings and blinding was not feasible due to nature of the intervention. 6/6,233 randomized (<0.1%) to screening were not screened 247/6,233 (4%) randomized to screening were high risk but did not have a DXA (157 declined, 81 were unable to have hip BMD measured, and 9 died) 2,817/6,233 randomized (45%) to screening were high-risk after FRAX screening and had a DXA.
Shepstone et al, 2018[120] Shepstone et al, 2012[121]						898/6,233 randomized (14.4%) to screening continued to be high risk after revised FRAX score with BMD and had treatment recommended.

Appendix D Table 20. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 3 Departures from Intended Interventions)

Author, Year Trial Name	Were the patients unaware of the assigned intervention status?	Were the trial personnel/ clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP (continued)						1,486/6,233 randomized (24%) received at least one prescription for treatment over the course of the study; 953/6,233 randomized (15%) received treatment in the first 12 months; of those considered high-risk, 703/898 (78%) received treatment in the first 6-months. Adherence among those taking medication at 6 months: 79.2% by 1 y, 65% by 2 y, 34.9% by 5 y. 6/6,250 randomized (<0.1%) to control did not participate Number randomized to control that received DXA through usual care was NR 982/6,250 randomized (16%) to control received treatment over the course of the study; 2,64/6,250 randomized (4%) in the first 12 months. Participants with prescriptions for anti-osteoporotic medication: End of first year: screened, 15%, not screened, 4% End of fifth year: overall, 11.5%; screened, 13%-14%, not screened, 9.7%

Abbreviations: ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study.

Appendix D Table 21. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 4 Outcome Measurement)

Author, Year Trial Name	Were benefit outcomes adequately described, valid, and reliable and was the duration of followup adequate?	Were harm outcomes adequately described, valid, and reliable with an adequate duration of followup?	Were outcome assessors masked to group assignment?	ROB: Outcome Measurement	Comments on Bias Arising From Measurement of Outcomes
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS	Yes	Yes	Yes	Low	Followup was through 36 months; this is a length of followup associated with treatment benefit in drugs trials of anti-osteoporosis medications.
Rubin et al, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE	Yes	NA-no harm outcomes	No information	Low	Administrative/registry data used to identify outcomes, formal masking of persons pulling and analyzing these data was NR.
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP	Yes	Yes	Yes	Low	Fracture outcomes were verified with medical records.

Abbreviations: NA=not applicable; NR=not reported; ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study.

Appendix D Table 22. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 5 Selective Outcome Reporting and Overall ROB)

Author, Year Trial Name	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study ROB	ROB Rating Justification	Does ROB rating vary by outcome?
Merlijn et al, 2019[124] Elders et al, 2017[125] SALT-SOS	Yes	Low	None	Fair	This was a pragmatic RCT but has moderate risk of bias because practitioners and patients were not blinded, only modest fidelity for the screening interventions, and some contamination of the usual-care group. Outcome assessment was blinded.	No
Rubin et al, 2018[126] Rothman et al, 2017[128] Hoiberg et al, 2019[129] ROSE	Yes	Low	None	Fair	Moderate risk of bias deviations from intended intervention; trial was not blinded, and there was contamination in the control group and poor fidelity to the intervention in the screening group; however, this was a large pragmatic trial so not entirely unexpected.	No

Appendix D Table 22. Risk of Bias for Included Trials in Key Questions 1 and 3 (Domain 5 Selective Outcome Reporting and Overall ROB)

Author, Year Trial Name	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study ROB	ROB Rating Justification	Does ROB rating vary by outcome?
Shepstone et al, 2018[120] Shepstone et al, 2012[121] McCloskey et al, 2018[122] Parsons et al, 2020[123] SCOOP	Yes	Low	None	Fair	Some risk of bias because of deviations from intended interventions and poor fidelity of intervention.	No

Abbreviations: RCT=randomized, controlled trial; ROB=risk of bias; ROSE=Risk-stratified Osteoporosis Strategy Evaluation Study; SALT-SOS=Stichting Artsen Laboratorium en Trombosedienst Osteoporosis Study; SCOOP=Screening in the Community to Reduce Fractures in Older Women study.

Appendix D Table 23. Risk of Bias for Systematic Reviews Included for Key Question 1 (Domain 1 Study Eligibility)

Author, Year	1.1 Did the review adhere to predefined objectives and eligibility criteria?	1.2 Were the eligibility criteria appropriate for the review question?	1.3 Were eligibility criteria unambiguous?	1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate (e.g., date, sample size, study quality, outcomes measured)?	1.5 Were any restrictions in eligibility criteria based on sources of information appropriate (e.g., publication status or format, language, availability of data)?	Concerns Regarding Specification of Study Eligibility Criteria	Rationale for Concern
Auais et al, 2023[133]	Probably yes	Probably yes	Probably yes	Probably yes	Probably yes	Low	None
Gates et al, 2023[131]	Yes	Yes	Yes	Yes	Yes	Low	None
Merlijn et al, 2020[130]	Yes	Yes	Probably yes	Yes	Yes	Low	None

Appendix D Table 24. Risk of Bias for Systematic Reviews Included for Key Question 1 (Domain 2 Identification and Selection of Studies)

Author, Year	2.1 Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?	2.2 Were methods additional to database searching used to identify relevant reports?	2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	2.4 Were restrictions based on date, publication format, or language appropriate?	2.5 Were efforts made to minimize error in selection of studies?	Concerns Regarding Methods Used to Identify and/or Select Studies	Rationale for Concern
Auais et al, 2023[133]	Probably yes	Probably yes	Probably yes	Probably yes	Yes	Low	None
Gates et al, 2023[131]	Yes	Yes	Yes	Yes	Yes	Low	None
Merlijn et al, 2020[130]	Probably yes	Yes	Probably no	Yes	No Information	Unclear	Search terms not comprehensive; only two databases searched, but studies of screening are unlikely to be found outside of these two databases.

Appendix D Table 25. Risk of Bias for Systematic Reviews Included for Key Question 1 (Domain 3 Data Collection and Study Appraisal)

Author, Year	3.1 Were efforts made to minimize error in data collection?	3.2 Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	3.3 Were all relevant study results collected for use in the synthesis?	3.4 Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	3.5 Were efforts made to minimize error in risk-of-bias assessment?	Concerns Regarding Methods Used to Collect Data and Appraise Studies	Rationale for Concern
Auais et al, 2023[133]	Yes	Yes	Probably yes	Probably no	Yes	Unclear	Used PEDro scale to assess ROB for RCTs
Gates et al, 2023[131]	Yes	Yes	Yes	Yes	Yes	Low	None
Merlijn et al, 2020[130]	Yes	Yes	Yes	Yes	No information	Low	None

Appendix D Table 26. Risk of Bias for Systematic Reviews Included for Key Question 1 (Domain 4 Synthesis and Findings)

Author, Year	4.1 Did the synthesis include all studies that it should?	4.2 Were all predefined analyses reported or departures explained?	4.3 Was the synthesis appropriate given the nature and similarity in the research questions, study designs, and outcomes across included studies?	4.4 Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	4.5 Were the findings robust (e.g., as demonstrated through funnel plot or sensitivity analyses)?	4.6 Were biases in primary studies minimal or addressed in the synthesis?	Concerns Regarding the Synthesis and Findings	Rationale for Concern
Auais et al, 2023[133]	Probably yes	Probably yes	Probably yes	Probably yes	No information	Probably yes	Low	None
Gates et al, 2023[131]	Yes	Yes	Yes	Probably Yes	Yes	Low	None	None
Merlijn et al, 2020[130]	Yes	Yes	Yes	Yes	Yes	Probably yes	Low	None

Appendix D Table 27. Risk of Bias for Systematic Reviews Included for Key Question 1 (Overall Risk of Bias)

Author, Year	A. Did the interpretation of findings address all of the concerns identified in Domains 1 through 4?	B. Was the relevance of identified studies to the review's research question appropriately considered?	C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?	ROB in the Review/Study Quality	Rationale for ROB
Auais et al, 2023[133]	Yes	Yes	Yes	Low/Good	None
Gates et al, 2023[131]	Yes	Yes	Yes	Low/Good	None
Merlijn et al, 2020[130]	Yes	Yes	Yes	Low/Good	None

Abbreviations: ROB=risk of bias.

Appendix D Table 28. Risk of Bias for Included Studies for Key Question 2a (Domain 1 Participants)

Author, Year	Risk Assessment Instrument	Population and Data Sources	Representation of Racial/Ethnic Minorities?	Domain 1 Comments
Azagra et al, 2015[168, 310]	FRAX	FRIDEX cohort; women ages 40 to 90 years referred for DXA by their physician; Persons with cancer or who were receiving osteoporosis medications were excluded.	No information	Cohort of Spanish women
Bolland et al, 2011[166]	FRAX, Garvan Fracture Risk Calculator	Healthy menopausal women age ≥55 years who were taking part in a 5-year placebo-controlled trial of calcium supplements; normal lumbar spine BMD for their age (Z-score >-2), not taking osteoporosis medication or vitamin D supplements in doses >1,000 IU/day, serum 25 [OH] D levels ≥25 nmol/L.	No information	Conducted in New Zealand
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159]	FRAX, OST	Manitoba BMD Registry, a population-based registry of all persons who received DXA testing in the province of Manitoba, Canada. See **Appendix D Table 2** for description of date and population criteria used in each of the analyses reported in the various articles cited.	One article notes that 98% of the cohort was White.	Conducted in Canada
Chapurlat et al, 2020[175]	FRAX	OFELY and QUALYOR; Retrospective analysis of 2 population-based cohorts (OFELY and QUALYOR). Postmenopausal women with a baseline bone measure obtained during 2006-2008 from OFELY, and women with T-scores at the hip of between -1.0 and -2.5 with clinical risk factors or <-3.0 without risk factors from QUALYOR.	No information	Study conducted among 2 French cohort studies
Cheung et al, 2012[150]	FRAX	Hong Kong Osteoporosis Study; Community-dwelling, ambulatory, postmenopausal women age ≥40 years recruited from different districts of Hong Kong between 1995 and 2009 during health fairs and road shows on osteoporosis; Women taking osteoporosis treatment were excluded.	No/Probably No	All participants were Southern Chinese post-menopausal women.
Collins et al, 2011[167]	QFracture	THIN Database; patients ages 30 to 85 years registered between 1994 and 2008 with records in the THIN database, a database of general practices that use INPS Vision system (20% of U.K. practices); no previously recorded fracture of hip, distal radius, or vertebra	No information	Study conducted in U.K.

Appendix D Table 28. Risk of Bias for Included Studies for Key Question 2a (Domain 1 Participants)

Author, Year	Risk Assessment Instrument	Population and Data Sources	Representation of Racial/Ethnic Minorities?	Domain 1 Comments
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143]	FRAX, Garvan, OST, SCORE	Retrospective analysis of participants assembled from the Women's Health Initiative Clinical Trials and Observational study; 87% White; 10.4% Black or Hispanic; postmenopausal, free from serious medical conditions; participants using osteoporosis medication or somatostatin agents at baseline were excluded as were participants with fewer than 10 years of followup time and who contributed incomplete information regarding risk factors.	No/Probably No	Conducted in U.S.
Dagan et al, 2017[169]	FRAX, QFracture	Electronic health record data for members ages 30 to 100 years (depending on tool validation) from one of four national healthcare insurer/providers; race/ethnicity NR; Continuous membership in the health plan for 3 years prior to index date and during followup period.	No/Probably No	98.8% of study population was White and 1.2% was Black
Davis et al, 2019[173]	QFracture	Fremantle Diabetes Study Phase 1; retrospective analysis of a longitudinal cohort of persons with known diabetes from an urban community in one region of the country; only cohort members between age 40 and 90 with type 2 diabetes were included in this analysis.	No/Probably No	Anglo-celt: 62%; southern European: 17% to 22%; Other European: 8% to 14%; Asian 0% to 3.5%; Indigenous Australian: 0% to 1.3%; Mixed other: 2.0% to 7.8%
Desbiens et al, 2020[177]	FRAX, QFracture	CARTaGENE; retrospective analysis of data from a population-based survey of adults ages 40 to 69 years in a single province; persons with history of dialysis or kidney transplant were excluded. Only the persons without chronic kidney disease from this cohort were included for this update review. Persons living in nursing home, correctional facilities, and First Nation Reserves were excluded.	No/Probably No	White participants made up roughly 89% of participants; other race/ethnicity groups were not reported.
Ensrud et al, 2009[152] Premaor et al, 2013[153]	FRAX	Study of Osteoporotic Fractures (SOF); Women age ≥65 years recruited between 1986 and 1988 from population-based listings in 4 U.S. areas; Black women were excluded because of low incidence of hip fracture; women who were unable to walk without assistance or had a history of bilateral hip replacement were also excluded.	No/Probably No	Only White women were included in the analysis. Black women were excluded due to "low incidence of hip fracture."
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146]	FRAX FRC QFracture	Retrospective analysis of the MrOs cohort of community-dwelling men age ≥65 years recruited from 6 clinical centers between March 2000 and April 2002; U.S. Cohort: 89.4% White; 4% Black; 3% Asian; 2% Hispanic, 1% Other; Men who used bisphosphonates in the 30 days prior to enrollment were excluded. Some analyses include participants from the MrOs cohort recruited from Hong Kong. Some analyses excluded participants with osteoporosis at baseline.	No/Probably No	89% White, 4% Black, 3% Asian, 2% Hispanic

Appendix D Table 28. Risk of Bias for Included Studies for Key Question 2a (Domain 1 Participants)

Author, Year	Risk Assessment Instrument	Population and Data Sources	Representation of Racial/Ethnic Minorities?	Domain 1 Comments
Fraser et al, 2011[154] Langsetmo et al, 2011[155]	FRAX Garvan	Canadian Multicentre Osteoporosis Study (CaMos); Data from the CaMos cohort which included persons living within proximity to 1 of 9 Canadian cities randomly selected from residential phone numbers	No information	Study conducted in Canada
Goldshtein et al, 2018[172] Goldshtein et al, 2018[172]	FRAX	Maccabi Healthcare Services (MHS); Retrospective cohort assembled from data from the computerized database of Maccabi Healthcare Services (MHS); a large government-funded health maintenance organizations. This analysis includes women ages 50 to 90 years in 2004 with at least 3 years of prior MHS membership; persons with osteoporosis treatment were included (19%) were on therapy before the index date	No information	Study conducted in Israel
Gonzalez-Macias et al, 2012[149]	FRAX	Caucasian women age 65 years or older recruited from 58 primary care centers of the National Health Services in Spain between March 2000 and June 2001, comprising ECOSAP cohort; Excluded women with metabolic bone disease, renal failure, hypercalcemia, therapeutic doses of fluoride for certain duration, life expectancy <3 years.	No/Probably No	All recruited participants were Caucasian women
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148]	QFracture	QResearch database of more than 13 million patients registered at more than 620 general practices in the U.K.	No information	Study conducted in the U.K.
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170]	QFracture FRAX	Clinical Research Data Link (CPRD); Retrospective analysis of data on participants ages 30 to 99 years from the Clinical Practice Research Database (CPRD), a database of patients from general practices in the U.K.; one analysis[171] limited to persons at 357 practices with links to the Office of National Statistics. The other analysis[170] involved persons ages 40 to 90 years between January 1987 and December 2013 from medical records of 625 primary care practices. Race/ethnicity NR; Persons exposed to osteoporosis drugs before the index date were excluded; the reported analysis compared persons with RA to the general population; only data for the general population is captured here.	No/Probably No	White, 95%; Indian, 1%; the remaining 4% included Pakistani, Bangladeshi, Other Asian, Carribean, Black African, Chinese, and “other ethnic group.”

Appendix D Table 28. Risk of Bias for Included Studies for Key Question 2a (Domain 1 Participants)

Author, Year	Risk Assessment Instrument	Population and Data Sources	Representation of Racial/Ethnic Minorities?	Domain 1 Comments
Jain et al, 2023[311]	FRAX, QFracture	Retrospective analysis of EHR data on men and women at least 50 years of age with at least 1 primary care provider visit per year between 2010 and 2018 and at least 2 full years of followup. Excluded participants with missing data for risk calculation or with prescription for osteoporosis medication.	Yes, but Asian participants (too few in number) and participants with unknown race or who were multiracial were excluded (no calculator specific to these groups)	White: 64.5 (9.9) Black: 61.2 (9.3) Hispanic: 60.2 (8.7)
Lo et al, 2011[73] Pressman et al, 2011[165]	FRAX, FRC	Kaiser Permanente Northern California; women ages 50 to 85 years who underwent first DXA scan between 1997 and 2003; 78% White; 12% Asian, 6% Hispanic, 4% Black; Excluded women without coverage 1 year prior to and after the DXA scan, without accessible data, or missing race/ethnicity. Women with a filled prescription for bisphosphonates in the year prior to DXA were also excluded.	Yes/Probably Yes	76% White, 4% Black, 6% Hispanic, 13.9% Asian
Lu et al, 2021[178]	FRAX	Retrospective analysis using data from 5 cohort studies (UK Biobank, MrOs US, MrOs Sweden, SOF, CKB). These were population-based cohorts with varying inclusion/exclusion criteria.	No information	Do not provide overall race/ethnicity percentages, note that one study from U.S. has 11% "visible minorities." Break down genetic results for European ancestry and Asian populations only
Marques et al, 2017[176]	FRAX	3 different Portuguese cohorts (SAOL, IPR, EPIPorto); Retrospective analysis using data from 3 Portuguese cohorts (SAOL, IPR, EPIPort) using participants age 40 years and older with complete FRAX data.	No information	Study conducted in Portugal.
Pluskiewicz et al, 2023[312]	FRAX, Garvan	Population-based random sample of postmenopausal women age 55 years or older in Poland supplemented with a nonrandom sample of additional volunteers but it is unclear how these additional volunteers were recruited; however no differences in baseline history of fracture of FRAX or Garvan fracture risk was observed between participants recruited randomly or nonrandomly.	No information	Study conducted in Poland
Tamaki et al, 2011[151]	FRAX	Population-based cohort of women ages 15 to 79 years randomly selected in 5-year age groups from resident registrations in municipalities in Japan starting in 1996.	None	Study conducted in Japan.

Appendix D Table 28. Risk of Bias for Included Studies for Key Question 2a (Domain 1 Participants)

Author, Year	Risk Assessment Instrument	Population and Data Sources	Representation of Racial/Ethnic Minorities?	Domain 1 Comments
Tanaka et al, 2010[156]	FRAX	Data from participants enrolled in two Japanese cohort studies (Miyama and Taiji); these cohorts randomly selected participants ages 40 to 79 years for recruitment from resident registration records in December 1988 and the Taiji cohort enrolled participants ages 40 to 79 years randomly selected from resident registration records in June 1992; only women from these cohorts were included in this analysis of the validation dataset.	None	Study conducted in Japan.
Tebe Cordomi et al, 2013[144]	FRAX	Random sample of women identified from a database of women ages 40 to 90 years with a first visit for DXA between January 1992 and February 2008.	None	Study conducted in Spain.
Zwart et al, 2023[313]	FRAX	Caucasian persons ages 40 to 90 years from different regions of Spain selected from lists of participating general practitioners. Persons who received bone-conserving drugs at baseline or during 10 years of followup were excluded.	None	Study conducted in Spain among Caucasian persons

Abbreviations: CaMos=Canadian Multicentre Osteoporosis Study; DXA=dual-energy x-ray absorptiometry; ECOSAP=Ecografia Osea en Atencio Primaria; FRIDEX=Fracture RIsk factors and bone DEnsitometry type central dual X-ray; MrOs=Osteoporotic Fractures in Men; NR=not reported; QUALYOR=QUalité Osseuse LYon Orléans; SOF=Study of Osteoporotic Fractures; THIN=The Health Improvement Network; U.K.=United Kingdom; U.S.=United States.

Appendix D Table 29. Risk of Bias for Included Studies for Key Question 2a (Domain 2 Predictors)

Author, Year	Risk Assessment Instrument	Deviations in Predictors Compared to Development Cohort	Domain 2 Comments
Azagra et al, 2015[168, 310]	FRAX	NR	None
Bolland et al, 2011[166]	FRAX, Garvan	None, predictors were collected by questionnaire	None
Brennan et al, 2014[158]	FRAX	Prior to 2000, height and weight were self-reported. Instead of interview data for smoking/alcohol intake, COPD was used as proxy for smoking status and diagnosis of alcohol/substance use used as proxy for alcohol	Use of ICD codes for smoking/alcohol makes it likely that subjects with more mild-moderate use were missed.
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159]	FRAX, OST	Prior to 2000, height and weight were self-reported. Instead of interview data for smoking/alcohol intake, COPD was used as proxy for smoking status and diagnosis of alcohol/substance use used as proxy for alcohol[158] Parental hip fracture, smoking, and alcohol were from ICD codes before 2005 then switched to self-report (data collected 1987-2016)[160] Proxies were used for smoking (COPD) and high alcohol intake (alcohol or substance abuse diagnosis). Parental hip fracture information was collected only in the last two years (2005 and onwards) and therefore was missing for earlier cases.[157]	Use of ICD codes for smoking/alcohol makes it likely that subjects with more mild-moderate use were missed.
Chapurlat et al, 2020[175]	FRAX	NR	None
Cheung et al, 2012[150]	FRAX	NR	None
Collins et al, 2011[167]	QFracture	NR	None
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143]	FRAX, Garvan, OST, SCORE	Used data collected at baseline enrollment into the WHI to determine risk factors; only 1 of the articles discusses availability of data[139] and reported that paternal hip fracture was missing for 7,519 participants and maternal hip fracture history was missing for 8,180 participants; missing information was less common for other factors (BMI, missing n=340; smoking, missing n=573; alcohol intake, missing n=145) for calculation of FRAX. For calculation of SCORE, authors substituted history of fracture age greater than 55 years for the factor of "history of fracture at age greater than 45 years" but performed a sensitivity analysis suggesting there was minimal impact of this substitution.	None
Crandall et al, 2019[141]	FRAX	NR, reported by self-assessment questionnaire.	None
Dagan et al, 2017[169]	FRAX, QFracture	NR for FRAX; QFracture includes 3 categories of "current smokers," whereas the present study only includes 1 category of current smokers. All current smokers in the present study were assigned to the middle "current smokers" category from QFracture (10-19 cigarettes daily). Alcohol consumption was dichotomized, rather than categorized, and based on diagnoses of alcoholism or alcohol-induced chronic complications, rather than alcohol intake - individuals with alcohol-related diagnoses were assigned to QFracture's fourth level of alcohol consumption (7-9 units daily) and those without alcohol-related diagnoses were assigned to the "none" (no alcohol intake) category.	None

Appendix D Table 29. Risk of Bias for Included Studies for Key Question 2a (Domain 2 Predictors)

Author, Year	Risk Assessment Instrument	Deviations in Predictors Compared to Development Cohort	Domain 2 Comments
Davis et al, 2019[173]	QFracture	NR	None
Desbiens et al, 2020[177]	FRAX, QFracture	Most predictors were collected via a questionnaire or by patient self-report at baseline rather than from electronic health record data.	None
Ensrud et al, 2009[152] Premaor et al, 2013[153]	FRAX	NR	None
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146]	FRAX FRC QFracture	If data from FRAX questionnaire was missing, characteristic is set to null; they also set secondary osteoporosis risk factors (e.g., steroid use) to null as they had no consensus on diagnosis History of steroid use identified by drugs used within preceding 30 days only; 23.9% missing parental history of hip fracture Predictors were obtained via patient self-report during a baseline survey.	Had missing data for >25% of FRAX calculations; they did not utilize secondary osteoporosis risk factors
Fraser et al, 2011[154] Langsetmo et al, 2011[155]	FRAX Garvan	From Fraser et al[154] History of parental hip fracture was used for everyone with year 5 data, whereas history of any parental osteoporotic fracture was used from the baseline questionnaire for those without year 5 data. From Langsetmo et al[155] Used number of falls in preceding one-month as opposed to one year due to what was collected on survey	None
Goldshtein et al, 2018[172]	FRAX	All data from chart review, BMI and smoking history utilized from any data recorded during the study period if not available at time of DXA. Family history of osteoporosis used as proxy for parental hip fracture	Smoking data missing for 1.5% of sample; those with BMI data missing were excluded
Gonzalez-Macias et al, 2012[149]	FRAX	NR	None
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148]			Study conducted in the U.K.
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170]	QFracture FRAX	Material deprivation was categorized, rather than continuous, due to limitations in the study dataset[171] Parental history of fracture was not available; the study instead used a calculated weighted average of risks when assuming a parental hip fracture and by assuming absence of parental hip fracture based on a prevalence of parental hip fracture of 12%. Oral glucocorticoid use was alternatively defined by mean daily dose in the year before (<2.5, 2.5–7.5, and >7.5 mg/day).[170]	None

Appendix D Table 29. Risk of Bias for Included Studies for Key Question 2a (Domain 2 Predictors)

Author, Year	Risk Assessment Instrument	Deviations in Predictors Compared to Development Cohort	Domain 2 Comments
Jain et al, 2023[311]	FRAX, QFracture	All predictors based on EHR data and ICD-9 or 10 diagnostic codes; nursing home residence and parental history of hip fracture were poorly captured; algorithm used to determine secondary osteoporosis; for QFracture "other race" was used for Hispanic participants.	None
Lo et al, 2011[73] Pressman et al, 2011[165]	FRAX, FRC	Parental history of fracture, smoking/alcohol were all obtained from chart and assumed null if missing. Previous fracture obtained by insurance claims which only required 1 year previous enrollment	If BMI missing, assumed 25 (average of cohort); was missing for 26.3% of cohort
Lu et al, 2021[178]	FRAX	Data was used from 5 cohorts (UK Biobank, MrOS US, MrOS Sweden, SOF, CKB) so variability in predictor acquisition; some did not have data about parental fracture, alcohol use available	None
Marques et al, 2017[176]	FRAX	No deviations, appear to be similar by cohort	None
Pluskiewicz et al, 2023[312]	FRAX, Garvan	Data collected prospectively as part of a longitudinal cohort study; no details on how predictors were measured.	None
Tamaki et al, 2011[151]	FRAX	Alcohol intake was switched from dichotomous to continuous based on daily intake.	None
Tanaka et al, 2010[156]	FRAX	NR	None
Tebe Cordomi et al, 2013[144]	FRAX	Self-reported on FRAX variables; no deviations noted.	None
Zwart et al, 2023[313]	FRAX	Data collected via structured questionnaire as specified in FRAX.	None

Abbreviations: BMD=bone mineral density; BMI=body mass index; COPD=chronic obstructive pulmonary disease; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; ICD=International Classification of Diseases; NR=not reported; U.K.=United Kingdom; U.S.=United States.

Appendix D Table 30. Risk of Bias for Included Studies for Key Question 2a (Domain 3 Outcomes)

Author, Year	Risk Assessment Instrument	Deviations in Outcome Assessment Compared to Development Cohort	Was outcome determined appropriately in a similar way for all patients using a standard measure or definition?	Domain 3 Comments
Azagra et al, 2015[168, 310]	FRAX	NR	Yes/Probably yes	None
Bolland et al, 2011[166]	FRAX	Mean followup 8.8 years	No/Probably no	For first 5 years, fractures were self-reported and then confirmed by physician; after that only self-report
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159]	FRAX, OST	None, was determined by ICD codes, hip and forearm fractures also required procedure codes[158] Only required minimum 5 years' followup; however, mean followup was 10.5 years Based fracture incidence only on medical records[160]	Yes/Probably Yes	None
Chapurlat et al, 2020[175]	FRAX	One cohort had followup for 5 years only, the other for median 9.4 years	Yes/Probably yes	Fractures were confirmed with radiographs
Cheung et al, 2012[150]	FRAX	NR	Yes/probably yes	None
Collins et al, 2011[167]	QFracture	NR	Yes/Probably Yes	None
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143]	FRAX, Garvan, OST, SCORE	Self-reported fractures, with hip fractures being confirmed by medical records	Yes/Probably Yes for hip fractures; no/probably no for other fracture types	Self-report for non-hip fractures.
Dagan et al, 2017[169]	FRAX, QFracture	5-year fracture risks were calculated instead of 10-year risks	Yes/Probably yes	Preliminary analysis found that cumulative incidence of fractures is linear over a 1-year period. To calculate the 5-year fracture risk under this assumption, the 10-year risk scores were multiplied by 0.5.
Davis et al, 2019[173]	QFracture	NR	Yes/Probably yes	None
Desbiens et al, 2020[177]	FRAX, QFracture	Fracture data were collected using claims.	Yes/Probably yes	None

Appendix D Table 30. Risk of Bias for Included Studies for Key Question 2a (Domain 3 Outcomes)

Author, Year	Risk Assessment Instrument	Deviations in Outcome Assessment Compared to Development Cohort	Was outcome determined appropriately in a similar way for all patients using a standard measure or definition?	Domain 3 Comments
Ensrud et al, 2009[152] Premaor et al, 2013[153]	FRAX	None reported	Yes/Probably yes	None
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146]	FRAX FRC QFracture	Included traumatic fractures, authors stated this is because trauma is difficult to quantify Mean followup 8.4 years Fracture incidence was assessed by patient self-report using a Tri-Annual Questionnaire (every 4 months) and validated using electronic health record data.	No/Probably no	Includes traumatic fractures.
Fraser et al, 2011[154] Langsetmo et al, 2011[155]	FRAX Garvan	From Fraser et al[154]: NR From Langsetmo et al[155] Mean followup 8.3 years; all fractures self-reported annually	Yes/Probably yes	From Fraser et al[154]: None From Langsetmo et al[155] Followup visits in year 3 for those ages 40-60 years only, and at years 5 and 10 with all participants, but all fractures defined the same way and sent survey annually, unclear why 40-60 had extra visit
Goldshtein et al, 2018[172]	FRAX	Fractures obtained from billing, if multiple fractures coded at same encounter or within 6 months of motor vehicle accident were not included as thought more likely to be traumatic fractures, to avoid double-counting fractures only included different classes of fractures (hip, vertebral, nonhip-nonvertebral) as new events	Yes/Probably yes	None
Gonzalez-Macias et al, 2012[149]	FRAX	Clinical vertebral fractures were not measured for the cohort and therefore not included in the count of major osteoporotic fractures. Fracture risk was calculated for 3-year followup, rather than 10-year followup.	No/Probably no	Potential bias due to exclusion of vertebral fractures from MOF and 3-year followup
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148] Poor				

Appendix D Table 30. Risk of Bias for Included Studies for Key Question 2a (Domain 3 Outcomes)

Author, Year	Risk Assessment Instrument	Deviations in Outcome Assessment Compared to Development Cohort	Was outcome determined appropriately in a similar way for all patients using a standard measure or definition?	Domain 3 Comments
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170]	QFracture FRAX	NR	Yes/Probably yes	None
Jain et al, 2023[311]	FRAX, QFracture	Identification based on billing codes; excluded fractures within 30 days of codes associated with trauma.	Yes/Probably yes	None
Lo et al, 2011[73] Pressman et al, 2011[165]	FRAX	Median followup 6.6 years, fractures obtained by ICD codes	Yes/Probably yes	Only studied hip fractures; unenrolled if completed 1 year of bisphosphonate therapy or insurance unenrollment/lapse
Lu et al, 2021[178]	FRAX	Fractures from ICD codes for UK Biobank cohort, X-ray archives used for MrOS Sweden	No/Probably no	Variable between cohorts
Marques et al, 2017[176]	FRAX	Mean followup of 9.12 years	No information	All self-reported, they report that SAOL cohort also confirmed by clinical file review in all but 2 of 52 fractures, but unclear if these were excluded
Pluskiewicz et al, 2023[312]	FRAX, Garvan	Self-reported fractures reported by phone; no indication that fractures were confirmed.	No information.	None
Tamaki et al, 2011[151]	FRAX	NR	Yes/Probably yes	None
Tanaka et al, 2010[156]	FRAX	NR	Yes/Probably yes	None
Tebe Cordomi et al, 2013[144]	FRAX	Location/cause of fractures self-reported, not confirmed in all cases but did not report how frequently were confirmed	No/probably no	Self-reported and it reported that “not all cases” were confirmed, unclear how many; self-reports also were at the end of the 10 years, increasing risk of recall bias
Zwart et al, 2023[313]	FRAX	Fractures measured through medical record review and self-report.	Yes/Probably yes	None

Abbreviations: FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; ICD=International Statistical Classification of Diseases and Related Health Problems; NR=not reported; U.K.=United Kingdom; U.S.=United States.

Appendix D Table 31. Risk of Bias for Included Studies for Key Question 2a (Domain 4 Analysis)

Author, Year	Risk Assessment Instrument	Adequate number of hip fractures?	Adequate number of MOF?	Were continous predictors handled appropriately?	Were all enrolled participants included in the analysis?	Were participants with missing data handled appropriately?	Were relevant model performance measures evaluated appropriately?
Azagra et al, 2015[168, 310]	FRAX	No/Probably no	No/Probably no	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes
Bolland et al, 2011[166]	FRAX, Garvan	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159]	FRAX, OST	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no	Yes/Probably yes
Chapurlat et al, 2020[175]	FRAX	No/Probably no	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no
Cheung et al, 2012[150]	FRAX	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no
Collins et al, 2011[167]	QFracture	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143]	FRAX, Garvan, OST, SCORE	Yes/Probably Yes	Yes/Probably yes	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes for FRAX and Garvan; No/probably no for OST and SCORE
Dagan et al, 2017[169]	FRAX, QFracture	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Davis et al, 2019[173]	QFracture	No/Probably no	No information	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Desbiens et al, 2020[177]	FRAX, QFracture	No information	No information	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes

Appendix D Table 31. Risk of Bias for Included Studies for Key Question 2a (Domain 4 Analysis)

Author, Year	Risk Assessment Instrument	Adequate number of hip fractures?	Adequate number of MOF?	Were continous predictors handled appropriately?	Were all enrolled participants included in the analysis?	Were participants with missing data handled appropriately?	Were relevant model performance measures evaluated appropriately?
Ensrud et al, 2009[152] Premaor et al, 2013[153]	FRAX	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no	No information	No/Probably no
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146]	FRAX FRC QFracture	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/probably yes	Yes/Probably yes	Yes/Probably yes
Fraser et al, 2011[154] Langsetmo et al, 2011[155]	FRAX Garvan	No information No/Probably no	No information Yes/Probably yes	Yes/Probably yes Yes/Probably yes	Yes/Probably yes No/Probably no	No/Probably No No/Probably no	Yes/Probably yes Yes/Probably yes
Goldshtein et al, 2018[172]	FRAX	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Gonzalez-Macias et al, 2012[149]	FRAX	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148] Poor	QFracture	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170]	QFracture FRAX	Yes/Probably yes	Yes/Probably yes	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes
Jain et al, 2023[311]	FRAX, QFracture	No/Probably no	No/Probably no	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes
Lo et al, 2011[73] Pressman et al, 2011[165]	FRAX, FRC	Yes/Probably yes	No/Probably no	Yes/Probably yes	No/Probably no	Yes/Probably yes	Yes/Probably yes

Appendix D Table 31. Risk of Bias for Included Studies for Key Question 2a (Domain 4 Analysis)

Author, Year	Risk Assessment Instrument	Adequate number of hip fractures?	Adequate number of MOF?	Were continous predictors handled appropriately?	Were all enrolled participants included in the analysis?	Were participants with missing data handled appropriately?	Were relevant model performance measures evaluated appropriately?
Lu et al, 2021[178]	FRAX	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no	Yes/Probably yes	Yes/Probably yes
Marques et al, 2017[176]	FRAX	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes
Pluskiewicz et al, 2023[312]	FRAX, Garvan	No/Probably no	No/Probably no	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no
Tamaki et al, 2011[151]	FRAX	No/Probably no	No/Probably no	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes
Tanaka et al, 2010[156]	FRAX	No/Probably no	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	Yes/Probably yes
Tebe Cordomi et al, 2013[144]	FRAX	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes
Zwart et al, 2023[313]	FRAX	No/Probably no	No/Probably no	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no

Abbreviations: BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator.

Appendix D Table 32. Risk of Bias for Included Studies for Key Question 2a (Overall)

Author, Year	Risk Assessment Instrument	Risk of Bias/Study Quality for Hip Fracture Outcomes	Risk of Bias/Study Quality for MOF Outcomes	Comments
Azagra et al, 2015[168, 310]	FRAX	High/Poor	High/Poor	Potential bias due to lack of representation/validation across racial/ethnic groups, low incidence of hip fracture and MOF, and insufficient handling of missing data.
Bolland et al, 2011[166]	FRAX	High/Poor	High/Poor	Changed how they measured fractures during the study, <10 years' followup; predictors measured per development cohort; low number of hip fractures.
Brennan et al, 2014[158] Leslie et al, 2010[157] Leslie et al, 2016[161] Leslie et al 2018[162] Crandall et al, 2019[160] Morin et al, 2009[159]	FRAX, OST	High/Poor	High/Poor	Use of diagnosis codes instead of participant report of smoking/alcohol use; only included subjects with all necessary data in retrospective design
Chapurlat et al, 2020[175]	FRAX	High/Poor	High/Poor	<100 fracture events and <10 years' followup in both cohorts.
Cheung et al, 2012[150]	FRAX	High/Poor	Unclear/Fair	Potential bias for both hip fracture and MOF due to lack of representation and validation across racial/ethnic groups and failure to report sufficient calibration measures. Additional bias for hip fractures due to insufficient fracture incidence.
Collins et al, 2011[167]	QFracture	Unclear/Fair	Unclear/Fair	Potential bias due to lack of representation and validation across racial/ethnic groups.
Crandall et al, 2014[139] Crandall et al, 2018[140] Crandall et al, 2019[141] Crandall et al, 2019[142] Crandall et al, 2023[143]	FRAX, Garvan, OST, SCORE	High/Poor	High/Poor	Mostly White sample, very little information on missing data for risk factors and excluded participants with less than 10 years of followup; only hip fractures verified
Dagan et al, 2017[169]	FRAX	Unclear/Fair	Unclear/Fair	Potential sources of bias include inappropriate categorization of smoking and alcohol intake predictor variables; lack of representation and validation across racial/ethnic groups.
Davis et al, 2019[173]	QFracture	High/Poor	Unclear/Fair	Potential bias due to lack of representation and validation across racial/ethnic groups and low incidence of hip fractures in the study population.

Appendix D Table 32. Risk of Bias for Included Studies for Key Question 2a (Overall)

Author, Year	Risk Assessment Instrument	Risk of Bias/Study Quality for Hip Fracture Outcomes	Risk of Bias/Study Quality for MOF Outcomes	Comments
Desbiens et al, 2020[177]	FRAX	Unclear/Fair	Unclear/Fair	Potential bias due to lack of representation and validation across racial/ethnic groups; lack of information about handling of predictors variables compared to the original development.
Ensrud et al, 2009[152] Premaor et al, 2013[153]	FRAX	High/Poor	High/Poor	Potential bias due to lack of representation and validation across racial/ethnic groups as well as inappropriate exclusion of Black women. Additional bias due to lack of reporting about the handling of missing data and insufficient calibration outcomes.
Ettinger et al, 2013[145] Ettinger et al, 2012[72] Gourlay et al, 2017[146] Gourlay et al, 2017[146]	FRAX FRC QFracture	High/Poor	High/Poor	Missing information for >25% of FRAX calculations. Included traumatic fractures in outcome, excluded data for persions missing a BMD measure at Year 7. Potential bias due to lack of accounting for missing data and exclusion of men with fracture or treatment at baseline, who were included in the QFracture development cohort.
Fraser et al, 2011[154] Langsetmo et al, 2011[155]	FRAX Garvan	High/Poor	High/Poor	Potential bias due to lack of representation and validation across racial/ethnic groups, presumably low fracture incidence, and insufficient handling of missing data. Did not have 10-year followup, had different definition for fall predictor (1 month vs. 1 year), excluded 15% for missing data.
Goldshtein et al, 2018[172]	FRAX	Unclear/Fair	Unclear/Fair	All data from chart review/claims data, data handled appropriately, many fractures.
Gonzalez-Macias et al, 2012[149]	FRAX	High/Poor	High/Poor	Potential bias due to 3-year followup for fracture incidence and inappropriate handling of participants with missing outcome data. Additional bias for hip fracture due to insufficient number of hip fracture incidences.
Hippisley-Cox et al, 2012[147] Hippisley-Cox et al, 2009[148] Poor	QFracture	Unclear/Fair	Unclear/Fair	None
Hippisley-Cox et al, 2014[171] Klop et al, 2016[170]	QFracture FRAX	Unclear/Fair	Unclear/Fair	Potential bias due to lack of representation of racial/ethnic groups and inappropriate categorization of the material deprivation predictor.
Jain et al, 2023[311]	FRAX, QFracture	High/Poor	High/Poor	Excluded participants with unknown or mixed race, and Asian participants. Did not have 10 years of followup data so adjusted risk based on amount of followup data available; all inputs based on data captured in EHR and billing codes; no information on missing data; insufficient number of fractures.

Appendix D Table 32. Risk of Bias for Included Studies for Key Question 2a (Overall)

Author, Year	Risk Assessment Instrument	Risk of Bias/Study Quality for Hip Fracture Outcomes	Risk of Bias/Study Quality for MOF Outcomes	Comments
Lo et al, 2011[73] Pressman et al, 2011[165]	FRC	High/Poor	High/Poor	Did not measure MOF, only 6.6 years of followup, and a lot of missing data although participants were still included in analysis. All of the variables for FRC were determined from chart review and assumed null if missing/BMI set to 25 if missing (for 26.3% of sample), making it more difficult to determine true value of FRC. Had significant number of hip fractures and had relatively diverse sample.
Lu et al, 2021[178]	FRAX	High/Poor	High/Poor	Data from multiple cohorts, which acquired data (both predictors and outcomes) in different ways, median followup not reported, although noted MrOS U.S. cohort had only 4 years' followup.
Marques et al, 2017[176]	FRAX	High/Poor	High/Poor	Significant number of participants excluded for loss to followup with no statistical attempts to account for missing data, unclear if some outcomes were confirmed by clinician, low number of hip fractures.
Pluskiewicz et al, 2023[312]	FRAX, Garvan	High/Poor	High/Poor	Insufficient number of fractures; no information about how data on predictors were collected; used self-reported fractures, does not account for complexities in data or analysis; insufficient calibration evaluation.
Tamaki et al, 2011[151]	FRAX	High/Poor	High/Poor	Potential bias due to lack of representation and validation across racial/ethnic groups, low incidence of hip and major osteoporotic fractures, inappropriate handling of missing data, and inappropriate exclusion of older participants.
Tanaka et al, 2010[156]	FRAX	High/Poor	High/Poor	Potential bias due to lack of representation and validation across racial/ethnic groups, low incidence of hip fracture and MOF, and inappropriate handling of missing data.
Tebe Cordomi et al, 2013[144]	FRAX	High/Poor	High/Poor	The majority of enrolled subjects were not included in analysis as they did not answer phone for survey, did not detail other ways to try to recover missing data or account for this in analysis. Fractures were all self-reported and not confirmed; participants were called at the end of the 10 years to discuss if did not followup. Only 13 hip fractures reported.
Zwart et al, 2023[313]	FRAX	High/Poor	High/Poor	Insufficient number of fractures; discrimination measures not reported.

Abbreviations: BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; FRC=Fracture Risk Calculator; MOF=major osteoporotic fracture; WHI=Women's Health Initiative.

Appendix D Table 33. Risk of Bias for Systematic Reviews Included for Key Question 2a (Domain 1 Study Eligibility)

Author, Year	1.1 Did the review adhere to predefined objectives and eligibility criteria?	1.2 Were the eligibility criteria appropriate for the review question?	1.3 Were eligibility criteria unambiguous?	1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate (e.g., date, sample size, study quality, outcomes measured)?	1.5 Were any restrictions in eligibility criteria based on sources of information appropriate (e.g., publication status or format, language, availability of data)?	Concerns Regarding Specification of Study Eligibility Criteria	Rationale for Concern
Beaudoin et al, 2019[135]	Yes	Yes	Yes	Yes	Yes	Low	None
Crandall, 2015[426]	Probably yes	Yes	Probably no	No	No information	High	Restricted to studies in U.S. or Canada, no specification on BMD T-score measurement or anatomical site, or parameters on fracture outcome measurement or length of time for prediction, no specification on referral clinic.
Jiang et al, 2017[136]	Yes	Probably yes	Probably yes	Probably yes	Probably yes	Low	English language only, required studies to report Sn and Sp or data able to derive these values; studies only reporting AUC were excluded. Since the objectives were to assess specific U.S. thresholds for FRAX, this restriction is probably reasonable.
Marques et al, 2015[134]	Yes	Yes	Yes	Yes	Yes	Low	None

Abbreviations: AUC=area under the curve; BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; Sn=sensitivity; Sp=specificity; U.S.=United States.

.

Appendix D Table 34. Risk of Bias for Systematic Reviews Included for Key Question 2a (Domain 2 Identification and Selection of Studies)

Author, Year	2.1 Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?	2.2 Were methods additional to database searching used to identify relevant reports?	2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	2.4 Were restrictions based on date, publication format, or language appropriate?	2.5 Were efforts made to minimize error in selection of studies?	Concerns Regarding Methods Used to Identify and/or Select Studies	Rationale for Concern
Beaudoin et al, 2019[135]	Yes	Yes	Yes	Yes	Yes	Low	None
Crandall, 2015[426]	No	No	No	Probably yes	No	High	Only 1 database searched, no supplemental methods used to identify relevant studies, unclear whether terms used were employed for controlled vocabulary or were used also as text words, single reviewer screened studies, which could lead to errors.
Jiang et al, 2017[136]	Yes	Probably yes	Probably yes	Yes	Yes	Low	None
Marques et al, 2015[134]	Yes	Yes	Yes	Yes	Yes	Low	None

Appendix D Table 35. Risk of Bias for Systematic Reviews Included for Key Question 2a (Domain 3 Data Collection and Study Appraisal)

Author, Year	3.1 Were efforts made to minimize error in data collection?	3.2 Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	3.3 Were all relevant study results collected for use in the synthesis?	3.4 Was ROB (or methodological quality) formally assessed using appropriate criteria?	3.5 Were efforts made to minimize error in ROB assessment?	Concerns Regarding Methods Used to Collect Data and Appraise Studies	Rationale for Concern
Beaudoin et al, 2019[135]	Yes	Probably yes	Probably yes	Probably no	Yes	Unclear	ROB assessed using an adapted version of QUADAS, which is designed for diagnostic accuracy, not predictive accuracy. The adaptations are likely not sufficient to address predictive accuracy.
Crandall, 2015[426]	No	Yes	Probably yes	No	No information	High	No assessment of ROB for included studies, a single reviewer extracted all data, which could lead to data errors.
Jiang et al, 2017[136]	No information	Yes	Yes	Probably no	No information	Unclear	No mention of dual independent ROB; used QUADAS to assess ROB, which is not designed for predictive accuracy.
Marques et al, 2015[134]	Yes	Probably yes	Yes	Yes	Yes	Low	None

Abbreviations: QUADAS=Quality Assessment of Studies of Diagnostic Accuracy; ROB=risk of bias.

Appendix D Table 36. Risk of Bias for Systematic Reviews Included for Key Question 2a (Domain 4 Synthesis and Findings)

Author, Year	4.1 Did the synthesis include all studies that it should?	4.2 Were all predefined analyses reported or departures explained?	4.3 Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?	4.4 Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	4.5 Were the findings robust (e.g., as demonstrated through funnel plot or sensitivity analyses)?	4.6 Were biases in primary studies minimal or addressed in the synthesis?	Concerns Regarding the Synthesis and Findings	Rationale for Concern
Beaudoin et al, 2019[135]	Yes	Yes	Yes	Yes	Yes	No information	Low	Done
Crandall, 2015[426]	Probably yes	Probably yes	Probably yes	No information	No information	No information	Unclear	Restricted to studies in U.S. or Canada, no specification on BMD T-score measurement or anatomical site, or parameters on fracture outcome measurement or length of time for prediction, no specification on referral clinic.
Jiang et al, 2017[136]	Yes	Yes	Yes	Yes	Yes	Yes	Low	English language only, required studies to report Sn and Sp or data able to derive these values; studies only reporting AUC were excluded. Since the objectives were to assess specific U.S. thresholds for FRAX, this restriction is probably reasonable.
Marques et al, 2015[134]	Yes	Probably yes	Yes	Probably no	Probably no	Probably yes	Low	None

Abbreviations: AUC=area under the curve; BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; Sn=sensitivity; Sp=specificity; U.S.=United States.

Appendix D Table 37. Risk of Bias for Systematic Reviews Included for Key Question 2a (Overall Risk of Bias)

Author, Year	A. Did the interpretation of findings address all of the concerns identified in Domains 1 through 4?	B. Was the relevance of identified studies to the review's research question appropriately considered?	C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?	ROB/Study Quality in the Review	Rationale for Risk
Beaudoin et al, 2019[135]	Yes	Yes	Yes	Low/Good	The only concern is that the authors used QUADAS to evaluate ROB; although it was adapted for this review, it may not be as appropriate as an ROB tool specifically designed for predictive accuracy/prognosis studies.
Crandall, 2015[426]	Probably no	Yes	Yes	High/Poor	Single author, which increases the chances for error in study selection and data extraction; no ROB assessment of included studies. Serious flaws in search strategy, restricted to studies in 2 countries without clear rationale.
Jiang et al, 2017[136]	Probably yes	Yes	Yes	Some concerns/Fair	ROB evaluated using QUADAS, which may not have been appropriate; excluded studies that reported AUC since primary interest was in evaluating Sn and Sp of a specific threshold.
Marques et al, 2015[134]	Yes	Yes	Yes	Low/Good	None

Abbreviations: AUC=area under the curve; QUADAS=Quality Assessment of Studies of Diagnostic Accuracy; ROB=risk of bias; Sn=sensitivity; Sp=specificity.

Appendix D Table 38. Risk of Bias of Included Studies for Key Question 2b (Domain 1 Participants)

Author, Year Fractures Reported	1.1 Were appropriate data sources used?	1.2 Were all inclusions and exclusions of participants appropriate?	Domain 1 ROB	Domain 1 ROB Rationale	Domain 1 Applicability: Concern that the included participants and setting do not match the review question?	Domain 1 Applicability Rationale
Baleanu et al, 2021[179]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Black et al, 2018[192]	Yes/Probably yes	Yes/Probably yes	Unclear	More than 25% had h/o prior fracture since age 50 years.	Low	None
Bolland et al, 2011[166]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Chapurlat et al, 2020[175]	Yes/Probably yes	Yes/Probably yes	Unclear	Analysis based on data collected from both 2 population-based cohorts.	Low	None
Cheung et al, 2012[150]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Fraser et al, 2011[154]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Goldshtein et al, 2018[172]	Yes/Probably yes	Yes/Probably yes	Unclear	Retrospective analysis of data from electronic health records of a government-funded health maintenance organization.	Low	None
Gourlay et al, 2017[146]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Iki et al, 2021[194]	Yes/Probably yes	yes/probably yes	Low	None	Low	None
Kwok et al, 2012[182]	yes/probably yes	Yes/Probably yes	Low	None	Low	None
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193]	Yes/probably yes	Yes/Probably yes	Unclear	Retrospective analysis based on data from a registry of persons who were referred for DXA within a single geographic area.	Unclear	Referral population

Appendix D Table 38. Risk of Bias of Included Studies for Key Question 2b (Domain 1 Participants)

Author, Year Fractures Reported	1.1 Were appropriate data sources used?	1.2 Were all inclusions and exclusions of participants appropriate?	Domain 1 ROB	Domain 1 ROB Rationale	Domain 1 Applicability: Concern that the included participants and setting do not match the review question?	Domain 1 Applicability Rationale
Marques et al, 2017[176]	No information	No information	Unclear	Analysis used data from preexisting cohort studies, not all of which appeared to have been designed to assess the relationship between BMD and fracture; one cohort included high proportion with secondary osteoporosis; DXA in some cohorts was at the discretion of clinicians, very little detail on inclusion/exclusion criteria for the 3 cohorts that were used in this analysis.	Low	None
Nguyen et al, 2004[186]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Prince et al, 2019[191]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Robbins et al, 2007[187]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Sornay-Rendu et al, 2010[188]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Stewart et al, 2006[190]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Sund et al, 2014[185]	No/Probably no	No/Probably no	Unclear	Retrospective analysis of data from a longitudinal cohort study; patients who died or had hip fracture before the first 5-year followup were excluded; women without FRAX variable information were excluded, and only a subset of women with BMD information were included (two-thirds of those with BMD were a random sample, the other third was not to ensure the inclusion of prespecified risk factors).	Low	None
Tamaki et al, 2011[151]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None

Appendix D Table 38. Risk of Bias of Included Studies for Key Question 2b (Domain 1 Participants)

Author, Year Fractures Reported	1.1 Were appropriate data sources used?	1.2 Were all inclusions and exclusions of participants appropriate?	Domain 1 ROB	Domain 1 ROB Rationale	Domain 1 Applicability: Concern that the included participants and setting do not match the review question?	Domain 1 Applicability Rationale
Tanaka et al, 2010[156]	Yes/Probably yes	Yes/Probably yes	Unclear	Datasets for the present analysis derived from three preexisting cohort studies, none of which were specifically focused on osteoporosis, BMD, or fracture. Although data were evaluated retrospectively, the predictor and outcome data were collected prospectively.	Low	None
Trajanoska et al, 2018[15]	Yes/Probably yes	Yes/Probably yes	Unclear	Used data from a prexisting cohort study that was designed to follow adults age 45 years or older for the development of a variety of conditions and was not necessarily focused specifically on osteoporosis or fractures specifically.	Low	None
Tremollieres et al, 2010[189]	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; FRAX=Fracture Risk Assessment Tool; h/o=history of; MOF=major osteoporotic fracture; ROB=risk of bias.

Appendix D Table 39. Risk of Bias of Included Studies for Key Question 2b (Domain 2 Predictors)

Author, Year	2.1 Were predictors defined and assessed in a similar way for all participants?	2.2 Were predictor assessments made without knowledge of outcome data?	2.3 Are all predictors available at the time the model is intended to be used?	Domain 2 ROB	Domain 2 ROB Rationale	Domain 2 Applicability: Concern that the definition, assessment or timing of predictors in the model do not match the review question?	Domain 2 Applicability Rationale
Baleanu et al, 2021[179]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Black et al, 2018[192]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Bolland et al, 2011[166]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Chapurlat et al, 2020[175]	No/Probably no	Yes/Probably yes	Yes/Probably yes	Unclear	Different DXA machines were used by the different cohorts; no discussion of whether cross-calibration occurred.	Low	None
Cheung et al, 2012[150]	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR but likely since prospectively conducted and BMD was collected at baseline.	Low	None
Fraser et al, 2011[154]	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR but likely since BMD was measured at baseline.	Low	None
Goldshtein et al, 2018[172]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Gourlay et al, 2017[146]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Iki et al, 2021[194]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Kwok et al, 2012[182]	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding of outcome data NR, but likely since BMD was captured at baseline in this prospective study.	Low	None

Appendix D Table 39. Risk of Bias of Included Studies for Key Question 2b (Domain 2 Predictors)

Author, Year	2.1 Were predictors defined and assessed in a similar way for all participants?	2.2 Were predictor assessments made without knowledge of outcome data?	2.3 Are all predictors available at the time the model is intended to be used?	Domain 2 ROB	Domain 2 ROB Rationale	Domain 2 Applicability: Concern that the definition, assessment or timing of predictors in the model do not match the review question?	Domain 2 Applicability Rationale
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193]	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR but likely since BMD was performed at entry into the registry.	Low	None
Marques et al, 2017[176]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Nguyen et al, 2004[186]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Robbins et al, 2007[187]	Yes/Probably yes	Yes/Probably Yes	Yes/Probably yes	Low	None	Low	None
Prince et al, 2019[191]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Sornay-Rendu et al, 2010[188]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Stewart et al, 2006[190]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Sund et al, 2014[185]	No information	No information	Yes/Probably yes	Unclear	No information about how BMD was assessed in this paper; assume it is described in the main papers describing the assembly of the cohort.	Unclear	No information about how BMD was assessed in this paper, assume it is describe in the main study papers.

Appendix D Table 39. Risk of Bias of Included Studies for Key Question 2b (Domain 2 Predictors)

Author, Year	2.1 Were predictors defined and assessed in a similar way for all participants?	2.2 Were predictor assessments made without knowledge of outcome data?	2.3 Are all predictors available at the time the model is intended to be used?	Domain 2 ROB	Domain 2 ROB Rationale	Domain 2 Applicability: Concern that the definition, assessment or timing of predictors in the model do not match the review question?	Domain 2 Applicability Rationale
Tamaki et al, 2011[151]	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR but likely since data on BMD was collected at baseline in this prospective study.	Low	None
Tanaka et al, 2010[156]	No/Probably no	No information	Yes/Probably yes	Unclear	Different DXA machines were used in the different cohorts; no information about reference ranges used to calculate T-scores in 2 of the cohorts; blinding not explicitly mentioned but likely since BMD was collected at baseline in all cohorts.	Low	None
Trajanoska et al, 2018[15]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Tremollieres et al, 2010[189]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None

Abbreviations: BMD=body mass index; DXA=dual-energy X-ray absorptiometry; NR=not reported; NV=nonvertebral; ROB=risk of bias; vs.=versus.

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Baleanu et al, 2021[179]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	No information about blinding	Low	None
Black et al, 2018[192]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no	Unclear	Prediction was made over 25 years; it is not clear whether this is an appropriate interval given the significant change in health status that could occur over that length of time.	Low	None
Bolland et al, 2011[166]	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	During the main trial, fractures were confirmed by radiographic reports, but during the extension part of the study, fractures were ascertained based on self-report.	Low	None
Chapurlat et al, 2020[175]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Cheung et al, 2012[150]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding was NR; since fractures were confirmed ascertainment unlikely to have been influenced by knowledge of BMD status.	Low	None
Fraser et al, 2011[154]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR but unlikely to impact since fractures required confirmation.	Low	None
Goldshtein et al, 2018[172]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR, but ascertain-ment based on clinical records, so likely minimal to no ROB.	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Gourlay et al, 2017[146]	No information	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	Specific fracture ascertainment methods NR in this paper, likely reported in other paper describing the MrOs cohort. No information about whether fracture ascertainment was blinded to baseline predictors (e.g., BMD, Fracture Risk Score).	Low	None
Iki et al, 2021[194] Hip Validation	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Self-reported fracture data, though it was based on participant report of X-ray confirmation via nurse interview.	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Kwok et al, 2012[182]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding to BMD data NR, but knowledge of it unlikely to affect fracture ascertainment since fractures were confirmed.	Low	None
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Blinding NR, but likely minimal impact since fractures based on claims data	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Marques et al, 2017[176]	No information	Yes/Probably yes	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes	High	Fractures were only confirmed by clinical review in 1 of the 3 cohorts; fracture confirmation not reported in the other 2 cohorts, some cohorts included traumatic fractures.	Low	None
Nguyen et al, 2004[186]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	No information	High	No information about length of followup over which fractures were being predicted	Low	None
Prince et al, 2019[191]	No information	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	Administrative hospital data used to identify fractures so unclear whether all relevant fractures would be identified	Low	None
Robbins et al, 2007[187] Hip	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Sornay-Rendu et al, 2010[188]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None
Stewart et al, 2006[190]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None
Sund et al, 2014[185]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None
Tamaki et al, 2011[151]	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	Self-reported fractures diagnosed by X-ray; unclear whether fractures were confirmed through medical records or radiographs	Low	None
Tanaka et al, 2010[156]	No information	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	Fractures assessed annually but no mention of whether reported fractures were confirmed with X-rays or medical record review	Low	None

Appendix D Table 40. Risk of Bias of Included Studies for Key Question 2b (Domain 3 Outcomes)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applicability Rationale
Trajanoska et al, 2018[15]	Yes/Probably yes	No information	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Unclear	No information about blinding of outcome assessment to BMD status, but given that data on fractures were obtained through existing records, likelihood of bias is low; did not discuss whether traumatic fractures were excluded	Low	None
Tremollieres et al, 2010[189]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	Low	None	Low	None

Abbreviations: BMD=bone mineral density; NR=not reported; ROB=risk of bias; vs.=versus.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Baleanu et al, 2021[179]	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	No information	No/Probably no	High	No calibration outcomes reported for BMD alone as a risk; participants with missing data were significantly older and had higher history of personal and parental fracture. Unclear whether complexities in the data were accounted for. Finally, only observed 47 hip fractures, so not enough events for hip fracture prediction.
Black et al, 2018[192]	Yes/Probably yes	Yes/Probably yes	No/Probably no	Yes/Probably yes	Yes/Probably yes	No/Probably no	Unclear	10.7% of enrolled participants terminated followup; study only reported on calibration outcomes; discrimination outcomes not reported.
Bolland et al, 2011[166]	No/Probably no	No information	No/Probably no	No information	No information	No/Probably no	Unclear	Only had 57 hip fractures, which is not sufficient; however, had sufficient number of fracture events for MOF and Garvan OF prediction; did not report calibration outcomes for BMD alone as a predictor; unclear whether complexities in the data were accounted for; no discussion of how/whether missing data were handled.
Chapurlat et al, 2020[175]	No/Probably no	Yes/Probably yes	No information	No information	No/Probably no	No/Probably no	High	Did not have sufficient number of events for MOF; complexities in the data not discussed; no calibration outcomes reported; however, this was not the main focus of the analysis.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Cheung et al, 2012[150]	Yes/Probably yes	Yes/Probably yes	No information	No information	Yes/Probably yes	Yes/Probably yes	Unclear	Not enough events for hip fracture prediction; barely enough events (N=106) for MOF prediction. T-score appears to have been modeled continuously. No information about missing data.
Fraser et al, 2011[154]	No/Probably no	Yes/Probably yes	No information	No information	Yes/Probably yes	No/Probably no	Unclear	Did not have enough fracture events for hip fractures in men, sufficient number of events for other fractures and hip fractures in women; no information about missing data; no calibration plots for BMD alone but this was not the focus of this analysis.
Goldshtein et al, 2018[172]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no	No/Probably no	Unclear	Analysis based on logistic regression and did not manage the complexities of the data (i.e., censoring, competing risk); no calibration plots for BMD alone as a predictor, but this was not the focus of the analysis.
Gourlay et al, 2017[146]	Yes/Probably yes	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes	Yes/Probably yes	Unclear	Some participants were excluded from the analysis with BMD because they developed osteoporosis and later had a hip fracture during study followup; it is unclear why such participants would be excluded.
Iki et al, 2021[194]	No/Probably no	Yes/Probably yes	No/Probably no	No/Probably no	Yes/Probably yes	No/Probably no	Unclear	Insufficient number of fracture events; excluded women because of no followup data; did not report calibration plots.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Kwok et al, 2012[182]	No/Probably no	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes	No/Probably no	Unclear	Did not have enough events for prediction of hip or MOF fragility fractures; did not report calibration plots for BMD alone for prediction but this was not the focus of the study; reports that persons with missing DXA were excluded but number of missing persons not quantified and no information comparing those excluded with those included.
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193]	Yes/Probably yes	Yes/Probably yes	No information	No information	Yes/Probably yes and no information for some of the reports	No/Probably no	Unclear	Some reports with no information about missing data or lost to followup; some reports with no information about how complexities in the data handled for the specific question related to BMD relationship to fracture risk; no calibration outcomes reported for most of these reports; however, this was not the focus of the study. Not enough fracture events for prediction of hip fracture in men in some of the analyses that reported on men.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Marques et al, 2017[176]	No/Probably no	Yes/Probably yes	No information	No information	Yes/Probably yes	No/Probably no	Unclear	Calibration outcomes not reported, although this was not a focus of the analysis; did not have enough fracture events for hip fractures; methods of handling missing data not reported in 2 cohorts; the third cohort specifically did not use imputation for missing data.
Nguyen et al, 2004[186]	No/Probably no	No/Probably no	Yes/Probably yes	No information	No/Probably no	No/Probably no	High	Only 77 fracture events, which is not sufficient; analysis not designed for complexities in the data (i.e., censoring, competing risks), categorized BMD predictor by SD.
Prince et al, 2019[191]	No/Probably no	Yes/Probably yes	Yes/Probably yes	No information	Yes/Probably yes	No/Probably no	Unclear	Did not have enough fracture events for vertebral fractures; had enough events for the other fracture types reported. No mention of missing data. Did not report any calibration outcomes for BMD alone, but this was not the focus of the study.
Robbins et al, 2007[187]	No/Probably no	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No/Probably no	No/Probably no	Unclear	Fewer than 100 fracture events; unclear whether accounted for competing risks; participants without 5 years of followup data were excluded; did not report calibration for model with BMD, but evaluation of BMD alone was not the primary purpose of the analysis.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Sornay-Rendu et al, 2010[188]	No/Probably no	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no	No/Probably no	Unclear	Insufficient number of fracture events for hip, forearm, and clinical vertebral fracture; not clear whether complexities in the data were handled appropriately (i.e., censoring, competing risks); no information about missing data; calibration outcomes for BMD alone not reported; however, this was not the main focus of the analysis.
Stewart et al, 2006[190]	Yes/Probably yes	Yes/Probably yes	Yes/Probably yes	No information	No/Probably no	No/Probably no	Unclear	No information about whether/how missing data handled; no calibration plots for BMD alone; however, this was not the focus of the study; unclear how complexities in data handled, specifically competing risks.
Sund et al, 2014[185]	No/Probably no	Yes/Probably yes	No information	No information	No information	No/Probably no	High	Only 21 hip fracture events in women with BMD measurement; analysis does not account for the complexities in data (e.g., censoring); unclear whether any missing data in the subset of 2,755 relevant to the assessment of BMD prediction of fracture risk.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Tamaki et al, 2011[151]	No/Probably no	Yes/Probably yes	No/Probably no	No information	No/Probably no	No/Probably no	High	No blinding (though likely minimal impact on ROB); rare fracture events, did not use Cox regression to account for censoring; many women not included for missing data at baseline or lost to followup and no analysis of impact on results; no calibration results reported for BMD alone as a predictor.
Tanaka et al, 2010[156]	Yes/ Probably yes	Yes/Probably yes	No information	No information	No/Probably no	No/Probably no	Unclear	No calibration outcomes reported, but focus of study was not on BMD alone; not enough fracture events for hip fracture prediction; did not use Cox regression or survival modeling to account for censoring; unclear whether there was any missing data for BMD or fracture ascertainment.
Trajanoska et al, 2018[15]	Yes/Probably yes	No/Probably no	No/Probably no	No information	No/Probably no	No/Probably no	High	No AUC reported so did not handle BMD continuously; however, we were able to calculate Sn and Sp based on data provided; no mention of how competing risks were handled; did not provide calibration plots; excluded 23% of participants because of missing data at baseline.

Appendix D Table 41. Risk of Bias of Included Studies for Key Question 2b (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Tremollieres et al, 2010[189]	Yes/probably yes	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes	No/Probably no	High	Missing data for nearly half of the population that was enrolled at baseline; no calibration plots for BMD alone as a predictor; however, this was not the focus of the study.

Abbreviations: AUC=area under the curve; BMD=bone mineral density; DXA=DXA=dual-energy X-ray absorptiometry; MOF=major osteoporotic fracture; NR=not reported; OF=osteoporotic fracture; ROB=risk of bias; SD=standard deviation.

Appendix D Table 42. Risk of Bias of Included Studies for Key Question 2b (Overall)

Author, Year	Overall Study Quality	Overall ROB Rationale	Concerns Overall Applicability	Overall Applicability Rationale
Baleanu et al, 2021[179]	Poor	Participants with missing data were significantly older and had higher history of personal and parental fracture. Unclear whether complexities in the data were accounted for. Finally, only observed 47 hip fractures, so not enough events for hip fracture prediction.	Low	None
Black et al, 2018[192]	Fair	None	Low	None
Bolland et al, 2011[166]	Fair	Unclear concerns in the analysis domain (insufficient number of events for hip fractures, missing data, and complexities in the data).	Low	None
Chapurlat et al, 2020[175]	Poor	Participant data from 2 preexisting cohorts; different DXA machines used in the different cohorts; did not have sufficient number of events for MOF; complexities in the data not discussed; no calibration outcomes reported; however, this was not the main focus of the analysis.	Low	None
Cheung et al, 2012[150]	Fair	No information about blinding; no information about missing data.	Low	None
Fraser et al, 2011[154]	Fair	No information about missing data; limited information about calibration, blinding NR.	Low	None
Goldshtein et al, 2018[172]	Fair	Unclear ROB in the analysis domain because complexities in the data were not managed.	Low	None
Gourlay et al, 2017[146]	Fair	Some concerns for bias in the analysis domain.	Low	None
Iki et al, 2021[194]	Poor	High ROB in the data analysis domain because of insufficient number of fracture events, exclusion of women with missing data, and limited calibration outcomes reported.	Low	None
Kwok et al, 2012[182]	Fair	Did not report calibration plots; blinding of data NR; no information about missing data; did not have enough events for prediction of major nonvertebral fractures.	Low	None
Leslie et al, 2010[157] Hans et al, 2011[183] Leslie et al, 2013[184] Leslie et al, 2016[161] Leslie et al, 2018[162] Crandall et al, 2019[160] Agarawal et al, 2022[193]	Fair	Unclear ROB because of retrospective cohort based on referral population; blinding NR; no mention of missing data in some reports; no calibration plots for BMD alone prediction; not enough events for prediction of hip fracture in men in some reports.	Low	None
Marques et al, 2017[176]	Poor	High ROB in the patient selection (preexisting cohorts used with little detail and did not appear focused on BMD/fracture relationship evaluation), outcome (clinical confirmation of fracture not reported in 2 of the cohorts), and data analysis domains (not enough hip fracture events, no mention of how missing data were handled, no calibration outcomes reported).	Low	None
Nguyen et al, 2004[186]	Poor	Interval for prediction NR: insufficient number of fracture events, insufficient analysis for complexities in data, continuous predictor (BMD) not handled correctly.	Low	None

Appendix D Table 42. Risk of Bias of Included Studies for Key Question 2b (Overall)

Author, Year	Overall Study Quality	Overall ROB Rationale	Concerns Overall Applicability	Overall Applicability Rationale
Prince et al, 2019[191]	Fair	Used administrative hospital data to identify fractures and unclear how accurate/reliable this method is; no mention of missing data; did not have enough fracture events for clinical vertebral fractures; no calibration outcomes reported.	Low	None
Robbins et al, 2007[187]	Fair	Some concerns for bias in the analysis domain.	Low	None
Sornay-Rendu et al, 2010[188]	Fair	Unclear ROB in the analysis domain (insufficient number of fracture events for some fracture types, complexities in the data, and missing data handling).	Low	None
Stewart et al, 2006[190]	Fair	Unclear ROB in the analysis domain (unclear how missing data and complexities in the data handled).	Low	None
Sund et al, 2014[185]	Poor	Retrospectively assembled data with unclear inclusions/exclusions; blinding NR; insufficient number of fracture events; did not account for the complexities of data analysis; unclear missing data.	Low	None
Tamaki et al, 2011[151]	Poor	High ROB in the analysis domain because of missing data, rare fracture events, failure to account for complexities in the data during analysis, and absence of calibration results for BMD alone as a predictor.	Low	None
Tanaka et al, 2010[156]	Poor	Unclear ROB across all domains, analyzed data from preexisting cohorts; different DXA machines used with no information about calibration or reference ranges used to calculate T-scores; no mention of whether self-reported fractures were confirmed; did not account for the complexities of the data in the analysis; unclear whether any missing data; insufficient number of fracture events for some outcomes.	Low	None
Trajanoska et al, 2018[15]	Poor	Unclear ROB in patient selection domain (use of data from a previous cohort study not focused on osteoporosis; Outcome domain (outcome definition unclear as to whether excluded traumatic fractures); high ROB in the analysis domain because of missing data (~23% excluded because of missing BMD data); lack of reporting of relevant measures and handling of competing risks in the analysis.	Low	None
Tremollieres et al, 2010[189]	Poor	High ROB in the analysis domain (missing data).	Low	None

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; NR=not reported; ROB=risk of bias.

Appendix D Table 43. Risk of Bias of Included Studies for Key Question 2c (Domain 1 Patient Selection)

Author, Year	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Could the selection of patients have introduced bias?	Comments
Adler et al, 2003[221]	Unclear	Yes	Yes	Unclear	Analysis of persons enrolled in two cross-sectional studies; participants recruited from pulmonary and rheumatology clinics at a single VA site so some risk of spectrum bias.
Bansal et al, 2015[220]	Yes	Yes	Yes	Unclear	Women of this age group likely had some recognized risk of osteoporosis or fracture risk (a majority [69.7%] had a previous DXA), so potential for spectrum bias
Brenneman et al, 2003[222]	Yes	Yes	Unclear	Low	Patients recruited by mailing to random sample
Cadarette et al, 2001[206]	Yes	Yes	Yes	Low	Population-based sample
Cadarette et al, 2004[223]	Yes	Yes	Yes	Low	None
Cass et al, 2013[197]	Unclear	Yes	Yes	Unclear	Unclear whether a consecutive or random sample
Cass et al, 2006[224]	Unclear	Yes	Yes	Unclear	Women recruited from a single-site family practice, but no details regarding consecutive or random sample
Cass et al, 2016[232] Shepherd et al, 2010[202]	Yes	Yes	Yes	Low	NHANES sample
Chan et al, 2006[207]	Unclear	Yes	Unclear	Unclear	No information on participant inclusion/exclusion criteria
Chang et al, 2016[240]	Unclear	Yes	Unclear	Unclear	Retrospective identification of men from a large teaching hospital who had a DXA done but otherwise no selection criteria or method reported, so unclear if was consecutive or random
Chao et al, 2015[427]	Unclear	Yes	Unclear	Unclear	Single-site enrollment without reported exclusion criteria. Used a convenience sample from health education workshops; only included women with intermediate (FRAX; 10%-20% MOF, 1.5%-3% hip) or high-risk fracture risk (FRAX; ≥20% MOF, ≥3% hip) so potential for spectrum bias
Chen et al, 2016[233]	Unclear	Yes	Yes	Unclear	Participants were not clearly consecutively or randomly sampled
Christodoulou et al, 2016[242]	Unclear	Yes	Yes	Unclear	Unclear whether a random or consecutive sample
Cook et al, 2005[208]	Unclear	Yes	Unclear	Unclear	Sample had potential for bias toward low BMD due to recruitment from DXA clinic (all patients referred by doctor for clinical risk factors)

Appendix D Table 43. Risk of Bias of Included Studies for Key Question 2c (Domain 1 Patient Selection)

Author, Year	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Could the selection of patients have introduced bias?	Comments
Crandall et al, 2014[195] Crandall et al, 2019[141]	Yes	Yes	Yes	Unclear	For this study, they used information from a subset of the WHI participants from 3 of the 40 centers who participated in the DXA substudy
D'Amelio et al, 2013[199]	Yes	Yes	Yes	Low	None
D'Amelio et al, 2005[225]	Unclear	Yes	Yes	Unclear	Potential for spectrum bias, given the study population was referred specifically for DXA testing, in some cases for suspected secondary osteoporosis
Diem et al, 2017[235] Lynn et al, 2008[218]	Yes	Yes	Yes	Low	Only exclusions listed were hip replacement and inability to walk without a cane[218]
Geusens et al, 2002[231]	Yes	Yes	Yes	Low	None
Gourlay et al, 2005[209] Ben Sedrine et al, 2001[211] Richy et al, 2004[210]	Unclear	Yes	Yes	Unclear	Potential for spectrum bias, given the study population was referred or consulted spontaneously for DXA testing
Gourlay et al, 2008[230]	Yes	Yes	Yes	Low	None
Hamdy et al, 2018[238]	Unclear	Yes	Yes	Unclear	Consecutive sample of patients referred to an osteoporosis center so potential for spectrum bias
Harrison et al, 2006[217]	Unclear	Yes	Unclear	Unclear	No details on setting or how participants were selected
Inderjeeth et al, 2020[239]	Unclear	Yes	Yes	Unclear	Unclear whether a random or consecutive sample
Jiang et al, 2016[236]	Unclear	Yes	Yes	Unclear	Participants originally recruited for a study of pregnancy, breastfeeding, and osteoporosis
Jimenez-Nunez et al, 2013[203]	Yes	Yes	Yes	Low	None
Kirilova et al, 2019[428]	Unclear	Yes	Unclear	Unclear	No information on where patients were recruited from and very few characteristics of the population described
Kung et al, 2005[212]	No	Yes	Yes	Unclear	Convenience samples of men recruited from community health fairs or health talks
Kung et al, 2003[213]	Unclear	Yes	Yes	Unclear	Convenience sample of patients recruited from the community

Appendix D Table 43. Risk of Bias of Included Studies for Key Question 2c (Domain 1 Patient Selection)

Author, Year	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Could the selection of patients have introduced bias?	Comments
Leslie et al, 2013[200]	Yes	Yes	Yes	Low	None
Machado et al, 2010[201]	Yes	Yes	Yes	Low	None
Martinez-Aguila et al, 2007[214]	No	Yes	Unclear	Unclear	Patients were all referred for DXA, so potential for spectrum bias
Mauck et al, 2005[226]	Yes	Yes	Yes	Low	None
McLeod et al, 2015[205]	Yes	Yes	Yes	Low	None
Moon et al, 2016[244]	Yes	Yes	Yes	Low	KNHANES data are considered representative of the entire Korean population but only included sample of men with reported DXA results; also excluded those who may be at increased risk for osteoporosis but included asthma and all thyroid disease. Also excluded anyone with foreign bodies in bones (surgical pins/cement), unclear how many persons this was.
Morin et al, 2009[159]	Yes	Yes	Unclear	Unclear	Population was younger women ages 40–59 years who received a DXA; however, in this province, younger women are only eligible to have coverage for DXA testing if they have clinical risks for secondary osteoporosis, history of prior fracture, or X-ray evidence of osteopenia.
Nguyen et al, 2004[227]	Yes	Yes	Yes	Low	None
Oh et al, 2013[204]	Yes	Yes	Yes	Low	None
Oh et al, 2016[229]	Yes	Yes	Yes	Low	Population-based sample (KNHANES)
Pang et al, 2014[196]	No	Yes	Yes	Unclear	Not a consecutive or random sample
Park et al, 2003[215]	Unclear	Yes	Yes	Unclear	Unclear whether a consecutive or random sample
Pecina et al, 2016[234]	Yes	Yes	Unclear	Unclear	Participants were identified retrospectively from a panel of patients at a single academic healthcare center who had undergone DXA measurement. Participants taking bone active drugs were excluded.
Richards et al, 2014[198]	Unclear	Yes	Yes	Unclear	Unclear whether a consecutive or random sample was enrolled
Rud et al, 2005[228]	Yes	Yes	Yes	Low	None

Appendix D Table 43. Risk of Bias of Included Studies for Key Question 2c (Domain 1 Patient Selection)

Author, Year	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Could the selection of patients have introduced bias?	Comments
Shepherd et al, 2010[202]	Yes	Yes	Yes	Low	None
Shuler et al, 2016[241]	Unclear	Yes	Yes	Unclear	Patients selected from EMR based on residence, history of prior fracture, and risk factors for secondary osteoporosis
Sinnott et al, 2006[219]	Unclear	Yes	Yes	Unclear	Selection of participants may be a convenience sample but unclear. Men were recruited from general medicine clinics.
Toh et al, 2019[245]	Yes	Yes	Yes	Low	None
Wang et al, 2021[243]	No	Yes	Yes	Unclear	Data were obtained from a convenience cohort; all participants were referred for a medical reason; 10% of subjects actually were already on osteoporotic treatment; excluded non-Caucasians; however, this was only 117 patients out of over 36,000
Williams et al, 2017[237]	Yes	Yes	Unclear	Unclear	Patients were identified through their designation of belonging to a bone health team at a single VA facility. This suggests they were already identified as being at high risk for osteoporosis, which may lead to spectrum bias.
Zimering et al, 2007[216]	Unclear	Yes	Yes	Unclear	Convenience sample 30% came from specialty clinics (endocrionology or osteoporosis) for total cohort, but unknown for validation cohort. Excluded those unable to assess risk factors or DXA, though did not exclude based on known medical comorbidities or bone active medications.

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; EMR=electronic medical records; FRAX=Fracture Risk Assessment Tool; KNHANES=Korean National Health and Nutrition Examination Survey; MOF=major osteoporotic fracture; NHANES= National Health and Nutrition Examination Survey; VA=Veterans Affairs; WHI=Women's Health Initiative.

Appendix D Table 44. Risk of Bias of Included Studies for Key Question 2c (Domain 2 Index Test)

Author, Year	Were the index test results interpreted without knowledge of the results of the reference standard?	If a threshold was used, was it prespecified?	Could the conduct or interpretation of the index test have introduced bias?	Comments on Rating for Index Test
Adler et al, 2003[221]	Unclear	Yes	Unclear	No masking; used three cutoffs for OST: two based on published literature and one based on what they thought was appropriate.
Bansal et al, 2015[220]	Unclear	Yes	Unclear	Masking NR
Brenneman et al, 2003[222]	Unclear	Yes	Unclear	SCORE cutoff was recalibrated using study data to achieve sensitivity of approximately 90%. Developer cutoff ≥6; study cutoff ≥8; masking NR
Cadarette et al, 2001[206]	Unclear	Yes	Unclear	Masking NR
Cadarette et al, 2004[223]	Unclear	Yes	Unclear	Masking NR
Cass et al, 2013[197]	Yes	Yes	Low	None
Cass et al, 2006[224]	Unclear	Yes	Unclear	Masking NR
Cass et al, 2016[232] Shepherd et al, 2010[202]	Yes	Unclear	Unclear	Threshold was determined in a split sample using a development cohort.
Chan et al, 2006[207]	Unclear	Yes	Unclear	Masking NR; study only reported outcomes for the femoral neck at the prespecified thresholds. The lumbar spine outcomes were reported using empirically derived thresholds.
Chang et al, 2016[240]	Unclear	No	Unclear	Masking NR; threshold for OST not prespecified, used a threshold to optimize Sn and Sp
Chao et al, 2015[427]	Yes	Yes	Low	None
Chen et al, 2016[233]	Unclear	Yes	Unclear	Masking NR
Christodoulou et al, 2016[242]	Unclear	No	Unclear	Masking NR; does not appear to use prespecified thresholds
Cook et al, 2005[208]	Unclear	Yes	Unclear	Used a 90% sensitivity threshold, but also created a cutoff level based on the highest combined value of Sn and Sp
Crandall et al, 2014[195] Crandall et al, 2019[141]	Unclear	Yes	Unclear	Masking NR
D'Amelio et al, 2013[199]	Unclear	Yes	Unclear	Masking NR; study was prospective but not clear when the risk assessments were calculated (before or after BMD); the thresholds mentioned in study do not correspond entirely to thresholds used by other studies.
D'Amelio et al, 2005[225]	Unclear	Yes	Unclear	Masking NR
Diem et al, 2017[235] Lynn et al, 2008[218]	Unclear	Yes	Unclear	Masking NR; some thresholds were prespecified[235]
Geusens et al, 2002[231]	Unclear	Yes	Unclear	Masking NR

Appendix D Table 44. Risk of Bias of Included Studies for Key Question 2c (Domain 2 Index Test)

Author, Year	Were the index test results interpreted without knowledge of the results of the reference standard?	If a threshold was used, was it prespecified?	Could the conduct or interpretation of the index test have introduced bias?	Comments on Rating for Index Test
Gourlay et al, 2005[209] Ben Sedrine et al, 2001[211] Richy et al, 2004[210]	Unclear	No	Unclear	Did not use prespecified cutoffs for ORAI, OST, or SCORE. Instead, picked cutoff to achieve Sn 90% for each age group younger and older than 65 years; masking NR[210, 211]
Gourlay et al, 2008[230]	Unclear	Yes	Unclear	Masking NR
Hamdy et al, 2018[238]	Unclear	Yes	Unclear	Masking NR; although 2 of the 3 identified thresholds were prespecified, the study also examined the impact of different thresholds based on the ROC curve.
Harrison et al, 2006[217]	Unclear	Yes	Unclear	Masking NR
Inderjeeth et al, 2020[239]	Unclear	Yes	Unclear	Masking NR
Jiang et al, 2016[236]	Yes	Yes	Low	BMI cut point was not predetermined, but other index test thresholds were.
Jimenez-Nunez et al, 2013[203]	Yes	Yes	Low	None
Kirilova et al, 2019[428]	Unclear	Yes	Unclear	Masking NR
Kung et al, 2005[212]	Unclear	Yes	Unclear	Masking NR
Kung et al, 2003[213]	Unclear	Yes	Unclear	Masking NR
Leslie et al, 2013[200]	Unclear	Yes	Unclear	Masking NR
[218]				
Machado et al, 2010[201]	Unclear	Yes	Unclear	Masking NR
Martinez-Aguila et al, 2007[214]	No	Yes	Unclear	Clinical risk factors assessed retrospectively by asking participants to answer them based on the date of their BMD testing
Mauck et al, 2005[226]	Unclear	Yes	Unclear	Masking NR
McLeod et al, 2015[205]	Yes	Yes	Low	None
Moon et al, 2016[244]	Unclear	No	Unclear	Masking NR; thresholds were not prespecified
Morin et al, 2009[159]	Unclear	Yes	Unclear	Masking NR; Sn and Sp reported for multiple thresholds. The threshold of ≤1 is what has been used in other studies.
Nguyen et al, 2004[227]	Unclear	Yes	Unclear	Masking NR
Oh et al, 2013[204]	Unclear	No	Unclear	The authors did not report findings for the prespecified OSTA threshold. Instead, they reported findings for a different threshold that they selected to maximize discriminative ability.
Oh et al, 2016[229]	Unclear	Unclear	Unclear	Masking NR, unclear whether threshold used was prespecified
Pang et al, 2014[196]	Unclear	No	Unclear	Masking NR; thresholds were not prespecified; rather, they were chosen to maximize discriminative ability.

Appendix D Table 44. Risk of Bias of Included Studies for Key Question 2c (Domain 2 Index Test)

Author, Year	Were the index test results interpreted without knowledge of the results of the reference standard?	If a threshold was used, was it prespecified?	Could the conduct or interpretation of the index test have introduced bias?	Comments on Rating for Index Test
Park et al, 2003[215]	Unclear	Yes	Unclear	Masking NR
Pecina et al, 2016[234]	Unclear	Yes	Unclear	Masking of risk assessment calculations to BMD results NR
Richards et al, 2014[198]	Unclear	Yes	Unclear	Masking NR
Rud et al, 2005[228]	Unclear	Yes	Unclear	Masking NR
Shepherd et al, 2010[202]	Unclear	Yes	Unclear	Masking NR
Shuler et al, 2016[241]	Unclear	Yes	Unclear	Masking NR
Sinnott et al, 2006[219]	Unclear	Unclear	Unclear	Masking NR; unclear whether threshold prespecified
Toh et al, 2019[245]	Unclear	Yes	Unclear	Masking NR
Wang et al, 2021[243]	Unclear	No	Unclear	Masking NR; thresholds not prespecified
Williams et al, 2017[237]	Unclear	Unclear	Unclear	Only the data for FRAX were collected prior to DXA; the others were collected from the EMR. Thresholds were mostly based on sensitivity analyses, and only one appears to have been prespecified (hip fx 3%).
Zimering et al, 2007[216]	Unclear	Yes	Unclear	Masking NR; threshold determined in the development cohort

Abbreviations: BMD=bone mineral density; BMI=body mass index; DXA=dual-energy X-ray absorptiometry; EMR=electronic medical records; FRAX=Fracture Risk Assessment Tool; fx=fracture; NR=not reported; ORAI= Osteoporosis Risk Assessment Instrument; OST=Osteoporosis Self-Assessment Tool; OSTA=Osteoporosis Self-Assessment Tool for Asians; ROC=receive operating characteristics curve; SCORE=Simple Calculated Osteoporosis Risk Estimation; Sn=sensitivity; Sp=specificity; USPSTF=U.S. Preventive Services Task Force.

Appendix D Table 45. Risk of Bias of Included Studies for Key Question 2c (Domain 3 Reference Standard)

Author, Year	Is the reference standard likely to correctly classify the target condition?	Were the reference standard results interpreted without knowledge of the results of the index test?	Could the reference standard, its conduct, or interpretation have introduced bias?	Comments on Rating for Reference Standard
Adler et al, 2003[221]	Yes	Unclear	Unclear	Masking NR
Bansal et al, 2015[220]	Yes	Unclear	Unclear	Masking NR
Brenneman et al, 2003[222]	Yes	Unclear	Unclear	Used NHANES III reference values, but age and sex of reference values used NR, masking NR
Cadarette et al, 2001[206]	Yes	Unclear	Unclear	Masking NR; used young healthy Canadian adult references ranges, which authors stated are similar to NHANES
Cadarette et al, 2004[223]	Yes	Unclear	Unclear	Masking NR; reference ranges used NR
Cass et al, 2013[197]	Yes	Yes	Low	None
Cass et al, 2006[224]	Unclear	Unclear	Unclear	Used manufacturer's reference ranges; masking NR
Cass et al, 2016[232] Shepherd et al, 2010[202]	Yes	Yes	Low	None
Chan et al, 2006[207]	Unclear	Unclear	Unclear	Masking NR; no information on the specific reference ranges used to determine T-score
Chang et al, 2016[240]	Unclear	Unclear	Unclear	Did not describe the reference ranges used to calculate T-score; given that this was a male sample, it is possible that male reference ranges were used instead of the range ISCD-recommended range (young, healthy female)
Chao et al, 2015[427]	Yes	Unclear	Unclear	Masking NR
Chen et al, 2016[233]	Unclear	Unclear	Unclear	Reference ranges used for T-scores NR, masking NR
Christodoulou et al, 2016[242]	Unclear	Unclear	Unclear	Masking NR; did not report reference values used for calculating T-scores
Cook et al, 2005[208]	Unclear	Unclear	Unclear	T-scores were computed using the databases supplied with the systems; masking NR
Crandall et al, 2014[195] Crandall et al, 2019[141]	Yes	Unclear	Unclear	Masking NR; DXA was conducted at time of enrollment; reference ranges used for T-scores reported in one of the articles as NHANES III normative reference database[195] and was NR in the other article.[141]
D'Amelio et al, 2013[199]	Unclear	Unclear	Unclear	Reference range for T-score NR; masking NR
D'Amelio et al, 2005[225]	Unclear	Unclear	Unclear	Masking NR; reference ranges used NR
Diem et al, 2017[235] Lynn et al, 2008[218]	Ye[235]s/Unclear[218]	Unclear	Unclear	Masking NR; Used male reference ranges to calculate T-score for U.S. participants, local Chinese reference ranges used for Hong Kong participants.[218]
Geusens et al, 2002[231]	Yes	Unclear	Unclear	Masking NR

Appendix D Table 45. Risk of Bias of Included Studies for Key Question 2c (Domain 3 Reference Standard)

Author, Year	Is the reference standard likely to correctly classify the target condition?	Were the reference standard results interpreted without knowledge of the results of the index test?	Could the reference standard, its conduct, or interpretation have introduced bias?	Comments on Rating for Reference Standard
Gourlay et al, 2005[209] Ben Sedrine et al, 2001[211] Richy et al, 2004[210]	Unclear	Unclear	Unclear	Used local reference values to calculate BMD[211] and NR[210]; masking NR[210, 211]
Gourlay et al, 2008[230]	Yes	Unclear	Unclear	Masking NR
Hamdy et al, 2018[238]	Yes	Unclear	Unclear	Masking NR
Harrison et al, 2006[217]	Yes	Unclear	Unclear	Masking NR; reference ranges used NR
Inderjeeth et al, 2020[239]	Unclear	Unclear	Unclear	Did not report reference values used for BMD calculations; unclear if reference standard assessors were blinded to index test results
Jiang et al, 2016[236]	Unclear	Yes	Unclear	Reference ranges used for T-scores NR
Jimenez-Nunez et al, 2013[203]	Unclear	Yes	Unclear	Manufacturer's reference ranges for the Spanish population for young Caucasian adults were used to calculate T-scores; masking NR
Kirilova et al, 2019[428]	Unclear	Unclear	Unclear	Masking NR; did not directly report what reference values were used to calculate T-score, but the description implies used a young healthy female reference range
Kung et al, 2005[212]	Unclear	Unclear	Unclear	Young healthy males recruited from the same community were the reference values used to compute T-scores; masking NR
Kung et al, 2003[213]	Unclear	Unclear	Unclear	Used reference range values from young healthy Chinese; masking NR
Leslie et al, 2013[200]	Yes	Unclear	Unclear	Masking NR
Lynn et al, 2008[218]	Unclear	Unclear	Unclear	Used male reference ranges to calculate T-score for U.S. participants, local Chinese reference ranges used for Hong Kong participants. Masking NR.
Machado et al, 2010[201]	Unclear	Unclear	Unclear	Masking NR: used NHANES reference ranges for hip but unclear whether used female or male ranges
Martinez-Aguila et al, 2007[214]	Unclear	Unclear	Unclear	Used a young healthy population for reference range but specific population used NR; masking NR
Mauck et al, 2005[226]	Unclear	Unclear	Unclear	Used a local reference range for T-score values; masking NR

Appendix D Table 45. Risk of Bias of Included Studies for Key Question 2c (Domain 3 Reference Standard)

Author, Year	Is the reference standard likely to correctly classify the target condition?	Were the reference standard results interpreted without knowledge of the results of the index test?	Could the reference standard, its conduct, or interpretation have introduced bias?	Comments on Rating for Reference Standard
McLeod et al, 2015[205]	Yes	Yes	Low	None
Moon et al, 2016[244]	Unclear	Unclear	Unclear	Reference ranges used to calculate T-scores NR; masking NR
Morin et al, 2009[159]	Yes	Unclear	Unclear	Masking NR
Nguyen et al, 2004[227]	Unclear	Unclear	Unclear	Local reference range for young Australian women used; masking NR
Oh et al, 2013[204]	Unclear	Unclear	Unclear	Masking NR; used reference valued from young Japanese women
Oh et al, 2016[229]	Unclear	Unclear	Unclear	Used gender-specific normal values for young Japanese men; masking NR
Pang et al, 2014[196]	Unclear	Unclear	Unclear	Reference range used to calculate T-score NR; masking NR
Park et al, 2003[215]	Unclear	Unclear	Unclear	Used reference ranges for young healthy Korean women; masking NR
Pecina et al, 2016[234]	Unclear	Unclear	Unclear	Reference ranges to calculate T-scores NR
Richards et al, 2014[198]	Unclear	Unclear	Unclear	Masking NR, used race-specific male reference ranges
Rud et al, 2005[228]	Yes	Unclear	Unclear	Masking NR
Shepherd et al, 2010[202]	Unclear	Unclear	Unclear	Masking NR, used young male reference range but unclear whether this was NHANES data or other data
Shuler et al, 2016[241]	Unclear	Unclear	Unclear	Masking NR; reference ranges used to calculate T-scores NR
Sinnott et al, 2006[219]	Unclear	Unclear	Unclear	Used young Caucasian male reference values; masking NR
Toh et al, 2019[245]	Unclear	Unclear	Unclear	Reference ranges used to calculate T-scores NR
Wang et al, 2021[243]	Yes	Yes	Low	Risk assessment only performed for this study and presumably long after T-scores calculated
Williams et al, 2017[237]	Unclear	Unclear	Unclear	No masking, did not describe the reference ranges used to calculate T-score; given that this was a male sample, it is possible that male reference ranges were used instead of the ISCD-recommended range (young, healthy female)
Zimering et al, 2007[216]	Yes	Unclear	Unclear	Masking NR; reference ranges used NR

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; ISCD=International Society for Clinical Densitometry; NHANES=National Health and Nutrition Examination Survey; NR=not reported.

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Adler et al, 2003[221]	NR	1 month	Yes	Yes	Yes	Yes	Low	None
Bansal et al, 2015[220]	NR	FRAX input collected at time of DXA or from review of medical records	Yes	Yes	Yes	Yes	Low	None
Brenneman et al, 2003[222]	1,986 recruited 428 consented 416 had complete data	Occurred concurrently	Yes	Yes	Yes	Yes	Low	None
Cadarette et al, 2001[206]	69 patients missing data to calculate clinical decision rules, 382 patients had an osteoporosis diagnosis, 20 patients were taking bone sparing medications, 158 had potential causes for secondary osteoporosis, and 294 were using HRT for less than 5 years. Altogether, 923 patients were excluded.	Not specifically reported. All baseline data collected 2/2016–9/2017, presumably includes questionnaire and DXA testing	Unclear	Yes	Yes	No	Unclear	Interval between questionnaires and DXA testing NR
Cadarette et al, 2004[223]	Only patients with DXA included	NR	Unclear	Yes	Yes	No	Unclear	Timing of risk assessment and DXA NR in the prospective or retrospective sample. Persons with missing data were excluded from the retrospective sample.

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Cass et al, 2013[197]	40 men excluded because they did not complete DXA	NR	Unclear	Yes	Yes	No	Unclear	Interval between risk assessment and DXA NR; 40 men excluded who did not complete DXA
Cass et al, 2006[224]	562 approached: 226 enrolled, 173 declined, 163 not eligible		Unclear	Yes	Yes	No	Unclear	23 enrolled patients did not undergo DXA scan so were not included
Cass et al, 2016[232] Shepherd et al, 2010[202]	NR	NR	Unclear	Yes	Yes	Yes	Low	Timing of assessments NR but likely close to concurrent based on NHANES methodology
Chan et al, 2006[207]	NR	NR	Unclear	Yes	Yes	Unclear	Unclear	Neither the number eligible nor the number of dropouts is reported. Only the final N analyzed is reported.
Chang et al, 2016[240]	821 analyzed, but unclear how many were eligible but excluded for not being a "valid" sample	NR	Unclear	Yes	Yes	Unclear	Unclear	Retrospective analysis, unclear who was eligible but excluded, no interval reported
Chao et al, 2015[427]	Women deemed to be a low risk based on fracture risk scores were not included in the analysis	NR	Unclear	No	Yes	No	High	Excluded women at low risk for fracture based on the index test. High potential for spectrum bias.
Chen et al, 2016[233]	All patients received all tests	Baseline assessments of physical measurements and personal interviews appear to have been conducted at the same time	Yes	Yes	Yes	Yes	Low	None

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Christodoulou et al, 2016[242]	1,000 patients included, all received reference and index test; did not include any patients who may have been approached but did not complete screens	NR	Unclear	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA NR; retrospectively conducted so unclear whether patients excluded because of missing data
Cook et al, 2005[208]	None	NR	Unclear	Yes	Yes	Yes	Low	None
Crandall et al, 2014[195] Crandall et al, 2019[141]	Only participants from the BMD substudy between ages 50 and 64 years were included in these analyses.	Risk assessment data collected at baseline and DXA was conducted at time of enrollment but specific interval between them was NR	Probably Yes	Yes	Yes	Yes	Low	None
Crandall et al, 2014[195]	NA	NR	Yes	Yes	Yes	Yes	Low	Main analysis was restricted to a subgroup of non-HRT users by design (supplemental analyses include HRT users and all women [including those with preventive use of HRT])

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
D'Amelio et al, 2013[199]	NR	NR	Unclear	Yes	Yes	No	Low	Some patients initially enrolled were excluded because it was determined they did not meet study criteria (3.4%); interval between risk assessment and BMD NR but presumably concurrent because study was prospective
D'Amelio et al, 2005[225]	NR	Clinical risk factors collected at the time of DXA scan	Yes	Yes	Yes	Yes	Low	None
Diem et al, 2017[235] Lynn et al, 2008[218]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between risk assessment and DXA NR; unclear whether participants were excluded from analysis[218]
Geusens et al, 2002[231]	NA	NR	Unclear	Yes	Yes	Yes	Unclear	Unclear because of lack of clarity on timing of the tests
Gourlay et al, 2005[209] Ben Sedrine et al, 2001[211] Richy et al, 2004[210]	Retrospectively conducted study, only participants with available data were included	NR	Unclear	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA NR; unclear how many eligible participants had missing data
Gourlay et al, 2008[230]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between collection of risk factors and DXA NR

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Hamdy et al, 2018[238]	726 men included in this retrospective analysis; no information on the number of men who were eligible but excluded for missing FRAX data or uninterpretable DXA scans	NR	Unclear	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA NR; no information on the number of eligible men excluded for missing data
Harrison et al, 2006[217]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Unclear whether any missing data
Inderjeeth et al, 2020[239]	The reported number of participants is unclear. 531 participants were included in the final analysis yet elsewhere the manuscript noted that the study sample included n=534	NR	Unclear	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA testing NR
Jiang et al, 2016[236]	Three of the 445 women surveyed failed to provide the researcher with their age and were consequently eliminated from the study	NR	Unclear	Yes	Yes	No	Unclear	Interval between risk assessment and DXA NR
Jimenez-Nunez et al, 2013[203]	NR	Same day	Yes	Yes	Yes	Yes	Low	None
Kirilova et al, 2019[428]	180 analyzed, but no mention of how many were eligible	NR	Unclear	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA NR

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Kung et al, 2005[212]	Excluded those with history or evidence of metabolic bone disease, history of cancer, evidence of significant renal impairment, both hips previously fractured or replaced, prior use of any bisphosphonates, fluoride or calcitonin, abnormal thyroid stimulating hormone	NR	Unclear	Yes	Yes	Yes	Unclear	Time frame between clinical assessment of risk factors and DXA unclear
Kung et al, 2003[213]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between risk assessment and DXA testing NR
Leslie et al, 2013[200]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between risk assessment and DXA NR
Machado et al, 2010[201]	73% of enrolled participants were excluded because of incomplete data or missing technical data for the DXA	NR	Unclear	Yes	Yes	No	Unclear	Interval between risk assessment and DXA NR; of 473 men, 202 were included that were age 50 years or older and had DXA data
Martinez-Aguila et al, 2007[214]	NR	NR	Unclear	Yes	Yes	No	Unclear	30 eligible patients were excluded for missing data
Mauck et al, 2005[226]	NR	NR	Yes	Yes	Yes	Yes	Low	None
McLeod et al, 2015[205]	3 patients were excluded because of previous diagnosis or progressive terminal illness	Within 3 weeks	Yes	Yes	Yes	Yes	Low	None

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Moon et al, 2016[244]	Does not describe how many persons excluded, but 2,519 were included with both index/reference test	NR	Unclear	Yes	Yes	Yes	Unclear	Unclear whether any missing data
Morin et al, 2009[159]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Unclear for timing between DXA and index test
Nguyen et al, 2004[227]	NR	Not explicit, but given study design presume it was concurrent	Yes	Yes	Yes	Yes	Low	None
Oh et al, 2013[204]	708 participants were excluded because of at least one of the following reasons: the absence of BMD measurement (n=149), a previous osteoporosis diagnosis or osteoporosis treatment (n=473), missing blood tests (n=199), or being in a bedridden state (n=36)	NR but because prospective, likely collected concurrently or in close proximity	Yes	Yes	Yes	Yes	Low	None

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Oh et al, 2016[229]	Excluded 252, at least one of the following reasons: absence of BMD measurement (n=149), previously diagnosed osteoporosis or treatment for osteoporosis (n=34), missing blood tests (n=144), and being in a bedridden state (n=14)	NR	Unclear	Yes	Yes	No	Unclear	Excluded some men for probably valid reasons, interval between risk assessment and DXA NR
Pang et al, 2014[196]	Unclear	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between risk assessment and DXA testing NR. Unclear whether all participants were included in the analysis.
Park et al, 2003[215]	NR	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between risk assessment and DXA NR
Pecina et al, 2016[234]	Retrospectively conducted study. Does not appear that any participants were excluded.	NR	Unclear	Yes	Yes	Yes	Unclear	Interval between index test and BMD testing not reported
Richards et al, 2014[198]	2 men excluded for not having a DXA test.	NR	Unclear	Yes	Yes	Yes	Unclear	Unclear because of lack of clarity on timing of the tests. Two patients were excluded from the analysis because no BMD tests were done.
Rud et al, 2005[228]	Data were available for 1,997 of the 2,009 participants	NR	Unclear	Yes	Yes	Yes	Unclear	Timing of risk assessment and DXA NR

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Shepherd et al, 2010[202]	Men who self-reported history of radiographic contrast material (barium) use during the past 7 days, nuclear medicine studies during the previous 3 days, a weight more than 300 pounds, or a height more than 6' 5" were excluded. 40 men (35 non-Hispanic White men and 5 men of unspecified race/ethnicity) were dropped from analysis because of missing values for essential variables.	NR	Yes	Yes	Yes	Unclear	Unclear	Excluded men without DXA available, though not specifically reported. NHANES enrolls subjects prospectively, so clinical risks and DXA likely collected concurrently.
Shuler et al, 2016[241]	55 patients completed FRAX; 45 patients completed DXA testing	NR	Unclear	No	Yes	No	Unclear	Interval between risk assessment and DXA NR; 10 patients excluded for not having DXA data
Sinnott et al, 2006[219]	NR	NR	Unclear	Yes	Yes	Yes	Low	None
Toh et al, 2019[245]	224 patients approached, 164 consented and received index test, 150/164 received reference test and were included for analysis	NR	Unclear	Yes	Yes	No	Unclear	Interval between risk assessment and DXA NR: small proportion did not get included because of missing data for DXA

Appendix D Table 46. Risk of Bias of Included Studies for Key Question 2c (Domain 4 Flow and Timing)

Author, Year	Describe the number of patients who did not receive the index test(s) and/or reference standard or who were excluded from the analysis?	Describe the time interval and any interventions between index test(s) and reference standard.	Was there an appropriate interval between index test(s) and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?	Could the patient flow have introduced bias?	Comments on Flow or Timing
Wang et al, 2021[243]	A total of 18,670 of the original 36,590 patients were interpreted. 117 patients were excluded for being non-Caucasian, 1,935 were excluded for age <40 years, and 15,868 were excluded for incomplete reference tests.	NR	No	Yes	Yes	Unclear	Unclear	Interval between risk assessment and DXA NR; many persons excluded for missing data at one or more sites on DXA but unclear whether this is missing at random or related to outcome
Williams et al, 2017[237]	965 enrolled in bone health team, 463 analyzed; the rest were either missing a DXA result or did not have weight documented	NR	Unclear	Yes	Yes	No	Unclear	The analysis was limited to participants with DXA and weight and who were male, all others excluded. No interval reported.
Zimering et al, 2007[216]	NR	Not reported, presumably concurrent testing	Unclear	Yes	Yes	No	Unclear	The flow was not specifically described, but appears sequence was clinical assessment followed by ultrasound and then DXA

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; FRAX=Fracture Risk Assessment Tool; HRT=hormone replacement therapy; N=number; NA=not applicable; NHANES= National Health and Nutrition Examination Survey; NR=not reported.

Appendix D Table 47. Risk of Bias of Included Studies for Key Question 2c (Overall Study Quality)

Author, Year	Overall Study Quality	Rationale for Overall Rating
Adler et al, 2003[221]	Some ROB/Fair quality	Unclear for domain of patient selection, no masking of index and reference test results
Bansal et al, 2015[220]	Some ROB/Fair quality	Potential for spectrum bias because younger women with DXA likely have had some unspecified risk factors. Some risk of bias introduced by retrospective design becuause women ages 50–64 years would typically not have DXA ordered in the absence of increased risks for osteoporosis.
Brenneman et al, 2003[222]	Some ROB/Fair quality	No masking of index and reference test results
Cadarette et al, 2001[206]	Some ROB/Fair quality	Masking NR; interval between questionnaire and DXA NR
Cadarette et al, 2004[223]	Some ROB/Fair quality	Masking NR, unclear timing of risk assessment and DXA; some participants with missing data were excluded from the retrospective sample.
Cass et al, 2013[197]	Some ROB/Fair quality	Unclear whether consecutive or random sample, interval between risk assessment and DXA NR; 40 men (10%) were excluded because they did not complete DXA test.
Cass et al, 2006[224]	Some ROB/Fair quality	Unclear whether sample was random or consecutive, masking of results not reported, some enrolled participants did not get DXA, timing of risk assessment and DXA NR, did not use NHANES young healthy reference ranges for T-scores
Cass et al, 2016[232] Shepherd et al, 2010[202]	Some ROB/Fair quality	Threshold determined in a split sample.
Chan et al, 2006[207]	Some ROB/Fair quality	Not clear whether a random or consecutive sample, masking NR, reference range values NR, unclear interval between risk assessment and DXA
Chang et al, 2016[240]	Some ROB/Fair quality	Unclear ROB for all domains mainly because of lack of detailed reporting
Chao et al, 2015[427]	High ROB/Poor quality	Subjects who were low risk on risk assessment were excluded from analysis leading to high risk for spectrum bias. The participants were from a single site with unclear exclusion criteria. The index test was done without knowledge of the reference test, but it is unclear if the reference test was interpreted without the index test results. The interval between risk assessment and DXA was NR.
Chen et al, 2016[233]	Some ROB/Fair quality	The method of patient selection is unclear, the blinding of index text and reference standard interpretation were unclear, reference ranges used for T-scores NR.
Christodoulou et al, 2016[242]	Some ROB/Fair quality	Unclear whether a random or consecutive sample, masking NR, did not use prespecified thresholds, interval between risk assessment and DXA NR, unclear whether patients excluded for missing data
Cook et al, 2005[208]	Some ROB/Fair quality	Patient selection has the potential to skew the sample toward low BMD, did not use a standard reference range for calculating T-scores
Crandall et al, 2014[195] Crandall et al, 2019[141]	Some ROB/Fair quality	Retrospectively assembled dataset based on participants from 3 of the 40 centers who participated in the DXA substudy; masking of index and reference test NR.
D'Amelio et al, 2013[199]	Some ROB/Fair quality	Masking NR, BMD reference range used NR, interval between risk assessment and BMD NR
D'Amelio et al, 2005[225]	Some ROB/Fair quality	Referral population so potential for spectrum bias, masking NR, unclear what reference ranges for T-scores were used
Diem et al, 2017[235] Lynn et al, 2008[218]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA testing NR; data were collected prospectively from MrOS study and then analyzed as part of these analyses, one analysis used reference ranges other than NHANES young healthy female range
Geusens et al, 2002[231]	Some ROB/Fair quality	Masking NR, unclear interval between risk assessment and DXA
Gourlay et al, 2005[209] Ben Sedrine et al, 2001[211] Richy et al, 2004[210]	Some ROB/Fair quality	Risk of spectrum bias due to referral population; index test thresholds not prespecified, interval between DXA and risk assessment NR; masking NR; no mention of who was excluded or if any dropped out

Appendix D Table 47. Risk of Bias of Included Studies for Key Question 2c (Overall Study Quality)

Author, Year	Overall Study Quality	Rationale for Overall Rating
Gourlay et al, 2008[230]	Some ROB/Fair quality	Masking NR, no information about interval between risk factor collection and DXA
Hamdy et al, 2018[238]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA testing NR, unclear whether all eligible subjects were included
Harrison et al, 2006[217]	Some ROB/Fair quality	Very little information about patient selection, no mention of results of DXA being blinded during calculation of risk assessment indices
Inderjeeth et al, 2020[239]	Some ROB/Fair quality	Unclear whether a random or consecutive sample was used, masking NR, interval between risk assessment and index test NR
Jiang et al, 2016[236]	Some ROB/Fair quality	Patient selection was unclear, reference ranges used for T-scores NR, a few missing participants and uncertain interval between the index and reference tests
Jimenez-Nunez et al, 2013[203]	Some ROB/Fair quality	Did not use NHANES reference ranges to calculate T-scores
Kirilova et al, 2019[428]	High ROB/Poor quality	Unclear risk of bias in all domains evaluated, no information on where or how patients were selected and no inclusion/exclusion criteria mentioned, no information on timing of index test with respect to BMD or how DXA was conducted or reference ranges used, no discussion of masking of index test and reference test results, unclear whether any eligible patients were excluded for missing data
Kung et al, 2005[212]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA NR, convenience sampling
Kung et al, 2003[213]	Some ROB/Fair quality	Sample not consecutive or random, masking NR, interval between risk assessment and DXA testing NR
Leslie et al, 2013[200]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA NR
Lynn et al, 2008[218]	Some ROB/Fair quality	Data were collected prospectively from MrOS study and then analyzed as part of this study focus. Unclear ROB from unclear masking of index and reference standard results, and use of reference ranges other than NHANES young healthy female range and timing of index test with respect to reference test
Machado et al, 2010[201]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA NR, unknown amount of missing data for men
Martinez-Aguila et al, 2007[214]	Some ROB/Fair quality	Not a random or consecutive population, index text data collected retrospectively after reference test results known, some patients excluded for missing data
Mauck et al, 2005[226]	Some ROB/Fair quality	Masking NR
McLeod et al, 2015[205]	Low ROB/Good quality	None
Moon et al, 2016[244]	Some ROB/Fair quality	Masking NR, thresholds not prespecified, interval between risk assessment and DXA NR, unclear whether persons eligible were excluded for missing data
Morin et al, 2009[159]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA NR
Nguyen et al, 2004[227]	Some ROB/Fair quality	Masking NR
Oh et al, 2013[204]	Some ROB/Fair quality	Masking NR, threshold did not appear to be prespecified
Oh et al, 2016[229]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA testing NR, did not use NHANES reference range
Pang et al, 2014[196]	Some ROB/Fair quality	Not a consecutive or random sample, masking NR, reference range values to calculate T-scores NR, interval between risk assessment and DXA testing NR
Park et al, 2003[215]	Some ROB/Fair quality	Unclear whether sample was random/consecutive; masking NR; interval between risk assessment and DXA NR

Appendix D Table 47. Risk of Bias of Included Studies for Key Question 2c (Overall Study Quality)

Author, Year	Overall Study Quality	Rationale for Overall Rating
Pecina et al, 2016[234]	Some ROB/Fair quality	Unclear risk of bias because of masking of results of index and reference tests NR, unclear interval between tests, and reference range for T-scores NR. Participants taking bone active drugs were excluded.
Richards et al, 2014[198]	Some ROB/Fair quality	Masking NR, did not use NHANES White young female reference range, no information on time interval between risk assessment and DXA
Rud et al, 2005[228]	Some ROB/Fair quality	Masking NR, timing of risk assessment and DXA not reported
Shepherd et al, 2010[202]	Some ROB/Fair quality	Masking NR, used male reference range for T-scores
Shuler et al, 2016[241]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA testing NR; some missing data
Sinnott et al, 2006[219]	Some ROB/Fair quality	Primarily due to 1) no information on the type of sampling; assumed convenience sampling; 2) not clear about the sequence of testing; and 3) results from index and reference standard not masked
Toh et al, 2019[245]	Some ROB/Fair quality	Masking NR, interval between risk assessment and DXA NR, references ranges used for T-scores NR
Wang et al, 2021[243]	Some ROB/Fair quality	Convenience sample, masking of index text NR, interval between risk assessment and DXA NR, index text thresholds not prespecified, many patients excluded for missing data
Williams et al, 2017[237]	Some ROB/Fair quality	Potential for spectrum bias because all patients were referred for DXA, masking NR, interval between risk assessment and DXA NR, participants excluded for missing DXA or weight data, reference ranges used for T-score calculations NR
Zimering et al, 2007[216]	Some ROB/Fair quality	Convenience sample, masking of index text and reference test NR, unclear timing between index test and reference test, unclear patient flow and timing

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; NHANES=National Health and Nutrition Examination Survey; NR=not reported; ROB=risk of bias.

Appendix D Table 48. Risk of Bias for Studies for Key Question 2d (Domain 1 Participants)

Author, Year	1.1 Were appropriate data sources used?	1.2 Were all inclusions and exclusions of participants appropriate?	Domain 1 ROB	Domain 1 ROB Rationale	Domain 1 Applicability: Concern that the included participants and setting do not match the review question?	Domain 1 Applicability Rationale
Berry et al, 2013[246] Framingham Osteoporosis Study	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Crandall et al, 2020 [249] WHI	Yes/Probably yes	Yes/Probably yes	Low	None	Some	Participants were enrolled in a clinical trial
Ensrud et al, 2022[250] Mr.Os	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Hillier et al, 2007[247] SOF	Yes/Probably yes	Yes/Probably yes	Low	None	Low	None
Leslie et al, 2017[248] Manitoba BMD Registry	No/Probably no	No	High	Retrospective registry of DXA results; only persons with at least 2 DXA measurements included; DXA measurements based on referral from usual care thus potential for selection bias	Low	None

Abbreviations: DXA=dual-energy X-ray absorptiometry; Mr.Os=Osteoporotic Fractures in Men (study); ROB=risk of bias; SOF=Study of Osteoporotic Fractures; WHI=Women's Health Initiative.

Appendix D Table 49. Risk of Bias for Included Studies for Key Question 2d (Domain 2 Predictors)

Author, Year	2.1 Were predictors defined and assessed in a similar way for all participants?	2.2 Were predictor assessments made without knowledge of outcome data?	2.3 Are all predictors available at the time the model is intended to be used?	Domain 2 ROB	Domain 2 ROB Rationale	Domain 2 Applicability: Concern that the definition, assessment or timing of predictors in the model do not match the review question?	Domain 2 Applicability Rationale
Berry et al, 2013[246] Framingham Osteoporosis Study	Yes/Probably yes	Yes/Probably yes	Yes/ Probably yes	Low	None	Low	None
Crandall et al, 2020 [249] WHI	Yes/Probably yes	Yes/Probably yes	Yes/ Probably yes	Low	None	Low	None
Ensrud et al, 2022[250] Mr.Os.	Yes/Probably yes	Yes/Probably yes	Yes/ Probably yes	Low	None	Low	None
Hillier et al, 2007[247] SOF	Yes/Probably yes	Yes/Probably yes	Yes/ Probably yes	Low	None	Low	None
Leslie et al, 2017[248] Manitoba BMD Registry	Yes/Probably yes	Yes/Probably yes	Yes/ Probably yes	Low	None	Low	None

Abbreviations: Mr.Os=Osteoporotic Fractures in Men (study); ROB=risk of bias; SOF=Study of Osteoporotic Fractures; WHI=Women's Health Initiative.

Appendix D Table 50. Risk of Bias for Studies for Key Question 2d (Domain 3 Outcome)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determin-ation appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applic-ability Rationale
Berry et al, 2013[246] Framingham Osteoporosis Study	Yes/ Probably yes for hip, but not for other fracture types	Yes/ Probably yes	Yes/ Probably yes	Yes/ Probably yes	No information	Yes/ Probably yes	Unclear	Blinding NR; self-reported non-hip fractures not confirmed	Low	None
Crandall et al, 2020 [249] WHI	Yes/ Probably yes for hip; No for other fractures	Yes/ Probably yes for hip	Yes/ Probably yes	Yes/ Probably yes	No information	Yes/ Probably yes	Unclear	Blinding NR; fracture self-reported; only hip verified with medical records	Low	None
Ensrud et al, 2022[250] Mr.Os	Yes/ Probably yes; traumatic fractures may have been included	Yes/ Probably yes	Yes/ Probably yes	Yes/ Probably yes	No information	Yes/ Probably yes	Unclear	Blinding, NR; included traumatic fractures based on other MrOs cohort analyses	Low	None
Hillier et al, 2007[247] SOF	Yes/Probably yes (for clinical fractures)	Yes/ Probably yes	Yes/ Probably yes	Yes/ Probably yes	No information	Yes/ Probably yes	Unclear	Blinding NR, includes radio-graphic vertebral fractures	Low	None

Appendix D Table 50. Risk of Bias for Studies for Key Question 2d (Domain 3 Outcome)

Author, Year Fracture Type	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determin-ation appropriate?	Domain 3 ROB	Domain 3 ROB Rationale	Domain 3 Applicability: Concern that the outcome, its definition, timing, or determination do not match the review question?	Domain 3 Applic-ability Rationale
Leslie et al, 2017[248] Manitoba BMD Registry	Yes/Probably yes	Yes/ Probably yes	Yes/ Probably yes	Yes/ Probably yes	No information	Yes/ Probably yes	Unclear	Blinding was NR; fracture ascertain-ment based on administra-tive data	Low	None

Abbreviations: NR=not reported; Mr.Os=Osteoporotic Fractures in Men (study); ROB=risk of bias; SOF=Study of Osteoporotic Fractures; WHI=Women's Health Initiative.

Appendix D Table 51. Risk of Bias for Studies for Key Question 2d (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Berry et al, 2013[246] Framingham Osteoporosis Study	Yes/ Probably yes for MOF; No for hip	Yes/ Probably yes	No/Probably no	No information	Yes/Probably yes	No/Probably no	Unclear	Did not have enough fracture events for hip; excluded persons with incident fracture between BMD measurements; no information about missing data
Crandall et al, 2020 [249]	Yes/ Probably yes	Yes/ Probably yes	No/Probably no	No/Probably no	Yes/Probably yes	Yes/Probably yes	Unclear	Exlcuded persons with MOF between BMD measurements and those missing covariate data on risk assessment tools
Ensrud et al, 2022[250] Mr.Os	Yes/ Probably yes	Yes/ Probably yes	No/Probably no	No information	Yes/Probably yes	Yes/Probably yes	Unclear	Excluded persons with missing BMD measurement at year 7; no information about missing data
Hillier et al, 2007[247] SOF	Yes/ Probably yes	Yes/ Probably yes	No/Probably no	No information	No/Probably no	No/Probably no	High	Used logistic regression and did not account for complexities in the data; excluded persons with incident fracture between BMD measurements; no infromation about missing data

Appendix D Table 51. Risk of Bias for Studies for Key Question 2d (Domain 4 Analysis)

Author, Year	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	Domain 4 ROB	Domain 4 ROB Rationale
Leslie et al, 2017[248] Manitoba BMD Registry	Yes/ Probably yes	Yes/Probably yes	No/Probably no	No information	Yes/Probably yes	No/Probably No	Unclear	Number of participants from other studies reporting using this registry have a much higher number of participants suggesting that not all were included; unclear

Abbreviations: MrOs=Osteoporotic Fractures in Men (study); ROB=risk of bias; ; SOF=Study of Osteoporotic Fractures; WHI=Women's Health Initiative.

Appendix D Table 52. Risk of Bias for Studies for Key Question 2d (Domain 5 Overall ROB)

Author, Year	Overall Study Quality	Overall Rationale	Overall Applicability Assessment	Overall Applicability Rationale
Berry et al, 2013[246] Framingham Osteoporosis Study	Fair	Some risk of bias because no infromation about missing data; unclear whether outcome ascertainment was blinded; borderline number of fracture events	Low concerns	None
Crandall et al, 2020 [249]	Fair	FRAX instrument itself was rated as high ROB in the development cohort and similarly the external validation in the WHI cohort was also rated as high ROB. Fractures other than hip were self-reported, participants with missing covariate data excluded. Unclear whether outcome ascertainment was blinded to BMD measures.	Some concerns	Participants were enrolled in a clinical trial
Ensrud et al, 2022[250]	Fair	May have included traumatic fractures; persons excluded for missing covariate information, excluded participants with no repeat BMD at year 7; unclear whether outcome ascertainment was blinded to BMD measures.	Low concerns	None
Hillier et al, 2007[247] SOF	Poor	Analysis did not account for complexities, no information on how missing data was handled; included radiographic vertebral fractures; unclear whether outcome ascertainment was blinded to BMD measures.	Low concerns	None
Leslie et al, 2017[248] Manitoba BMD Registry	Poor	Only participants with at least 2 DXA measurements in a BMD registry were included; potential for selection bias; no information on how missing data handled, and unclear whether outcome ascertainment was blinded to BMD measures; outcomes based on administrative data.	Low concerns	None

Abbreviations: BMD=bone mineral density; DXA=dual-energy X-ray absorptiometry; Mr.Os=Osteoporotic Fractures in Men (study); ROB=risk of bias; SOF=Study of Osteoporotic Fractures; WHI=Women's Health Initiative.

Appendix D Table 53. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 1 Randomization and Allocation Concealment)

Author, Year	Was method of randomization adequate?	Was allocation concealment adequate?	Were there baseline imbalances between groups that suggest a problem with randomization?	ROB: Randomization or Selection	Comments on Bias Arising From Randomization or Selection
Adachi et al, 2009[296]	Yes	Yes	Yes	Some or unclear	Alendronate group had greater proportion of patients with history of UGI disease, active UGI disease, esophageal disease; no statistical comparison is given, but the differences are large enough to warrant some concern for ROB because it does not appear that these differences were corrected for in the analysis.
Ascott-Evans et al, 2003[252]	Yes	Yes	No	Low	None
Bala et al, 2014[321]	No information	No information	No	Some or unclear	Method of randomization and allocation concealment NR.
Bell et al, 2002[276]	No information	No information	No	Some or unclear	Method of randomization and allocation concealment NR
Bone et al, 1997[277]	No information	No information	No	Some or unclear	Method of randomization and allocation concealment NR.
Bone et al, 2008[285]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
Boonen et al, 2012[251]	Yes	Yes	No	Low	None
Chapurlat et al, 2013[288]	Yes	Yes	No	Low	None
Chesnut et al, 1995[253]	No information	No information	No	Some or unclear	Unclear how the randomization was generated; allocation concealment not described.
Cryer et al, 2005[297]	Yes	Yes	No	Low	None
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255]	Yes	Yes	No	Low	None
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283]	No information	No information	No	Some or unclear	Randomization and allocation concealment not described.

Appendix D Table 53. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 1 Randomization and Allocation Concealment)

Author, Year	Was method of randomization adequate?	Was allocation concealment adequate?	Were there baseline imbalances between groups that suggest a problem with randomization?	ROB: Randomization or Selection	Comments on Bias Arising From Randomization or Selection
Devogelaer et al, 1996[302]	No information	No information	No	Some or unclear	No information about method of randomization or allocation concealment.
Eisman et al, 2004[298]	Yes	Yes	No	Low	None
Fogelman et al, 2000[429]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR
Greenspan et al, 2002[299]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
Greenspan et al, 2003[300]	Yes	Yes	No	Low	None
Grey et al, 2010[262] Grey et al. 2009[263]	Yes	Yes	Yes	Some or unclear	28% with prior fractures in the zoledronic acid arm, 56% in the placebo arm, no sensitivity analyses.
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	Yes	Yes	No	Low	Statistician was unblinded to treatment allocation but had no access to patients. Staff member preparing infusions also had access to unblinded treatment allocation, was stated to have no access to patients.
Hosking et al, 2003[267]	No information	No information	No	Some or unclear	No details on randomization or allocation concealment.
Johnell et al, 2002[294]	Yes	Yes	No	Low	None
Koh et al, 2016[286]	No information	No information	No	Some or unclear	No details provided on randomization or allocation concealment.
Lewiecki et al, 2007[284] McClung et al, 2006[304]	Other	No information	No	Some or unclear	Details on allocation concealment NR.
Liberman et al, 1995[256]	No information	No information	No	Some or unclear	None
McClung et al, 2001[257]	No information	No information	No	Some or unclear	No information on randomization or allocation .
McClung et al, 2004[291]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
McClung et al, 2009[275]	No information	No information	No	Some or unclear	Method of randomization and allocation concealment NR.

Appendix D Table 53. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 1 Randomization and Allocation Concealment)

Author, Year	Was method of randomization adequate?	Was allocation concealment adequate?	Were there baseline imbalances between groups that suggest a problem with randomization?	ROB: Randomization or Selection	Comments on Bias Arising From Randomization or Selection
McClung et al, 2009[287]	Yes	Yes	No	Low	None
Mortensen et al, 1998[258]	No information	No	Yes	Some or unclear	No details on randomization or allocation concealment.
Nakamura et al, 2012[279]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment unclear.
Orwoll et al, 2012[278]	No information	Yes	No	Low	Method of randomization not explicitly reported, but use of an IVRS and permuted blocks suggest some method of computerized randomization was used.
Pols et al, 1999[259]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
Ravn et al, 1996[264]	No information	No information	No	Some or unclear	No information on randomization or allocation concealment.
Reginster et al, 2005[266]	No information	No information	No	Some or unclear	Details of randomization and allocation concealment NR.
Reid et al, 2002[260]	No information	No information	No	Some or unclear	No details on randomization or allocation concealment.
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]	Yes	Yes	No	Low	None
Riis et al, 2001[265]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
Shiraki et al, 2003[295]	No	No	No	Some or unclear	Details on randomization and allocation concealment NR.
Tanko et al, 2003[292]	No information	No information	No	Low	None
Thiebaud et al, 1997[293]	No information	No	No	Some or unclear	Details on randomization and allocation concealment NR.
Tucci et al, 1996[268]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.
Valimaki et al, 2007[261]	No information	No information	No	Some or unclear	Details on randomization and allocation concealment NR.

Abbreviations: NR=not reported; IVRS=interactive voice response system; ROB=risk of bias; UGI=upper gastrointestinal.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Adachi et al, 2009[296]	Overall: 16.2% G1: 18.6% G2: 11.6% Vary by outcome? No	No	Yes	No information	Low	None
Ascott-Evans et al, 2003[252]	No information	No information	No information	No information	Some or unclear	Unclear information on attrition.
Bala et al, 2014[321]	7% to 16% overall across the 2 studies	No	No	No information	Some or unclear	Unclear how missing data for harms was handled.
Bell et al, 2002[276]	1/65 was missing from the ITT	No	Yes	Other	Low	None
Bone et al, 1997[277]	Not clear	Other	Yes	Other	Some or unclear	Unclear how missing fracture/safety data was handled.
Bone et al, 2008[285]	Overall attrition: 3/332=0.09% G1: 2/166 (1.2%) G2: 1/166 (0.06%)	No	Yes	Other	Low	None

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Boonen et al, 2012[251]	58 (9.9%) vs. 71 (11.6%) discontinued the study	No	Yes	Yes	Low	The reasons for withdrawal look similar between the arms and some sensitivity analyses, including different types of imputation were done and the results suggest similar outcomes to the modified ITT efficacy analyses. Harms were reported on the full sample, but how data were obtained from those who withdrew consent is unclear; however, given the similar rates, there is no evidence suggesting bias.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Chapurlat et al, 2013[288]	Overall: 0.67% G1: 0% G2: 1.3% Overall: Unclear G1: Unclear G2: Unclear	No	Yes	Yes	Low	None
Chesnut et al, 1995[253]	No information	No information	No information	No information	Some or unclear	34/188 participants withdrew, leaving 154 participants; 168 available for intent-to-treat analyses, and 133 for per-protocol analysis. No details provided on proportion missing by arm.
Cryer et al, 2005[297]	Overall: 62/454 (13.7%) Alendronate: 31/224 (13.8%) Placebo: 31/230 (13.5%) No	No	Yes	Other	Low	None
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255]	Cummings, 1998 (participants without prior fracture): 5% without final followup radiographs; NR in other eligible publications	No	No information	No information	Low	None
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283]	Attrition varies by outcome, lowest for fractures: 475/7,868 (6.03%) G1: 231/3,933 (5.87%) G2: 244/3,935 (6.20%)	No	No information	Other	Some or unclear	Limited information on attrition and intent-to-treat analysis.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Devogelaer et al, 1996[302]	0%	No	Other	Other	Low	None
Eisman et al, 2004[298]	Overall: 6.2% Alendronate: 8.0% Placebo: 4.5% Vary by outcome? No	No	Yes	Other	Low	None
Fogelman et al, 2000[429]	Placebo: 33% Risedronate 2.5 mg: 67% Risedronate 5 mg: 37%	Yes	No	NR	High	High and differential attrition
Greenspan et al, 2002[299]	Overall: 6.9% Alendronate: 6.3% Placebo: 7.5% Vary by outcome? No	No	Yes	Other	Low	None
Greenspan et al, 2003[300]	Overall: 6.9% G1: 6.3% G2: 7.5% No	No	Yes	Yes	Low	None
Grey et al, 2010[262] Grey et al. 2009[263]	Overall: 2% Zoledronic acid: 4% Placebo: 0% Unclear whether outcomes were reported for the entire sample	No	Yes	No information	Some or unclear	Denominator used for outcomes is unclear.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	4.4% (8/180) did not receive study medication and 2.7% (5/180) withdrew Denominator not reported for harms so % with missing data for harms not available for each group; for benefits, data were missing for 2/45 for 1 mg arm, 2/45 for 2.5 mg arm, 2/45 for 5 mg arm, and 2/45 for placebo arm Unclear if % missing varied by reason for harms because the denominators were not reported	No	Yes	Other	Some or unclear	Denominator not reported for harms.
Hosking et al, 2003[267]	Overall at 3 months: 20% Alendronate: 21.5% Risedronate: 19.8% Placebo 17.6% Vary by outcome? No, study reports fractures as harms and uses the full sample Attrition at 12 months NR	Yes	Yes	No information	Some or unclear	Study lists full denominator in adverse events table, but unclear whether they obtained data on adverse events from all participants.
Johnell et al, 2002[294]	17% overall completed the study, but N missing outcomes by arm not reported	No information	No information	Yes	Some or unclear	Study notes that the analyses were based on intention to treat; denominators for harms appear to be the whole sample; attrition unclear.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Koh et al, 2016[286]	10/135 lost to followup but outcomes reported for all included participants (N=135), appears to be no missing data	No	Other	Other	Low	None
Lewiecki et al, 2007[284] McClung et al, 2006[304]	18.2% did not complete study; details of participants with missing outcomes NR	No information	Yes	Other	Some or unclear	Details on attrition NR.
Liberman et al, 1995[256]	Nonvertebral fractures and adverse events Overall: 0% G1: 0% G2: 0% Vertebral fractures Overall: 11.4% G1: 10.6% G2: 12% Vary by outcome: Yes	No	No information	Yes	Some or unclear	None
McClung et al, 2001[257]	36% overall Risedronate: 35% Placebo: 36% Similar reasons for discontinuation (details not reported)	Yes	Other	No	Some or unclear	High but nondifferential attrition.
McClung et al, 2004[291]	Overall: 1% Ibandronate 0.5 mg: 0.6% Ibandronate 1 mg: 0.6% Ibandronate 2 mg: 0% Placebo: 1.9% No	No	Yes	Yes	Low	None
McClung et al, 2009[275]	0%	No	Other	Other	Low	None

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
McClung et al, 2009[287]	Overall: 10% G1: 8.6% G2: 14.9% G3: 6.9% Vary by outcome: No	No	No	No	Some or unclear	ROB for harms data because it is limited to ITT analysis.
Mortensen et al, 1998[258]	Unclear	No	Yes	No	Some or unclear	14% did not complete treatment overall, but N for analysis unclear.
Nakamura et al, 2012[279]	Overall: 8.0% G1: (5/53) 9.4% G2: (4/54) 7.4% G3: (5/50) 10% G4: (3/55) 5.5% Probably no	No	Yes	Other	Low	None
Orwoll et al, 2012[278]	6% overall; 3% in control and 8% in active drug group	No	Yes	No information	Low	Slight difference in missing data between groups, but not enough to raise serious concerns for bias.
Pols et al, 1999[259]	NR	No information	No information	No information	Some or unclear	Details on attrition NR.

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Ravn et al, 1996[264]	Overall: 39/180, 22% G1: 4/30,13% G2: 8/30, 27% G3: 4/30, 13% G4: 6/30, 20% G5: 12/30, 40% G6: 5/30, 17% Yes	Yes	Yes	No information	Some or unclear	High overall and differential attrition; however, most safety outcomes appear to have been collected and reported on a larger subset of the population.
Reginster et al, 2005[266]	Overall: 3% Ibandronate 50 mg: 0 Ibandronate 50/100 mg: 0 Ibandronate 100 mg: 0 Ibandronate 150 mg: 3% Placebo: 8%	No	Yes	Yes	Low	None
Reid et al, 2002[260]	10% withdrew overall, details by arm NR	No	No information	Yes	Some or unclear	Distribution of loss to followup NR by arm; intention-to-treat analysis conducted but details NR.
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]		No	Yes	Yes	Low	None

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Riis et al, 2001[265]	Overall: 14% Ibandronate 2.5 continuously: 15% Ibandronate 20 mg intermittently: 15% Placebo: 11% Missing outcome data varying: no	No	Yes	Yes	Low	None
Shiraki et al, 2003[295]	3.8% overall Risedronate 1 mg: 3.8% Risedronate 2.5 mg: 0 Risedronate 5 mg: 5.3% Placebo: 5.5%	No	Yes	No	Low	Missing participants not included in the analyses but low overall rates.
Tanko et al, 2003[292]	Overall: 14% G1: NR G2: NR G3: NR G4: NR G5: NR No	No	Yes	Yes	Low	Unable to calculate group attrition.
Thiebaud et al, 1997[293]	Overall: 10% Ibandronate 0.25 mg: 12.5% (3/24) Ibandronate 0.50 mg: 3.7% (1/27) Ibandronate 1.0 mg: 11.5% (3/26) Ibandronate 2.0 mg: 8.7% (2/23) Placebo: 7.7% (2/26) Vary by outcome? No	No	Yes	Yes	Low	None

Appendix D Table 54. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 2 Missing Outcome Data)

Author, Year	What percentage of participants had missing outcome data overall? What percentage of participants had missing outcome data in each group?	Did the study have a percentage of participants with missing data that would raise concern for bias?	Are the proportion of participants and reasons for missing data similar across groups?	If a study had participants with missing data, were appropriate statistical methods used to evaluate the effect of the missing data?	ROB: Missing Outcome Data	Comments on Bias Arising From Missing Data
Tucci et al, 1996[268]	Overall: 29/478=6.0% (from Ns in Table IV) G1: 9.2% G2: 6.4% G3: 8.5% G4: 3.1% No	No	Yes	Other	Low	None
Valimaki et al, 2007[261]	Unclear	No information	No information	Other	Some or unclear	One crossover mentioned; attrition not described but modified ITT conducted.

Abbreviations: ITT=intention to treat; IV=intravenous; N=number; NR=not reported; ROB=risk of bias; vs.=versus.

Appendix D Table 55. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 3 Departures from Intended Interventions)

Author, Year	Were the patients unaware of the assigned intervention status?	Were the trial personnel/clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Adachi et al, 2009[296]	Yes	Yes	No information	No information	Low	No data on adherence; authors did not specifically say they performed an intention-to-treat analysis.
Ascott-Evans et al, 2003[252]	Yes	Yes	No information	Other	Low	None
Bala et al, 2014[321]	Yes	Yes	No information	No information	Some or unclear	No information about adherence or contamination.
Bell et al, 2002[276]	Yes	Yes	No information	No	Some or unclear	Compliance/adherence mentioned in methods, but not reported.
Bone et al, 1997[277]	Yes	Yes	No information	No	Some or unclear	Compliance/adherence mentioned in methods, but not reported.
Bone et al, 2008[285]	No information	Yes	Yes	No	Low	None
Boonen et al, 2012[251]	Yes	Yes	Other	No information	Low	None
Chapurlat et al, 2013[288]	Yes	Yes	Yes	No	Low	None
Chesnut et al, 1995[253]	Yes	Yes	No information	No information	Low	None
Cryer et al, 2005[297]	Yes	Yes	No information	No	Low	None
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255]	Yes	Yes	Yes	No	Low	None

Appendix D Table 55. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 3 Departures from Intended Interventions)

Author, Year	Were the patients unaware of the assigned intervention status?	Were the trial personnel/clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283]	No information	No information	Other	No information	Some or unclear	Blinding not described (although data monitoring and safety are described as being unblinded, suggesting that the rest of the operations were blinded).
Devogelaer et al, 1996[302]	Yes	Yes	No information	No information	Some or unclear	Methods mention tablet counting and questioning subject to measure adherence, but adherence data is not reported.
Eisman et al, 2004[298]	Yes	Yes	No information	No	Low	None
Fogelman et al, 2000[429]	Yes	Yes	No information	No	Low	None
Greenspan et al, 2002[299]	Yes	Yes	Yes	No	Low	None
Greenspan et al, 2003[300]	Yes	Yes	Yes	No	Low	None
Grey et al, 2010[262] Grey et al. 2009[263]	Yes	Yes	Yes	No	Low	None
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	Yes	Other	Other	No	Low	None
Hosking et al, 2003[267]	Yes	Yes	No information	No	Low	None
Johnell et al, 2002[294]	Yes	Yes	No information	No	Low	None
Koh et al, 2016[286]	Yes	Yes	Other	Other	Low	None

Appendix D Table 55. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 3 Departures from Intended Interventions)

Author, Year	Were the patients unaware of the assigned intervention status?	Were the trial personnel/clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Lewiecki et al, 2007[284] McClung et al, 2006[304]	Yes	Yes	No information	No information	Low	None
Liberman et al, 1995[256]	Yes	Yes	No information	No information	Some or unclear	None
McClung et al, 2001[257]	Other	No information	No information	No information	Some or unclear	No details on blinding.
McClung et al, 2004[291]	Yes	Yes	No information	No information	Low	None
McClung et al, 2009[275]	Yes	Yes	No information	No information	Some or unclear	No information about adherence to study medication
McClung et al, 2009[287]	Yes	Yes	Yes	No	Low	None
Mortensen et al, 1998[258]	Yes	Yes	No information	No information	Low	None
Nakamura et al, 2012[279]	Yes	Yes	No information	No	Low	None
Orwoll et al, 2012[278]	Yes	Yes	Yes	No	Low	None
Pols et al, 1999[259]	Yes	Yes	No information	No	Low	None
Ravn et al, 1996[264]	Yes	No	No information	No	Some or unclear	Data safety review committee was not blinded to treatment, and they monitored adverse events during each step. Information on compliance was not provided.
Reginster et al, 2005[266]	Yes	Yes	No information	No	Low	None
Reid et al, 2002[260]	Yes	Yes	Yes	No	Low	None

Appendix D Table 55. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 3 Departures from Intended Interventions)

Author, Year	Were the patients unaware of the assigned intervention status?	Were the trial personnel/clinicians unaware of the assigned intervention status?	Was intervention fidelity adequate?	Did the study have enough crossovers or contamination that would raise concern for bias?	ROB: Departures From Intended Interventions	Comments on Bias Arising From Departure From Intended Interventions
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]	Yes	Yes	Yes	No	Low	None
Riis et al, 2001[265]	Yes	Yes	No information	No	Low	None
Shiraki et al, 2003[295]	Yes	Yes	No information	No	Low	None
TankoD et al, 2003[292]	Yes	Yes	No information	No	Low	Large proportion of patients in each study group took ≥75% of study medication: 89% placebo, 88.8% (5 mg), 90.1% (10 mg) and 88.7% (20 mg) patients.
Thiebaud et al, 1997[293]	Yes	No	No information	No	Some or unclear	Intervention only partly blinded to investigators.
Tucci et al, 1996[268]	Yes	Yes	No information	No	Low	None
Valimaki et al, 2007[261]	Yes	Yes	Yes	Yes	Low	None

Abbreviations: ROB=risk of bias.

Appendix D Table 56. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 4 Outcome Measurement)

Author, Year	Were benefit outcomes adequately described, valid, and reliable and was the duration of followup adequate?	Were harm outcomes adequately described, valid, and reliable with an adequate duration of followup?	Were outcome assessors masked to group assignment?	ROB: Outcome Measurement	Comments on Bias Arising From Measurement of Outcomes
Adachi et al, 2009[296]	NA: No benefit outcomes	Yes	Yes	Some or unclear	There was not specific information about how often patients were assessed for harms, though did describe adequate blinding of patients.
Ascott-Evans et al, 2003[252]	Yes	Yes	Yes	Low	None
Bala et al, 2014[321]	NA: no benefit outcomes	No	No information	High	No information about how harms were ascertained, and no information about whether outcome assessors were masked to treatment allocation.
Bell et al, 2002[276]	NA: no benefit outcomes	Yes	Other	Some or unclear	Outcome assessment blinding was not explicitly reported, but assessors were probably masked given the objective nature of most of the outcomes assessed (i.e., lab or radiographic outcomes), double-blind implies study personnel conducting the study were blind and since patients were not aware of allocation; AEs were likely blindly assessed.
Bone et al, 1997[277]	Yes	Yes	No information	Some or unclear	Outcome assessment blinding was not explicitly reported, but assessors were probably masked given the objective nature of most of the outcomes assessed (i.e., lab or radiographic outcomes), double-blind implies study personnel conducting the study were blind and since patients were not aware of allocation so AEs were likely blindly assessed.

Appendix D Table 56. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 4 Outcome Measurement)

Author, Year	Were benefit outcomes adequately described, valid, and reliable and was the duration of followup adequate?	Were harm outcomes adequately described, valid, and reliable with an adequate duration of followup?	Were outcome assessors masked to group assignment?	ROB: Outcome Measurement	Comments on Bias Arising From Measurement of Outcomes
Bone et al, 2008[285]	Yes	Yes	Yes	Low	Assessors blinded to assignment when making determination that adverse event was treatment related, but other details on outcome assessors NR.
Boonen et al, 2012[251]	Yes	Yes	Yes	Low	None
Chapurlat et al, 2013[288]	NA: No benefit outcomes	Yes	Yes	Low	None
Chesnut et al, 1995[253]	Yes	NA: No harm outcomes	No information	Some or unclear	Unclear if outcome assessors were blinded; harms not reported by study arm.
Cryer et al, 2005[297]	NA: No benefit outcomes	Yes	No information	Some or unclear	No details on masking of outcome assessors.
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255]	Yes	Yes	Yes	Low	None
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283]	Yes	Yes	Yes	Low	None
Devogelaer et al, 1996[302]	NA: no benefit outcomes	Yes	No information	Some or unclear	Unclear whether outcome assessors were blinded to treatment allocation.

Appendix D Table 56. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 4 Outcome Measurement)

Author, Year	Were benefit outcomes adequately described, valid, and reliable and was the duration of followup adequate?	Were harm outcomes adequately described, valid, and reliable with an adequate duration of followup?	Were outcome assessors masked to group assignment?	ROB: Outcome Measurement	Comments on Bias Arising From Measurement of Outcomes
Eisman et al, 2004[298]	NA: No benefit outcomes	Yes	Yes	Low	None
Fogelman et al, 2000[429]	No	Yes	No information	High for fractures, some or unclear for harms	Masking of outcome assessors NR.
Greenspan et al, 2002[299]	NA: No benefit outcomes	Yes	No information	Some or unclear	Masking of outcome assessor unclear.
Greenspan et al, 2003[300]	NA: No benefit outcomes	Yes	Yes	Low	None
Grey et al, 2010[262] Grey et al. 2009[263]	No	Yes	Yes	Some or unclear	Outcomes not well specified (for fractures).
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	Yes	Other	No	Some or unclear	Timing of data collection unclear for some harms.
Hosking et al, 2003[267]	Yes	Yes	No information	Some or unclear	No details on masking of outcome assessors.
Johnell et al, 2002[294]	NA: No benefit outcomes	Yes	No information	Some or unclear	Masking of outcome assessors NR.
Koh et al, 2016[286]	Yes	Yes	Yes	Low	None
Lewiecki et al, 2007[284] McClung et al, 2006[304]	Yes	Yes	No information	Some or unclear	Masking of outcome assessors unclear.
Liberman et al, 1995[256]	Yes	Yes	Other	Some or unclear	None
McClung et al, 2001[257]	Yes	Yes	No information	Some or unclear	No details on masking.
McClung et al, 2004[291]	NA: No benefit outcomes	Yes	Yes	Low	None
McClung et al, 2009[275]	Yes	Yes	No information	Some or unclear	Unclear whether outcome assessors were masked.

Appendix D Table 56. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 4 Outcome Measurement)

Author, Year	Were benefit outcomes adequately described, valid, and reliable and was the duration of followup adequate?	Were harm outcomes adequately described, valid, and reliable with an adequate duration of followup?	Were outcome assessors masked to group assignment?	ROB: Outcome Measurement	Comments on Bias Arising From Measurement of Outcomes
McClung et al, 2009[287]	NA: No benefit outcomes	Yes	Yes	Low	None
Mortensen et al, 1998[258]	Yes	Yes	No information	Low	None
Nakamura et al, 2012[279]	Yes	Yes	Yes	Low	None
Orwoll et al, 2012[278]	No	Yes	Yes	Low	Fractures were captured as adverse events; details about ascertainment NR.
Pols et al, 1999[259]	Yes	Yes	No information	Some or unclear	Details on masking of outcome assessors NR.
Ravn et al, 1996[264]	Yes	Yes	No information	Some or unclear	No information on masking of outcome assessors.
Reginster et al, 2005[266]	Yes	Yes	No information	Some or unclear	Masking of outcome assessors unclear.
Reid et al, 2002[260]	Yes	Yes	No information	Some or unclear	Masking of outcome assessors NR.
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]	Yes	Yes	Yes	Low	None
Riis et al, 2001[265]	Yes	Yes	No information	Some or unclear	None
Shiraki et al, 2003[295]	NA: No benefit outcomes	Yes	No information	Some or unclear	Masking of outcome assessors unclear.
Tanko et al, 2003[292]	NA: No benefit outcomes	Yes	Yes	Low	None
Thiebaud et al, 1997[293]	NA: No benefit outcomes	Yes	Yes	Low	None
Tucci et al, 1996[268]	Yes	Yes	Yes	Low	None
Valimaki et al, 2007[261]	Yes	Yes	No information	Some or unclear	Details on masking of outcome assessors NR.

Abbreviations: NA=not applicable; NR=not reported; ROB=risk of bias.

Appendix D Table 57. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 5 Selective Outcome Reporting and Overall Risk of Bias)

Author, Year	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study Quality	ROB Rating Justification
Adachi et al, 2009[296]	Yes	Low	None	Fair	Baseline differences between groups raise some concerns for risk of bias, moderate risk of bias related to outcome measurement.
Ascott-Evans et al, 2003[252]	Yes	Low	None	Fair	Unclear information on attrition.
Bell et al, 2002[276]	No information	Some or unclear	No trial registration or published protocol.	Fair	Method of randomization and allocation concealment NR; no mention of whether outcome assessors were blinded, but presumably they were since study personnel asked patients about AE and both were blinded; no reporting of fidelity; no trial registration or published protocol to evaluate selective outcome reporting.
Bone et al, 1997[277]	No	Some or unclear	Authors selected to present mostly per protocol analyses for several outcomes not of relevance to our review. No trial registration or protocol publication.	Fair	Method of randomization and allocation concealment NR: unclear how missing data was handled; fractures reported as safety events, so unclear whether ITT used. No trial registration or published protocol; authors report per protocol analyses for outcomes not relevant to our use; but unclear whether the relevant outcomes we are interested in are per protocol or ITT.
Bone et al, 2008[285]	Yes		None	Fair	Details on randomization and allocation concealment unclear.
Bala et al, 2014[321]	No information	Some or unclear	No published study protocol.	Poor	Reporting very incomplete; unable to fully assess most domains. The only outcomes this study reports that would be eligible are AEs, and the method of ascertainment for AEs is not described at all and it is unclear whether outcome assessors were masked.
Boonen et al, 2012[251]	Yes	Low	None	Good	None
Chapurlat et al, 2013[288]	Yes	Low	None	Fair	Considering interactive voice response allocation with minimization scheme to be just adequate and unclear way dropouts handled.
Chesnut et al, 1995[253]	No information		None	Fair	Limited or no details on randomization, allocation concealment, or attrition.

Appendix D Table 57. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 5 Selective Outcome Reporting and Overall Risk of Bias)

Author, Year	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study Quality	ROB Rating Justification
Cryer et al, 2005[297]	Yes	Some or unclear	None	Fair	None
Cummings et al, 1998[254] Bauer et al, 2000[289] Cummings et al, 2007[290] Quandt et al, 2005[255]	Yes	Low	None	Good	None
Cummings et al, 2009[280] Watts et al, 2012[303] Simon et al, 2013[281] McCloskey et al, 2012[282] Palacios et al, 2015[283]	Yes	Low	None	Fair	Some uncertainties in reporting of randomization, allocation concealment, blinding, and attrition.
Devogelaer et al, 1996[302]	Yes	Low		Fair	Some concerns for bias because method of randomization/allocation concealment NR; adherence to intervention NR; and outcome assessor masking NR.
Eisman et al, 2004[298]	Yes	Low	None	Good	None
Fogelman et al, 2000[429]	Yes	Low	None	Poor	Details on randomization, allocation concealment, and masking of outcome assessors unclear; high and differential attrition.
Greenspan et al, 2002[299]	Yes	Low	Details on randomiz-ation, allocation concealment, and masking of outcome assessor unclear	Fair	None
Greenspan et al, 2003[300]	Yes	Low	None	Good	None

Appendix D Table 57. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 5 Selective Outcome Reporting and Overall Risk of Bias)

Author, Year	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study Quality	ROB Rating Justification
Grey et al, 2010[262] Grey et al. 2009[263]	No	Low	None	Fair	Differences at baseline on prior fractures; no sensitivity analyses; denominator for outcomes unclear; fractures outcomes not clearly specified.
Grey et al, 2012[272] Grey et al, 2014[273] Grey et al, 2017[274]	Yes	Low	None	Fair	Attrition for harms unclear; timing of data collection unclear for some harms.
Hosking et al, 2003[267]	Yes	Low	None	Fair	No details on randomization, allocation concealment, and masking of outcome assessors; some details on attrition NR.
Johnell et al, 2002[294]	Yes	Some or unclear	Details on masking and attrition NR	Fair	None
Koh et al, 2016[286]	Yes	Low	None	Fair	Details on randomization and allocation concealment NR.
Lewiecki et al, 2007[284] McClung et al, 2006[304]	Yes	Low	None	Fair	Details on allocation concealment, attrition, and masking of outcome assessors unclear.
Liberman et al, 1995[256]	Yes	Some or unclear	None	Fair	None
McClung et al, 2001[257]	Yes	Low	None	Fair	Details on randomization, allocation concealment, blinding of staff, and masking of outcome assessors NR.
McClung et al, 2004[291]	Yes	Low	None	Fair	Details on randomization and allocation concealment NR.
McClung et al, 2009[275]	Yes	Low		Fair	Unclear methods of randomization, allocation concealment. Outcome assessor masking NR; fidelity to intervention NR.
McClung et al, 2009[287]	Yes	Low	None	Fair	None
Mortensen et al, 1998[258]	Yes	Low	None	Fair	No details on randomization or allocation concealment or masking of outcome assessors.
Nakamura et al, 2012[279]	Yes	Low	None	Fair	Details on randomization and allocation concealment NR.

Appendix D Table 57. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 5 Selective Outcome Reporting and Overall Risk of Bias)

Author, Year	Are the reported effects unlikely to be selected on the basis of the results from multiple outcome measurements within the domain, multiple analyses or different subgroups?	ROB: Selective Outcome Reporting	Comments on Bias Arising From Selective Reporting	Overall Study Quality	ROB Rating Justification
Orwoll et al, 2012[278]	Yes	Low	None	Fair for fractures; low for harms and mortality	Fractures were captured as adverse events; details about ascertainment NR.
Pols et al, 1999[259]	Yes	Low	None	Fair	Details on randomization, allocation concealment, masking, and attrition NR.
Ravn et al, 1996[264]	Yes	Low	None	Fair	No information on randomization, allocation concealment, or masking of outcome assessors.
Reginster et al, 2005[266]	Yes	Low	None	Fair	Details on randomization, allocation concealment, and masking unclear.
Reid et al, 2002[260]	Yes	Low	None	Fair	Details NR on randomization, allocation concealment, attrition by arm, and masking of outcome assessors.
Reid et al, 2018[269] Reid et al, 2019[270] Reid et al, 2020[301] Reid et al, 2021[271]	Yes	Low	None	Good	None
Riis et al, 2001[265]	Yes	Low	None	Fair	Details on randomization, allocation concealment, and masking unclear.
Shiraki et al, 2003[295]	Yes	Low	None	Fair	No details on randomization, allocation concealment, and masking of outcome assessors.
Tanko et al, 2003[292]	Yes	Low	None	Fair	No information provided on method of randomization or concealment. Not able to calculate group attrition.
Thiebaud et al, 1997[293]	Yes	Low	None	Fair	Details on randomization and allocation concealment NR; only some arms blinded from investigators.
Tucci et al, 1996[268]	Yes	Low	None	Fair	Details on randomization and allocation concealment NR.
Valimaki et al, 2007[261]	Yes	Low	None	Fair	Details on randomization, allocation concealment, and masking of outcome assessors NR; attrition not described.

Appendix D Table 57. Risk of Bias of Included Trials for Key Questions 4 and 5 (Domain 5 Selective Outcome Reporting and Overall Risk of Bias)

Abbreviations: AE=adverse event; NR=not reported; ROB=risk of bias.

Appendix D Table 58. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 1 Bias Due to Confounding)—Part 1

Author, Year	1.1 Is confounding of the effect of intervention likely in this study?	1.2 Was the analysis based on splitting participants' followup time according to intervention received?	1.3 Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?	1.4 Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	1.5 Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?
Abrahamsen et al, 2011[430]	Yes	No	N/A	Probably no	N/A
Black , 2020[431]	Yes	No	N/A	Yes	Probably no
Cardwell, et al, 2010[432]	Yes	No	N/A	Yes	Yes
Chiang, et al, 2012[433]	Yes	No	N/A	No	N/A
Choi, et al, 2020[434]	Yes	No	N/A	Probably no	N/A
Kim, et al, 2021[435]	Yes	No	N/A	Probably no	N/A
Lee et al, 2012[436]	Yes	No	N/A	No	N/A
Lee et al, 2019[307]	Yes	No	N/A	Probably yes	Probably no
Nordström et al, 2020[437]	Yes	No	N/A	Probably no	N/A
Passarelli et al, 2013[438]	Yes	Yes	Probably no	Probably yes	Probably yes
Pazianas et al, 2012[305]	Yes	No	N/A	Probably yes	Probably yes
Rodriguez et al, 2020[439]	Yes	Yes	Yes	No	N/A
Rubin et al, 2020[306]	Yes	No	N/A	Probably yes	Probably no
Thadani et al, 2016[440]	Yes	Yes	Probably yes	Probably yes	Probably yes
Vestergaard et al, 2011[441]	Yes	No	N/A	Probably no	N/A
Wang et al, 2016[442]	Yes	No	N/A	No	N/A
Yang et al, 2018[443]	Yes	No	N/A	Probably no	N/A
Yuh et al, 2014[444]	Yes	No	N/A	Probably no	N/A

Abbreviations: N/A=not applicable.

Appendix D Table 59. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 1 Bias Due to Confounding)—Part 2

Author, Year	1.6 Did the authors control for any post-intervention variables that could have been affected by the intervention?	1.7 Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	1.8 Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	Risk of bias-Confounding	Support for Judgment
Abrahamsen et al, 2011[430]	No	N/A	N/A	Serious	Secondary data sources used; did not control for smoking, alcohol use, or GERD
Black et al, 2020[431]	No	N/A	N/A	Moderate	All covariates measured from electronic health record data; unclear how accurate or complete such data are for things like smoking status and self-reported race/ethnicity
Cardwell et al, 2010[432]	No	N/A	N/A	Moderate	Some risk for residual confounding, but appears to have included most important confounders, including smoking and alcohol
Chiang et al, 2012[433]	No	N/A	N/A	Serious	Used secondary data sources for covariate measures, did not have any information about smoking or alcohol use, did not have any information about hormone use, or comorbidities important for some cancers (like ulcerative colitis)
Choi et al, 2020[434]	No	N/A	N/A	Moderate	Claims data used for most covariates, self-report for others but which ones not specified; stratified analyses conducted based on some covariates but no adjusted results reported overall
Kim et al, 2021[435]	Probably no	N/A	N/A	Moderate	Claims/administrative data used for all confounders
Lee et al, 2012[436]	No	N/A	N/A	Serious	Did not include BMI, family history, tobacco use, or alcohol use and hormone use (for women) in the analysis; these are all important confounders for evaluating the risk of developing cancer across various organ systems

Appendix D Table 59. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 1 Bias Due to Confounding)—Part 2

Author, Year	1.6 Did the authors control for any post-intervention variables that could have been affected by the intervention?	1.7 Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	1.8 Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	Risk of bias-Confounding	Support for Judgment
Lee et al, 2019[307]	No	N/A	N/A	Moderate	Retrospective design, all secondary data sources
Nordström et al, 2020[437]	No	N/A	N/A	Moderate	Used matching on age, sex, origin, history of prior fracture, or hip surgery; reported estimates adjusted for confounders but did not specify what covariates were used for the adjustment
Passarelli et al, 2013[438]	No	N/A	N/A	Moderate	Some baseline differences and potential for residual confounding
Pazianas et al, 2012[305]	No	N/A	N/A	Moderate	Relied on secondary data sources to measure confounders, age-matched comparison group, propensity matching; adjusted for age, comorbidities, GI disease, HRT, drug use. Did not adjust for smoking status.
Rodriguez et al, 2020[439]	No	Yes	Probably no	Serious	All data, including those for confounding variables were from secondary data sources and health registries. Did not include smoking or alcohol use, which would be critical for cardiovascular outcomes.
Rubin et al, 2020[306]	No	N/A	N/A	Moderate	Confounders measured entirely through claims and administrative data. Did not control for baseline history of calcium and vitamin D levels, bone density, body mass index, smoking and alcohol exposure, hypertension and metabolic syndrome

Appendix D Table 59. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 1 Bias Due to Confounding)—Part 2

Author, Year	1.6 Did the authors control for any post-intervention variables that could have been affected by the intervention?	1.7 Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	1.8 Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	Risk of bias-Confounding	Support for Judgment
Thadani et al, 2016[440]	No	Probably yes	Probably yes	Moderate	Risk of unmeasured confounding; authors conducted a secondary analysis to evaluate effect of time-varying confounding from bisphosphonate use
Vestergaard et al, 2011[441]	Probably no	N/A	N/A	Serious	Used secondary data sources, did not adjust directly for alcohol use or smoking, used proxy measures for those variables
Wang et al, 2016[442]	No	N/A	N/A	Serious	The analysis failed to control for tobacco use, BMI, anticoagulant use, CVD medication use (as a proxy for severity of disease), all critical confounders when considering the type of CVD outcomes reported by the study. All measures based on claims/administrative data.
Yang et al, 2018[443]	No	N/A	N/A	Serious	Claims data used for all confounder measures; other than matching, the results did not control for any variables (including smoking, alcohol use, which are important confounders for association with atrial fibrillation)
Yuh et al, 2014[444]	No	N/A	N/A	Moderate	Secondary data analysis, all covariate measurement through claims data

Abbreviations: BMI=body mass index; CVD=cardiovascular disease; GERD=gastroesophageal reflux disease; GI=gastrointestinal; HRT=hormone replacement therapy; N/A=not applicable.

Appendix D Table 60. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 2 Bias in Selection of Participants into the Study)

Author, Year	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	2.2. Were the post-intervention variables that influenced selection likely to be associated with intervention?	2.3 Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	2.4 Do start of followup and start of intervention coincide for most participants?	2.5 Were adjustment techniques used that are likely to correct for the presence of selection biases?	Risk of Bias—Selection	Support for Judgment
Abrahamsen et al, 2011[430]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None
Black et al, 2020[431]	No	Not applicable	Not applicable	Probably no	No	Serious	Not limited to new users, includes prevalent users
Cardwell et al, 2010[432]	Probably yes	Probably yes	Probably yes	Probably no	Probably yes	Serious	Not entirely clear whether this was an inception cohort. The first 6 months of followup excluded as any cancers diagnosed during this time would be unlikely attributable to bisphosphonate exposure. Control group persons were selected without regard to bisphosphonate use.
Chiang et al, 2012[433]	No	Not applicable	Not applicable	Yes	Not applicable	Low	Inception cohort
Choi et al, 2020[434]	Yes	Yes	Yes	No	No	Serious	Followup observation for outcome did not start until after 2–4 years of exposure; participants who died or who were diagnosed with cancer before the start of followup observation were excluded
Kim et al, 2021[435]	Yes	Yes	Yes	Yes	Probably no	Serious	Participants who died within 1 year of index or who were diagnosed with ONJ within 6 months of the index date were excluded from the exposed group
Lee et al, 2012[436]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None
Lee et al, 2019[307]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None

Appendix D Table 60. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 2 Bias in Selection of Participants into the Study)

Author, Year	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	2.2. Were the post-intervention variables that influenced selection likely to be associated with intervention?	2.3 Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	2.4 Do start of followup and start of intervention coincide for most participants?	2.5 Were adjustment techniques used that are likely to correct for the presence of selection biases?	Risk of Bias—Selection	Support for Judgment
Nordström et al, 2020[437]	Yes	No information	Not applicable	Yes	No information	Moderate	Nonusers who died before the users last dispensed dose of drug were excluded from analysis and replaced with a new nonuser
Passarelli et al, 2013[438]	No	Not applicable	Not applicable	No		Serious	About a third of the user cohort were using at baseline; the rest were new users over the duration of followup. Thus, this is not an inception cohort.
Pazianas et al, 2012[305]	No	Not applicable	Not applicable	Yes	Not applicable	Low	Inception cohort but not restricted to those with diagnosis of osteoporosis. However, alendronate has no other indications so likely not important.
Rodriguez et al, 2020[439]	No	Not applicable	Not applicable	No information	Not applicable	Moderate	Lack of clarity regarding whether only new users or whether also contained some prevalent users
Rubin et al, 2020[306]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None
Thadani et al, 2016[440]	No	Not applicable	Not applicable	No	No	Serious	Included both prevalent users and incident users; therefore, not an inception cohort. Also, did not exclude persons with known history of atrial fibrillation.
Vestergaard et al, 2011[441]	No	Not applicable	Not applicable	Probably yes	Not applicable	Low	None
Wang et al, 2016[442]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None
Yang et al, 2018[443]	No	Not applicable	Not applicable	Yes	Not applicable	Low	None

Appendix D Table 60. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 2 Bias in Selection of Participants into the Study)

Author, Year	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	2.2. Were the post-intervention variables that influenced selection likely to be associated with intervention?	2.3 Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	2.4 Do start of followup and start of intervention coincide for most participants?	2.5 Were adjustment techniques used that are likely to correct for the presence of selection biases?	Risk of Bias—Selection	Support for Judgment
Yuh et al, 2014[444]	No	Not applicable	Not applicable	Probably no	No	Serious	Unclear whether the BP-exposed cohort were new users or prevalent users

Abbreviations: BP=bisphosphonate; ONJ=osteonecrosis of the jaw.

Appendix D Table 61. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 3 Bias in Classification of Intervention)

Author, Year	3.1 Were intervention groups clearly defined?	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	Risk of Bias—Classification of Interventions	Support for Judgment
Abrahamsen et al, 2011[430]	Yes	Yes	No	Low	None
Black et al, 2020[431]	Probably yes	Probably yes	No	Moderate	Unclear what the category of “not yet used” refers to. They are classified as users in the analysis, but it is not clear they used the drug.
Cardwell et al, 2010[432]	Yes	Yes	No	Low	None
Chiang et al, 2012[433]	Yes	Yes	No	Low	None
Choi et al, 2020[434]	Yes	Yes	No	Low	None
Kim et al, 2021[435]	Yes	Yes	Probably no	Low	None
Lee et al, 2012[436]	Yes	Yes	Probably no	Low	None
Lee et al, 2019[307]	Yes	Yes	Probably no	Low	None
Nordström et al, 2020[437]	Probably yes	Probably no	No	Serious	Potential for reverse causation
Passarelli et al, 2013[438]	Yes	Yes	No	Low	None
Pazianas et al, 2012[305]	Yes	Yes	No	Low	None
Rodriguez et al, 2020[439]	Yes	Yes	No	Low	None
Rubin et al, 2020[306]	Yes	Yes	No	Low	None
Thadani et al, 2016[440]	Yes	Yes	No	Low	None
Vestergaard et al, 2011[441]	Yes	Yes	No	Low	None
Wang et al, 2016[442]	Yes	Yes	No	Low	None
Yang et al, 2018[443]	Yes	Yes	No	Low	None
Yuh et al, 2014[444]	Probably no	Yes	Probably no	Serious	No information about how bisphosphonate users were defined by the large database used as a data source. Some control patients might have used over-the-counter bisphosphonates

Appendix D Table 62. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 4 Bias Due to Deviations From Intended Intervention)

Author, Year	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	4.2. Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	4.3. Were important co-interventions balanced across intervention groups?	4.4. Was the intervention implemented successfully for most participants?	4.5. Did study participants adhere to the assigned intervention regimen?	4.6 Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Risk of Bias—Deviations From Intended Intervention	Support for Judgment
Abrahamsen et al, 2011[430]	Probably yes	Yes	Not applicable	Not applicable	Not applicable	Not applicable	Moderate	Alendronate-exposed individuals were more likely to have undergone upper GI endoscopy, which could lead to surveillance bias.
Black et al, 2020[431]	No	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Cardwell et al, 2010[432]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Moderate	9% of the control group subsequently received an RX for BP at a date later than the index date of their matched case. These were not excluded because the RX could have been for cancer-related osteoporosis or metastases and excluding them would have reduced the risk of cancer in the control cohort. However, leaving them in could result in a bias to the null.
Chiang et al, 2012[433]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Choi et al, 2020[434]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Kim et al, 2021[435]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Lee et al, 2012[436]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None

Appendix D Table 62. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 4 Bias Due to Deviations From Intended Intervention)

Author, Year	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	4.2. Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	4.3. Were important co-interventions balanced across intervention groups?	4.4. Was the intervention implemented successfully for most participants?	4.5. Did study participants adhere to the assigned intervention regimen?	4.6 Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Risk of Bias—Deviations From Intended Intervention	Support for Judgment
Lee et al, 2019[307]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Nordström et al, 2020[437]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Passarelli et al, 2013[438]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Pazianas et al, 2012[305]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Rodriguez et al, 2020[439]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Rubin et al, 2020[306]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Thadani et al, 2016[440]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Vestergaard et al, 2011[441]	No information	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	No information about lower or upper endoscopy rates during period of followup
Wang et al, 2016[442]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Yang et al, 2018[443]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None
Yuh et al, 2014[444]	Probably no	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Low	None

Abbreviations: BP=bisphosphonate; GI=gastrointestinal; RX=prescription.

Appendix D Table 63. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 5 Bias Due to Missing Data)

Author, Year	5.1 Were outcome data available for all, or nearly all, participants?	5.2 Were participants excluded due to missing data on intervention status?	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	5.4 Are the proportion of participants and reasons for missing data similar across interventions?	5.5 Is there evidence that results were robust to the presence of missing data?	Risk of Bias—Missing Data	Support for Judgment
Abrahamsen et al, 2011[430]	Yes	No	No	Not applicable	Not applicable	Low	None
Black et al, 2020[431]	Probably yes	Probably no	Probably no	Not applicable	Not applicable	Moderate	Unclear whether complete covariate data and outcome data were available for all users of bisphosphonates. Users could have experienced outcomes not captured in the health systems data systems.
Cardwell et al, 2010[432]	Probably no	No	Probably yes	No information	Yes	Moderate	Noted 46,000 eligible but only reported on 41,000 with at least 6 months followup data. Some missing data on smoking, alcohol use; reported sensitivity analyses to account for missing confounder data and no impact on outcomes.
Chiang et al, 2012[433]	Yes	Probably no	Probably no	Not applicable	Not applicable	Low	None
Choi et al, 2020[434]	No information	No	Yes	No information	No information	Moderate	7% (3,367) excluded for missing covariate information in first cohort, 18% (5,612) excluded for missing covariate information in second cohort. Given rare outcome (GI cancer), this level of missing data is concerning for introducing selection bias. Although authors evaluated robustness of findings from exclusion of persons who died or were diagnosed with cancer before the index date, no such analysis was done for persons excluded for missing covariate information.
Kim et al, 2021[435]	Probably yes	No	Probably yes	No information	No information	Moderate	Persons with missing data on smoking, alcohol, or BMI at baseline were excluded from analysis.
Lee et al,	Probably yes	Probably no	Probably no	Not applicable	Not applicable	Low	None

Appendix D Table 63. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 5 Bias Due to Missing Data)

Author, Year	5.1 Were outcome data available for all, or nearly all, participants?	5.2 Were participants excluded due to missing data on intervention status?	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	5.4 Are the proportion of participants and reasons for missing data similar across interventions?	5.5 Is there evidence that results were robust to the presence of missing data?	Risk of Bias—Missing Data	Support for Judgment
2012[436]							
Lee et al, 2019[307]	Probably yes	No information	No information	No information	No information	Low	None
Nordström et al, 2020[437]	Probably yes	Probably no	Probably no	Not applicable	Not applicable	Low	None
Passarelli et al, 2013[438]	Probably no	No	Yes	No information	No information	Moderate	139 persons reporting a diagnosis of CRC were excluded from the analysis because the diagnosis could not be verified or the diagnosis was adenocarcinoma in situ. No sensitivity analyses conducted with these persons.
Pazianas et al, 2012[305]	Yes	No	No information	Not applicable	Not applicable	Low	No discussion of any missing data
Rodriguez et al, 2020[439]	Probably yes	No	No	Not applicable	Not applicable	Low	None
Rubin et al, 2020[306]	Probably yes	Probably no	No information	Not applicable	Not applicable	Low	None
Thadani et al, 2016[440]	Probably yes	No	Yes	No information	No information	Moderate	Data available for 91.2% of the cohort; however, no information on missing data reported by group.
Vestergaard et al, 2011[441]	Yes	Probably no	Probably no	Not applicable	Not applicable	Low	None
Wang et al, 2016[442]	Probably yes	Probably no	No information	Not applicable	Not applicable	Low	None
Yang et al, 2018[443]	Probably yes	Probably no	No information	Not applicable	Not applicable	Low	None
Yuh et al, 2014[444]	Probably yes	Probably no	Probably no	Not applicable	Not applicable	Low	None

Abbreviations: BMI=body mass index; CRC=colorectal cancer; GI=gastrointestinal.

Appendix D Table 64. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 6 Bias in Measurement of Outcomes)

Author, Year	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	6.2 Were outcome assessors aware of the intervention received by study participants?	6.3 Were the methods of outcome assessment comparable across intervention groups?	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Risk of Bias—Measurement of Outcomes	Support for Judgment
Abrahamsen et al, 2011[430]	Probably no	Yes	Yes	No	Low	None
Black et al, 2020[431]	Probably no	Yes	Yes	No	Low	None
Cardwell et al, 2010[432]	Probably no	Probably yes	Yes	No	Moderate	Clinicians made diagnoses and were not masked to drug use; however, it is unlikely that this knowledge would have influenced the diagnosis, although it may have led to increased endoscopy surveillance in the exposed group.
Chiang et al, 2012[433]	Probably no	Yes	Yes	No	Low	None
Choi et al, 2020[434]	Probably no	Probably yes	Yes	No	Low	Outcome assessment masking not explicitly stated, but because claims were used likely not an issue.
Kim et al, 2021[435]	Yes	Yes	Yes	Probably no	Moderate	Clinicians diagnosed ONJ and could have been influenced by knowledge of treatment.
Lee et al, 2012[436]	Probably no	No information	Probably yes	Probably no	Low	None
Lee et al, 2019[307]	Probably yes	Probably yes	Yes	No	Moderate	AFF based on diagnostic codes; it is possible AFF was more likely to be diagnosed if clinicians and radiologists were aware of drug exposure given it has a known association.
Nordström et al, 2020[437]	Probably no	Probably yes	Yes	No	Moderate	Clinicians made diagnosis of ONJ and may have been more attuned to making this diagnosis for persons who take BP because it is a known adverse event.
Passarelli et al, 2013[438]	Probably no	Yes	Yes	No	Low	Diagnoses made by clinicians not masked, unlikely however that they would associate BP use with CRC diagnosis.
Pazianas et al, 2012[305]	Probably no	Probably yes	Yes	No	Low	Clinicians made diagnoses, and this drug would not be expected to be associated so it would be unlikely to have influenced the diagnosis of colon cancer.

Appendix D Table 64. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 6 Bias in Measurement of Outcomes)

Author, Year	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	6.2 Were outcome assessors aware of the intervention received by study participants?	6.3 Were the methods of outcome assessment comparable across intervention groups?	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Risk of Bias—Measurement of Outcomes	Support for Judgment
Rodriguez et al, 2020[439]	Probably no	Probably yes	Yes	Probably no	Low	Diagnoses were made by clinicians aware of drug exposure, but these outcomes are not well-known risks for the drug, so low risk.
Rubin et al, 2020[306]	Probably no	Probably yes	Yes	Probably no	Moderate	Clinicians assigned diagnoses and would have been aware of drug exposure, but this is likely not problematic for the outcomes in this study because they are not well-known for causing CVD adverse events.
Thadani et al, 2016[440]	No	No information	Yes	No	Moderate	No information about whether outcome assessment was masked.
Vestergaard et al, 2011[441]	Probably no	Probably yes	Yes	No	Low	Clinicians made diagnoses and were not blinded to drug use. However, this is unlikely to have influenced a cancer diagnosis.
Wang et al, 2016[442]	Probably no	Yes	Yes	Probably no	Moderate	Diagnoses made by clinicians who were aware of drug use, but these drugs are not traditionally associated with CVD outcomes so probably minimal bias. However, outcomes all based on claims data, not centrally adjudicated events as is typical in trials involving CVD outcomes.
Yang et al, 2018[443]	Probably yes	Yes	Yes	No	Low	Clinicians made diagnoses of atrial fibrillation, but this outcome is not well-known for being associated with drug exposure so probably low risk of bias.
Yuh et al, 2014[444]	Probably no	Probably yes	Yes	No	Moderate	Unclear whether outcome assessment masked, recognition of ONJ likely to occur more readily in persons with known BP use.

Abbreviations: AFF=atypical femur fracture; BP=bisphosphonate; CRC=colorectal cancer; CVD=cardiovascular disease; ONJ=osteonecrosis of the jaw.

Appendix D Table 65. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 7 Bias in the Selection of the Reported Result and Overall Risk of Bias)

Author, Year	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain?	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple analyses of the intervention-outcome relationship?	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different subgroups?	Risk of Bias — Selection of Reported Result	Overall Risk of Bias	Support of Overall Judgment
Abrahamsen et al, 2011[430]	No	No	No	Low	Serious (Poor quality)	Serious ROB from confounding, no adjustment for smoking or alcohol use, two very critical risks for upper GI cancer; moderate ROB due to deviations (alendronate users had higher rate of upper GI endoscopy); though this should bias results away from the null, but this was not observed
Black et al, 2020[431]	No	No	No	Low	Serious (Poor quality)	Serious ROB for selection bias from inclusion of prevalent users; moderate RoB for confounding, exposure classification, and missing data; low ROB for outcome measurement and selective outcome reporting
Cardwell et al, 2010[432]	No	No	No	Low	Serious (Poor quality)	Serious ROB for not being a new user cohort; moderate ROB due to confounding, missing data, and measurement of outcomes; however no critical flaws concerning for serious ROB
Chiang et al, 2012[433]	No	No	No	Low	Serious (Poor quality)	Serious ROB due to confounding
Choi et al, 2020[434]	No	No	No	Low	Serious (Poor quality)	Serious risk of selection bias from the way in which the analytic cohort was assembled. Moderate ROB from confounding and missing data.
Kim et al, 2021[435]	No	No	No	Low	Serious (Poor quality)	Serious risk of selection bias because of exclusion of person in the exposed group who died within 1 year or developed ONJ within 6 months of the index date; moderate ROB due to confounding, missing data, and outcome measurement

Appendix D Table 65. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 7 Bias in the Selection of the Reported Result and Overall Risk of Bias)

Author, Year	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain?	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple analyses of the intervention-outcome relationship?	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different subgroups?	Risk of Bias — Selection of Reported Result	Overall Risk of Bias	Support of Overall Judgment
Lee et al, 2012[436]	No	No	No	Low	Serious (Poor quality)	Did not adjust for important confounders relevant to development of cancer in various organ systems (e.g., smoking, alcohol use, hormone use, family history)
Lee et al, 2019[307]	No	No	No	Low	Moderate (Fair quality)	Moderate risk for bias because of confounding and measurement of outcome
Nordström et al, 2020[437]	No	No	No	Low	Serious (Poor quality)	Serious ROB from potential reverse causation arising from ambiguity in timing of exposure (unclear whether osteonecrosis was always a consequence of bisphosphonate exposure, could have been cause in some instances); moderate ROB from confounding, selection, and measurement of outcome
Passarelli et al, 2013[438]	No	No	No	Low	Serious (Poor quality)	Serious ROB from selection, moderate ROB from confounding and missing data
Pazianas et al, 2012[305]	No	No	No	Low	Moderate (Fair quality)	Moderate ROB from confounding, all measures based on claims/administrative data, also potential for residual confounding and no adjustment for smoking status, but that is probably less critical for colon cancer than other GI cancers
Rodriguez et al, 2020[439]	No	No	No	Low	Serious (Poor quality)	Seriuous ROB from confounding (did not adjust for smoking), moderate ROB from selection
Rubin et al, 2020[306]	No	No	No	Low	Moderate (Fair quality)	Moderate ROB for confounding and outcome measurement

Appendix D Table 65. Risk of Bias of Included Cohort Studies for Key Question 5 (Domain 7 Bias in the Selection of the Reported Result and Overall Risk of Bias)

Author, Year	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain?	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple analyses of the intervention-outcome relationship?	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different subgroups?	Risk of Bias — Selection of Reported Result	Overall Risk of Bias	Support of Overall Judgment
Thadani et al, 2016[440]	No	No	No	Low	Serious (Poor quality)	Serious risk of selection bias because did not use an inception cohort and did not exclude persons with known atrial fibrillation. Moderate ROB from missing data; had data overall for 91% but no information about differential missing data and events were somewhat rare; moderate ROB from outcome measurement.
Vestergaard et al, 2011[441]	No	No	No	Moderate	Serious (Poor quality)	Serious ROB from confounding, inadequate control for smoking and alcohol use, key covariates for upper GI cancers
Wang et al, 2016[442]	No	No	No	Low	Serious (Poor quality)	Serious ROB because of confounding; moderate ROB because of outcome measurement
Yang et al, 2018[443]	No	No	No	Low	Serious (Poor quality)	HIgh ROB for confounding (failure to control for confounding)
Yuh et al, 2014[444]	No	No	No	Low	Serious (Poor quality)	High ROB of selection because sample is not restricted to new users and classification of exposure; moderate ROB for confounding and outcome measurement

Abbreviations: GI=gastrointestinal; ONJ=osteonecrosis of the jaw; ROB=risk of bias.

E.1 Detailed Findings for Key Question 1

Detailed study characteristics are reported in **Appendix D Table 1**, and detailed findings are reported in **Appendix D Table 8**.

The ROSE Trial

The ROSE RCT randomly selected women ages 65 to 80 years living in southern Denmark to receive an invitation to participate in a two-step screening process (n=34,229).[126-128] Before recruitment, these women were randomized to either screening (n=17,072) with FRAX followed by DXA and vertebral fracture assessment (VFA) if 10-year FRAX MOF risk was greater than or equal to 15 percent or to a control group that continued to receive usual care as directed by their primary care provider (PCP), with no routine screening offered by the study (n=17,157).[126] Because study participants were identified through the Danish Civil Registration system, study authors applied no clinical exclusion criteria. Results of the DXA test in the screening group were sent to the participant and her general practitioner, which included recommendations based on national guidelines, while control group participants received no further followup. Screening guidelines at the time included a recommendation for measuring BMD if one or more clinical risk factors were present.[445] Treatment guidelines at the time of the study called for the initiation of treatment for 1) a fragility fracture of the hip or spine, or 2) T-score less than -2.5 with one clinical risk factor, or 3) T-score between -1.0 and -2.5 if on glucocorticoid therapy, or 4) if T-score less than 4.0 with no clinical risk factors.[445] Of participants randomized who returned the initial questionnaire with no missing data (N=20,905), the mean (SD) 10-year FRAX risk was 23.2 (11.0) percent for MOF and 10.0 (9.1) percent for hip. Further, 12.3 percent reported a history of a fragility fracture and 9.5 percent reported already being treated for osteoporosis. Of the women who completed a DXA scan in the screening group (N=5,064), the mean T-score was -1.2 (SD 1.0) at the TH and was -1.3 (SD 1.4) at the LS. Further, 3.7 percent had prevalent vertebral fractures.

We assessed the ROSE study as fair quality. Because the study was pragmatic in nature, the intervention was not blinded to participants. No missing data were reported because the analysis was intent-to-treat based on all participants randomized, although 45.6 percent of participants did not receive screening with FRAX (1,132 already on treatment, 2,894 returned questionnaire blank, 104 returned questionnaire with data missing to calculate FRAX, and the rest did not return the questionnaire). Significant differences were found between responders and nonresponders. In the intervention group, 12 percent (2,047/17,072 randomized) were high risk but did not receive a DXA (830 were not interested in a DXA and 1,217 dropped out; or 29% of those with high-risk FRAX scores [2,047/7,056]). Only 7 percent (1,236/17,072 randomized) had a positive DXA for osteoporosis after a FRAX risk above the study threshold, and only 6 percent (986/17,072, or 80% [986/1,236] of those with an indication) received treatment. The authors stated that 23 percent of the screening group received medication after the index date (mailing of questionnaire), which we assumed is the 986 participants started on medication and the 1,132 women who were already receiving medication on the baseline questionnaire along with an unknown number of women who were randomized to screening but who did not return the questionnaire but who may have been prescribed medication by their PCPs through the course of usual care outside of this study.

Similarly, in the control group, 45.6 percent (7,831/17,157 randomized) did not participate (1,168 were already on treatment, 3,143 returned a blank questionnaire, 111 returned a questionnaire with missing data to calculate FRAX, and the rest did not return the questionnaire). Additionally, there was contamination in the control group such that 25 percent of the control group received a DXA at some point after the study index date, possibly from increased awareness after completing the baseline questionnaire. The overall difference in the use of osteoporotic medications after the study index date was 5 percent (23% in the intervention vs. 18% in the control group), although it is unclear who was included in the denominators the authors used to report these percentages. The authors did not specify whether these data included those on treatment from the index date (mailing of the questionnaire). Outcome ascertainment was through the national health registry, and persons retrieving data from these health registries were not formally blinded to group allocation.[126, 127]

At a median followup of 5.0 years, the incidence of MOF for the intent-to-treat analysis (which was the primary study endpoint) was not significantly different in the invitation-to-screening group (9.9%) compared with the control group (10.0%) with an adjusted subhazard ratio (aSHR) of 0.986 (95% CI, 0.922 to 1.055).[126] The subhazard ratios (SHRs) for hip fracture and all osteoporotic fractures (excluding fingers, toes, skull, and face), both as unadjusted and aSHRs between groups, were also not significantly different. Mortality outcomes were not reported but used as competing outcomes in their SHRs and noted to be virtually complete because the national health registries were used.

Given the potential challenges with the study design (e.g., participants were randomized before giving consent), the authors prespecified a per-protocol analysis to examine fracture outcomes between the screening and control groups with completed FRAX calculations and not in current osteoporotic treatment. The per-protocol incidence of MOF was 725/9,279 (7.8%) in the completed FRAX screening group compared with 786/9,326 (8.4%) in the completed FRAX control group with an aSHR of 0.914 (95% CI, 0.827, 1.011). The per-protocol incidence of hip fracture was 169/9,279 (1.8%) in the completed FRAX screening group compared with 202/9,326 (2.2%) in the control group with an aSHR of 0.82 (95% CI, 0.670 to 1.007), p=0.059.[126] The per-protocol incidence of all fractures was 996 (10.7%) in the completed FRAX screening group compared with 1,025 (11.0%) in the control group with an aSHR of 0.968 (95% CI, 0.887 to 1.056).[126]

In a second, post hoc, per-protocol analysis comparing persons with high-risk FRAX who were DXA scanned with high-risk controls, the aSHR for hip fracture was 0.741 (95% CI, 0.578 to 0.950). However, this per-protocol analysis should be interpreted with caution because the women in the DXA-scanned group showed significant differences in baseline characteristics compared with the high-risk controls (e.g., they were younger, had higher rates of previous fractures, and were less likely to smoke), although some of these differences were of uncertain clinical significance. In another analysis with the second per-protocol population, authors excluded hip fractures from the MOF outcome and the MOF results became nonsignificant, suggesting that most of the differences observed for MOF were being driven by differences in hip fractures.

The SCOOP Trial

The SCOOP RCT randomly selected women ages 70 to 85 years from 100 general practices in England and randomized them to either screening (n=6,233) with a FRAX assessment and invitation to DXA if risk was greater than or equal to an age-based threshold or to routine care as directed by the participant's PCP (n=6,250).[120] Participants were excluded if they were on treatment for osteoporosis (other than calcium and vitamin D) or had known comorbidity or another factor that might make participation inappropriate (e.g., advanced cancer or recent bereavement).[121] Of participants randomized who returned the initial questionnaire with no missing data (N=12,483), the mean (SD) 10-year FRAX risk was 19.3 (8.9) percent for MOF and 8.5 (7.4) percent for hip, and 23 percent had a history of a broken bone since age 50 years. Among those who completed DXA (2,817/6,233 randomized), the mean T-score was -2.6 at the FN. Although the two randomized groups were similar in baseline demographic characteristics, those who participated in the study had higher education, higher socioeconomic status, and more frequent history of previous fractures or parental hip fracture than nonparticipants.[120]

We assessed the SCOOP study as fair quality. Because the study was pragmatic in nature, the intervention was not blinded to participants. There was minimal missing data because the analysis was conducted on an intent-to-treat basis. Of the 49 percent of women in the screening group deemed initially high risk based on FRAX hip fracture risk, 4 percent (247/6,233 randomized) were high risk and did not have a DXA (157 declined, 81 were unable to have hip BMD measured, and 9 died); 45 percent (2,817/6,233 randomized) were high risk and had a DXA. In the screening group, 14 percent (898/6,233 randomized) had a high-risk FRAX after recalculation with the FN BMD. In the screening group, over the course of the study, 24 percent (1,486/6,233 randomized) received at least one prescription for treatment, with 15 percent (953/6,233 randomized) having at least one prescription for treatment in the first 12 months. Of the high-risk screening group (703/898), 78 percent had at least one prescription in the first 6 months.[123] In the control group, over the course of the study, 16 percent (982/6,250 randomized) received at least one prescription for treatment, suggesting evidence of contamination.[120] Outcome ascertainment was verified with medical records, and assessors were blinded to study group assignment.[121]

At 5 years followup, the incidence of fractures excluding hands, feet, nose, skull, and cervical vertebrae and without regard to trauma (the study's primary endpoint) for the intent-to-treat analysis was not significantly different in the invitation-to-screening group (12.9%) compared with the control group (13.6%) with an adjusted hazard ratio (aHR) of 0.94 (95% CI, 0.85 to 1.03).[126]Authors reported several prespecified secondary endpoints. The aHR for any clinical fracture (not excluding any site) was not significant, but the incidence of hip fracture was significantly lower in the screening group (2.6%) compared with the control group (3.5%; aHR: 0.72, 0.59 to 0.89).[446] All-cause mortality was not significantly different between groups.

In a post hoc analysis evaluating the association between baseline 10-year FRAX hip risk without BMD risk and fracture incidence, there were no significant differences between the screening group and the control group at the 10th, 25th, and 50th percentile of 10-year FRAX hip risk (2.6%, 3.8%, and 6.3%, respectively) for any clinical fracture (with or without selected sites excluded).[122] There were also no significant differences in any clinical fractures (with or without selected sites excluded) at the 75th and 90th percentiles of 10-year FRAX hip risk (10.5% and

16.8%, respectively), but there were significant differences for hip fracture incidence and when considering FRAX risk as a continuous measure, a significant interaction was observed for the association between FRAX score and hip fracture but not for any clinical fracture (with or without selected sites excluded.[122]

The SOS Trial

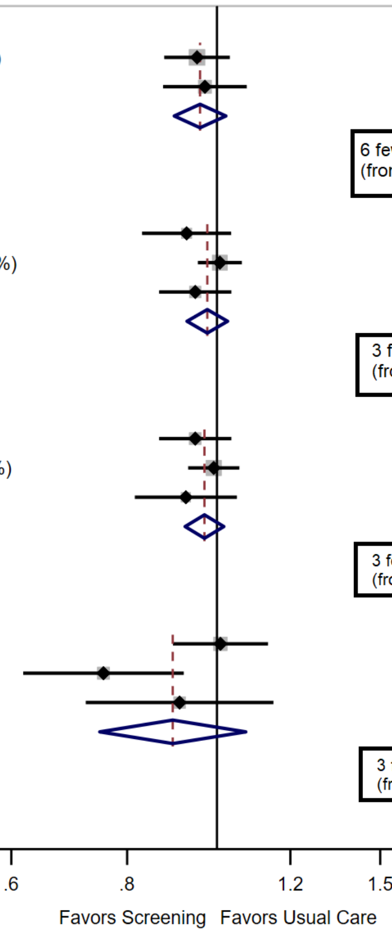
The SOS RCT randomly assigned women ages 65 to 90 years from general practice registries in the Netherlands who had one or more clinical risk factors for osteoporosis and completed baseline information (N=11,032).[124, 125] Participants were excluded if they were on treatment for osteoporosis currently or in the preceding 5 years or took prednisone. Participants assigned to the screening group (n=5,575) received a multicomponent screening intervention (FRAX [without BMD], DXA, VFA, falls risk assessment, and blood chemistries to exclude secondary osteoporosis), while those assigned to the control group (n=5,457) received routine care as directed by their PCP. The mean (SD) 10-year FRAX risk of participants was approximately 24 percent (10) for MOF and 11 percent (10) for hip, and 43 percent reported a fracture after age 50 years.

We assessed the SOS study as fair quality. Because the study was pragmatic in nature, the intervention was not blinded to participants. There were little missing data because authors used an intention-to-treat analysis. Twenty-four percent of participants invited to screening (1,347/5,575 randomized) did not participate.[124] Twenty-five percent randomized to screening (1,417/5,575 randomized) had an indication for treatment, but 31 percent of those did not start treatment.[124] In the screening group, 21 percent (1,154/5,575 randomized) received treatment over the course of the study, with 18 percent (982/5,575 randomized) reporting starting treatment and 12 percent (657/5,575 randomized) reporting still being on treatment at 36 months.[124] In the control group, 6 percent (316/5,457 randomized) received a DXA/VFA over the course of the study; 2 percent (112/5,457 randomized) received DXA/VFA within 3 months of randomization. About 5 percent (291/5,457 randomized) of the control group received treatment over the course of the study—3 percent (167/5,457 randomized) by 18 months.[124] Outcome ascertainment was blinded, and fractures were confirmed with medical records.[125]

Over a mean followup of 3.7 years, no statistically significant differences were found on the primary outcome of time to first incident fracture of any type. In total, 626 (11.3%) persons in the intervention group had a fracture vs. 632 (11.7%) in the control group (aHR, 0.97 [95% CI 0.87 to 1.08]).[124] Additionally, no statistically significant differences were found on any secondary fracture measures or mortality.[124]Authors also reported no significant interaction effects with age, history of prior fracture, or recency of prior fracture for the outcome of “all fractures.” However, there was a significant interaction with recency of prior fracture (within 2 years of baseline) for MOF and hip fracture, although these analyses were post hoc.[124]

Appendix E.1 Figure 1. Randomized, Controlled Trials of Screening vs. Usual Care: Fracture Outcomes (KQ 1)—Sensitivity Analysis Using the ROSE Intention-to-Treat Sample

Fracture and Study	Age	PriorFx	Mean FRAX MOF/Hip Risk	Length of F/U	Screened n with Events/Total N(%)	Control n with Events/Total N(%)	Risk Ratio (95% CI)
All							
SCOOP	76	22%	19.3/85	5 y	951/6,233 (15.3%)	1,002/6,250 (16.0%)	0.95 (0.88, 1.03)
SOS	75	43%	24.5/11.5	3.7 y	626/5,516 (11.3%)	632/5,405 (11.7%)	0.97 (0.87, 1.08)
Subgroup, DL							0.96 (0.90, 1.02)
(I^2 = 0.0%, p = 0.771)							6 fewer per 1,000 (from 14 fewer to 3 more)
Osteoporotic							
SOS	75	43%	24.5/11.5	3.7 y	547/5,516 (9.9%)	578/5,405 (10.7%)	0.93 (0.83, 1.04)
ROSE	71	12%	20/6.7	5 y	2,238/17,072 (13.1%)	2,233 /17,157 (13.0%)	1.01 (0.95, 1.06)
SCOOP*	76	22%	19.3/85	5 y	805/6233 (12.9%)	852/6250 (13.6%)	0.95 (0.87, 1.04)
Subgroup, DL							0.98 (0.93, 1.03)
(I^2 = 18.9%, p = 0.291)							3 fewer per 1,000 (from 9 fewer to 4 more)
MOF							
SCOOP*	76	22%	19.3/85	5 y	805/6233 (12.9%)	852/6250 (13.6%)	0.95 (0.87, 1.04)
ROSE	71	12%	20/6.7	5 y	1,697/17,072 (9.9%)	1,719/17,157 (10.0%)	0.99 (0.93, 1.06)
SOS	75	43%	24.5/11.5	3.7 y	427/5,516 (7.7%)	452/5,405 (8.3%)	0.93 (0.82, 1.05)
Subgroup, DL							0.97 (0.92, 1.02)
(I^2 = 0.0%, p = 0.530)							3 fewer per 1,000 (from 8 fewer to 2 more)
Hip							
ROSE	71	12%	20/6.7	5 y	534/17,072 (3.1%)	532/17,157 (3.1%)	1.01 (0.90, 1.14)
SCOOP	76	22%	19.3/85	5 y	164/6,233 (2.6%)	218/6,250 (3.5%)	0.75 (0.62, 0.92)
SOS	75	43%	24.5/11.5	3.7 y	133/5,516 (2.4%)	143/5,405 (2.6%)	0.91 (0.72, 1.15)
Subgroup, DL							0.90 (0.75, 1.07)
(I^2 = 67.1%, p = 0.048)							3 fewer per 1,000 (from 8 fewer to 2 more)
Heterogeneity between groups: p = 0.827							



* SCOOP reported an outcome entitled "osteoporotic fractures," which were defined as clinical fractures excluding hand, foot, skull, and cervical vertebrae. It is not entirely clear how this definition differs from the definition of MOF used by the other two studies (hip, clinical vertebral, distal forearm, and humerus); as such, we have included SCOOP "osteoporosis" outcome in the estimate for both "osteoporotic fractures" and for "MOF."

Appendix E.1 Figure 1. Randomized, Controlled Trials of Screening vs. Usual Care: Fracture Outcomes (KQ 1)—Sensitivity Analysis Using the ROSE Intention-to-Treat Sample

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FRAX=Fracture Risk Assessment Tool; F/U=followup;Fx=fracture; KQ=key question; MOF=major osteoporotic fracture; N/n=number; ROSE=Risk-stratified Osteoporosis Strategy Evaluation; RR=risk ratio; SCOOP=Screening in the Community to Reduce Fractures in Older Women; SOS=Stichting Artsen Laboratorium enTrombosedienst (SALT) Osteoporosis Study; vs.=versus.

E.2 Detailed Calibration Outcomes (Key Question 2b)

Appendix E.2 Table 1. Calibration Outcomes From 12 Unique Cohorts Reported for the Accuracy of Bone Mineral Density to Predict Fracture (KQ 2b)

Sex and Fracture Type	Gradient of Risk HR (95% CI) per SD decrease in FN BMD	Observed/Expected Ratio*	Calibration Plots	Hosmer-Lemeshow Goodness of Fit†
Men				
MOF	NR	NR	NR	p=0.1672[146]
Hip	2.30 (1.89 to 2.82)[15]	NR	NR	p=0.2655[146]
Major nonvertebral	2.31 (1.79 to 3.00)[182]	NR	NR	NR
Nonvertebral	1.37 (1.25 to 1.49)[15]	NR	NR	NR
Women				
MOF	1.68 (1.58 to 1.78)[183] 1.97 (1.91 to 2.03)[160] 1.78 (1.43 to 2.2)[190] 1.94 (1.81 to 2.08)[172]	NR	Dose-response observed in a plot by quartile of predicted risk; no other statistics reported[150]	
Hip	2.60 (2.23 to 3.03)[183] 2.99 (2.84 to 3.15)[160] 2.47 (NR)[185] 1.97 (1.76 to 2.21)[15] 3.82 (3.17 to 4.61)[172] 2.0 (1.9 to 2.1)[192] 3.22 (2.47 to 4.20)[194]	Across all quintiles of risk: 0.83 (95% CI, 0.65 to 1.04)[185]	Dose-response observed in a plot by quartile of predicted risk; no other statistics reported[150]	p=0.015[185]
Fragility	2.11 (1.62 to 2.73)[186]	NR	NR	NR
Hip or MOF	1.70 (1.35 to 2.14)[189]	NR	NR	NR
Nonvertebral	1.42 (1.35 to 1.50)[15] 1.5 (1.4 to 1.5)[192]	NR	NR	NR
Mixed Population				
MOF	1.56 (1.42 to 1.71)[154] 1.58 (1.50 to 1.66)[157]	NR	NR	NR
Hip	1.96 (1.62 to 2.37)[154] 2.19 (1.97 to 2.43)[157]	NR	NR	NR

* Ratios close to 1 indicate good agreement between observed and predicted.

† P values <0.05 indicate poor fit.

Abbreviations: BMD=bone mineral density; CI=confidence interval; FN=femoral neck; HR=hazard ratio; KQ=key question; MOF=major osteoporotic fracture; NR=not reported; SD=standard deviation.

E.3 Detailed Results for Diagnostic Accuracy (KQ 2c)

Age, Body Size, No Estrogen

Study Characteristics

Five fair-quality studies (total N=4,203 participants) reported on the accuracy of Age, Body Size, No Estrogen (ABONE);[206, 207, 233, 242, 245] three studies were new to this update.[233, 242, 245] Three studies were conducted in Asian countries,[207, 233, 245] one was conducted in Greece,[242] and one was conducted in Canada.[206] One study included men;[233] the rest were conducted exclusively among women. The mean age across studies ranged from 62 to 68 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 4.0 percent to 24.4 percent across four of the included studies; the prevalence was not reported in one study.[242]

Findings

The reported AUC across cohorts ranged from 0.62 to 0.78 using a reference BMD measurement at the FN, LS, or both. The exception to this range was one cohort[245] that was an outlier with respect to values reported for four different risk assessment tools, suggesting something unique about the underlying study population or study approach.

The most common score threshold reported was a score greater than or equal to 2, which was reported by four cohorts.[206, 207, 233, 245] Other included studies reported using a score threshold of greater than 1.5[242] or greater than or equal to 3.[207] The sensitivities ranged from 66 percent to 100 percent, and the specificities ranged from 16.7 percent to 60 percent, excluding the outlier study.[245]

One study reported findings separately for men vs. women.[233] The AUC was 0.78 (95% CI, 0.64 to 0.93) in men and 0.70 (95% CI, 0.61 to 0.77) in women. The sensitivity was the same among men and women (100%); the specificity was 28 percent among men and 10 percent among women.[233]

Age, Menopause, Menarche, BMI

Study Characteristics

Two fair-quality studies (total N=1,520 participants) reported on the accuracy of Age, Menopause, Menarche, BMI (AMMEB);[199, 225] neither was new to this update. Both studies were conducted in Italy, and both were exclusively conducted among postmenopausal women. The mean age of participants in one study[199] was 65 years and in the other study[225] ranged from 57 (normal BMD), 60 (osteopenia), to 62 years (osteoporosis) depending on BMD status. The prevalence of osteoporosis as measured by DXA BMD T-scores less than -2.5 at the FN or LS were 33.7 percent in the study[199] enrolling postmenopausal women from general practices (race/ethnicity NR) and 47.4 percent in the study[225] enrolling Caucasian women referred to a university bone metabolic unit for DXA, of which 13 percent were noted to have secondary osteoporosis.

Findings

The reported AUCs were 0.63[199] and 0.71,[225] both using reference BMD measurements at the FN or LS. Neither study reported sensitivity or specificity.

Fracture Risk Assessment Tool

Study Characteristics

Fifteen fair-quality studies reporting on 12 unique cohorts[141, 143, 195, 196, 200, 220, 232-239, 241] (total N=37,756 participants) reported on the accuracy of FRAX. Ten articles were new to this update.[141, 143, 233-239, 241] One study was conducted in Canada,[200] one study was conducted in Taiwan,[233] and two studies were conducted in Australia.[196, 239] The rest of the studies were conducted in the United States. Of studies in the United States, most had a high percentage of White participants, with all but two reporting greater than 85 percent White participants.

Four studies included both men and women, with three including 44 to 65 percent male participants;[196, 233, 239] another study[241] included only 13 percent male participants. Four studies included exclusively men.[232, 235, 237, 238] The other studies included exclusively women.[141, 195, 200, 220, 234, 236] The mean age across studies ranged from 57 years to 80 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 4.5 percent to 25.9 percent. One study had an outlying prevalence of osteoporosis of 51.1 percent; this study had a small sample size (N=45) from a rural area in the United States, included those at increased risk for osteoporosis (e.g., on chronic steroids), and did not report what site or T-score reference range was used to define osteoporosis.[241]

Findings

All but three studies[141, 195, 238, 241] reported AUCs based on FRAX MOF risk, FRAX hip fracture risk, or both. We included only FRAX risk estimates calculated without the use of BMD because BMD is the reference test for this KQ. Over all studies reporting AUCs, the AUCs ranged from 0.55 to 0.86 using a reference BMD measurement at the FN only or at the lowest site from among the TH, LS, or FN. One study[239] also considered BMD measured at the forearm in addition to the three usual sites. Two studies[236, 241] did not report the site for the reference BMD measurement used. When limited to AUCs based on FRAX MOF risk only, AUCs ranged from 0.55 to 0.79.[143, 195, 196, 200, 220, 232-237, 239] When limited to AUCs based on FRAX hip fracture only, AUCs ranged from 0.70 to 0.86.[196, 233, 239] In the one study that reported AUCs based on either MOF or hip fracture risk, the AUC was 0.65 (95% CI, 0.59 to 0.71) when calculated based on patient characteristics derived from electronic health record data and was 0.72 (95% CI, 0.67 to 0.78) when based on data collected directly from participants.[237] Four studies conducted exclusively in men or that included results separately for men reported AUCs, and these results ranged from 0.62 to 0.86.[232, 233, 235, 237]

All but two studies[200, 239] reported sensitivity and specificity or provided data for us to calculate these estimates. For determining sensitivity and specificity, the thresholds authors used varied by type of fracture risk (MOF vs. hip vs. both) and numeric value used. Some studies reported on multiple thresholds using the same data.

Authors used an MOF risk greater than or equal to 9.3 percent (the threshold suggested for use by the USPSTF's 2011 recommendation) in six studies.[195, 220, 232, 234-236] The sensitivity for this threshold among the four studies conducted exclusively in women ranged from 24 to 37 percent and the specificity ranged from 73 to 86 percent.[195, 220, 234, 236] Within the two studies conducted exclusively in men, the sensitivity was 39 percent[232] and 59 percent,[235] and the specificity was 89 percent[232] and 59 percent.[235] An MOF risk greater than or equal to 8.4 percent (suggested by the 2018 USPSTF recommendation) was used in one study conducted exclusively in women and was reported for a variety of age ranges.[141] The sensitivity was 5.2 percent for women ages 50 to 54 years, 16.9 percent for women ages 55 to 59 years, and 48.5 percent for women ages 60 to 64 years. The specificity ranged from 95.8 percent for the youngest age group to 63.4 percent for the oldest age group.[141]

An MOF risk greater than or equal to 20 percent (a commonly used threshold for initiating treatment) was reported in one study.[233] The sensitivity for this threshold was 0 percent for men and 17 percent for women, and the specificity for this threshold was 99 percent for men and 96 percent for women.

Accuracy was also reported for MOF risk thresholds between 6.5 and 10 percent in three studies.[196, 235, 238] Sensitivity ranged from 53 percent to 90 percent, and specificity ranged from 32 percent to 65 percent in these studies.

Two studies used a hip fracture risk threshold of 3 percent or greater (a commonly used threshold for initiating treatment).[196, 233] The sensitivity for this threshold ranged from 80 percent to 92 percent, and the specificity ranged from 37 percent to 71 percent.

Three studies defined a positive screening test based on having either a hip fracture risk greater than 3 percent or having an MOF risk greater than 20 percent (both commonly used thresholds for initiating treatment).[237, 238, 241] In the studies conducted exclusively in men, sensitivity was 27 percent[238] and 69 percent,[237] and the specificity was 88 percent[238] and 54 percent.[237] In the study conducted predominantly in women (87%), the sensitivity was 100 percent, and the specificity was 91 percent.[241]

One study conducted exclusively in men also evaluated other approaches based on either MOF or hip fracture risk (**Appendix D Table 14**).[238]

Garvan Fracture Risk Calculator: Hip and MOF Risk

Study Characteristics

Two fair-quality studies (total N=1,084 participants) reported the accuracy of the Garvan Fracture Risk Calculator;[233, 239] both studies were new to this update. One study was conducted in Taiwan,[233] and the other was conducted in Australia.[239] Both studies included men and women: 55.8 percent of participants across both studies were women. The mean age in one study[233] was 67.4 years, and the mean age in the other study[239] was 78 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 was 17.5 percent in one study[233] and was 24.5 percent in the other study, which had the higher mean age.[239]

Findings

The reported AUCs for the Garvan Fracture risk ranged from 0.72 to 0.80 for hip fracture and 0.71 to 0.75 for any osteoporotic fracture or MOF using a reference BMD measurement at the FN in one study[233] or the lowest BMD at the FN, TH, LS, or forearm in the other study.[239]

For determining sensitivity and specificity, one study reported score thresholds of greater than or equal to 3 percent for hip fracture risk and greater than or equal to 20 percent for any MOF.[233] The other study used an empirically derived, age-stratified risk threshold for both hip and MOF risk.[239] The sensitivities ranged from 20 percent to 72 percent, and the specificities ranged from 73 percent to 96 percent across these risk thresholds.

One study reported findings separately for men vs. women for both hip fracture risk and MOF risk.[233] The AUC for hip fracture risk was 0.72 (95% CI, 0.44 to 1.0) in men and 0.80 (95% CI, 0.73 to 0.88) in women. The AUC for MOF risk was 0.72 (95% CI, 0.46 to 0.98) in men and 0.75 (95% CI, 0.66 to 0.85) in women. The sensitivity for hip fracture risk was 60 percent among men and 28 percent among women; the specificity was 79 percent among men and 95 percent among women. The sensitivity for MOF risk was 20 percent among men and 55 percent among women; the specificity was 96 percent among men and 73 percent among women.

Male Osteoporosis Risk Estimation Score

Study Characteristics

Two fair-quality studies and 1 good-quality study (total N= 4,788 participants) reported on the accuracy of the Male Osteoporosis Risk Estimation Score (MORES)[197, 202, 232]; none of the studies were new to this update. All three studies were conducted in the United States, and all studies were conducted exclusively among predominantly White men (76% to 81% of participants). The mean age for the subjects varied from 63 to 70 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 4.3 percent[197] and 4.5 percent[232] in two studies, but was 10 percent in one study.[202] The study with the highest prevalence used the BMD reference values from White men ages 20 to 29 years to generate T-scores from, compared to the use of values from NHANES III for young, non-Hispanic women in the other two studies.

Findings

The reported AUCs ranged from 0.66 to 0.87 using reference BMD measurements at the FN, TH, or LS.[197, 202, 232] One study also evaluated efficacy when reference BMD was any site (thoracic vertebra, LS, arms, ribs, pelvis, legs), and the reported AUC was 0.73.[202]

For determining sensitivity and specificity, all studies reported results based on a score threshold of greater than or equal to 6. The sensitivities ranged from 58 to 96 percent, and the specificities ranged from 61 to 70 percent.

One study[202] reported on the sensitivity and specificity of identifying osteoporosis at the LS stratified by age bands of 5 years starting at age 50 years through age 89 years. Sensitivity was highest in the age group of 80 to 84 years (8%) and lowest in the age group of 50 to 54 years (29%).[202] Specificity was highest in the age group of 50 to 54 years (90%) and lowest in the age group of 85 to 89 years (23%).[202] This study also reported on the sensitivity and specificity for identifying osteoporosis at the LS stratified by race/ethnicity. Sensitivity was lowest in White participants (51% [95% CI, 38% to 64%]) and highest in participants of other (i.e., not African American or Mexican American) ethnicity (90% [95% CI, 66% to 98%]). Specificity was lowest in participants of other ethnicities (50% [95% CI, 40% to 60%]) and highest in White patients (67% [95% CI, 65% to 70%]).[202]

Male Osteoporosis Screening Tool

Study Characteristics

One fair-quality study (total N=4,658 participants) reported on the accuracy of the Male Osteoporosis Screening Tool (MOST);[218] it was not new to this update. This study was conducted among men in the United States and Hong Kong from the MrOs cohort study. The mean age of enrolled participants was not reported, but only men age 65 years or older were enrolled in the MrOs cohort study. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 at the FN or at the LS was 5.0 percent among U.S. participants.[218]

Findings

Data were analyzed separately for participants in the United States and Hong Kong. BMD reference measurements were reported for both FN alone and for the lowest T-score from either the FN or LS or TH.[218] The reported AUCs for U.S. participants were 0.799 (FN or LS or TH) and 0.807 (FN only) and for Hong Kong participants were 0.831 (FN or LS or TH) and 0.876 (FN only).[218]

For determining sensitivity and specificity, the data for participants from the United States were reported based on a score threshold of less than or equal to 26. The sensitivity was 89 percent, and the specificity was 50 percent based on lowest site BMD. The data for participants from Hong Kong were reported based on a score threshold of less than or equal to 21. The sensitivity was 87 percent, and the specificity was 59 percent based on lowest site BMD.

Male Simple Calculated Osteoporosis Risk Estimation

Study Characteristics

One fair-quality study (total n=197 participants) reported on the accuracy of the Male Simple Calculated Osteoporosis Risk Estimation (MSCORE);[216] it was not new to this update. This study was conducted exclusively in the United States; all participants were men age 40 years or older (94% Caucasian) enrolled from Veterans Affairs general medical or specialty clinic sites. The mean age for participants was 68 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 at the FN was 11.2 percent. This study also reported on a separate cohort of 134 African American men (mean age, 61 years) comprising a convenience sample recruited separately from the original development and validation cohorts.

Findings

The reported AUC for MSCORE was 0.84 (95% CI, 0.74 to 0.95) using a reference BMD measurement at the FN. For determining sensitivity and specificity, the study reported score thresholds of greater than or equal to 9. The sensitivity was 88 percent, and the specificity was 57 percent. In the separate African American convenience sample, the sensitivity was 93 and 100 percent depending on whether a Caucasian or African American BMD reference range was used to calculate the T-score, respectively. Similarly, the specificity was 73 or 79 percent.

National Osteoporosis Foundation Risk Score

Study Characteristics

Four fair-quality studies (total N=4,087 participants) reported on the accuracy of the National Osteoporosis Foundation (NOF) risk score.[199, 206, 225, 226] No new studies were included in this update. Two studies were conducted in Italy,[199, 225] one was conducted in Canada,[206] and one was conducted in the United States.[226] Participants from the studies in Canada and the United States were recruited from the general population,[206, 226] while the participants in the Italian studies were from either general practice clinics[199] or referred to an osteoporosis clinic from general practice clinics or gynecologists.[225] All studies included only postmenopausal women. The mean age across studies ranged from 60.5 to 69 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 at the FN only or at the FN or LS ranged from 10 percent to 47 percent across studies.

Findings

The reported AUCs across studies ranged from 0.60 to 0.70 using a reference BMD measurement at the FN only or measurement at the FN and LS. All studies used a threshold score of 1 or more. Only two studies reported sensitivity and specificity.[206, 226] The sensitivities were 96 percent and 100 percent, and the specificities were 10 percent and 18 percent in those studies.[206, 226]

One study reported findings separately for different age groups.[226] The group ages 45 to 64 years had an AUC of 0.69; sensitivity was 100 percent and specificity was 19 percent. The group age 65 years or older had an AUC of 0.60; sensitivity was 100 percent and specificity was 0 percent.

Osteoporosis Risk Assessment Instrument

Study Characteristics

Twenty-two publications[199, 203, 206-210, 214, 217, 223-228, 230, 231, 233, 234, 236, 242, 245] covering 21 unique cohorts (total N=24,427 participants) reported on the accuracy of the Osteoporosis Risk Assessment Instrument (ORAI) for 28 comparisons. We rated all analyses as fair quality. Five new studies were included in this update.[233, 234, 236, 242, 245] Six studies were conducted in the United States,[224, 226, 230, 231, 234, 236] five in Commonwealth countries,[206, 208, 217, 223, 227] seven in Europe,[199, 203, 208-210, 214, 217, 225, 228, 242] and three studies in Asia.[207, 233, 245] Twelve studies included only perimenopausal or postmenopausal women;[199, 203, 207, 208, 214, 224-226, 228, 236, 242, 245] the one study that included both men and women was 66 percent women.[233] Five studies included participants referred for BMD testing or to an osteoporosis-related clinic.[208, 210, 214, 217, 242] Three studies recruited participants from the general population,[206, 226, 233] and one study recruited participants from both primary care and specialty clinics.[203] The mean age across studies ranged from 50.5 to 70.5 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 (at one or at least one site if multiple sites measures) had a wide range from 4.6 percent to 47.4 percent, although half were between 10 percent and 30 percent,[203, 206-208, 214, 223, 224, 230, 231, 233, 234, 245] and one study did not report osteoporosis prevalence.[242]

Findings

Twenty publications covering 19 unique cohorts reported AUC for women.[199, 203, 206-210, 214, 217, 223-226, 228, 230, 233, 234, 236, 242, 245] These AUCs ranged from 0.32 to 0.84 excluding one extreme outlier.[245] In this study, only 15 percent of its 150 participants with osteoporosis reported an AUC of 0.047 and 0.129.[245] In another study of 525 participants, half of whom had osteoporosis, authors reported an AUC of 0.32.[225] The AUCs in the rest of the studies ranged from 0.60 to 0.84. [227, 231]

For determining sensitivity and specificity, studies used a variety of thresholds ranging from greater than or equal to 8 up to greater than 15. Two studies examined a range of thresholds: greater than 9, 16, and 20[207] and greater than 8 and 13,[209] respectively. Twelve studies used a threshold greater than or equal to 9.[203, 206, 207, 214, 224, 226, 230, 231, 234, 236, 245] Five studies used a threshold greater than or equal to 8.[199, 209, 210, 223, 225] One study did not report a threshold nor did it report sensitivity or specificity.[217] Two other studies did not report sensitivity or specificity either.[225, 447] The sensitivities ranged from 52 percent to 100 percent for studies restricted to those using a threshold of 8 or 9[199, 203, 206, 207, 209, 210, 214, 223-226, 230, 231, 234, 236, 245] and from 43 percent to 89 percent for the remainder of thresholds.[207-209, 217, 227, 228, 242] The specificities ranged from 5 percent to 100 percent (restricted to studies using thresholds of 8 or 9) and from 44.7 percent to 86 percent (for the remainder).

Five studies reported results in women younger than age 65 years; however, these studies did not use the same score threshold, which may partially explain the variation in results for sensitivity and specificity.[209, 226, 228, 234, 236] The AUCs ranged from 0.60 to 0.84, the sensitivities ranged from 44 percent to 99 percent, and the specificities ranged from 36 percent to 77 percent.[209, 226, 228, 234, 236]

The one study that included men reported an AUC of 0.87 for men and, using a score threshold of greater than or equal to 9, reported a sensitivity of 100 percent and a specificity of 19 percent.[233]

Osteoporosis Index of Risk

Study Characteristics

Seven fair-quality studies (total N=7,173 participants) reported on the accuracy of the Osteoporosis Index of Risk (OSIRIS).[203, 208, 210, 214, 217, 233, 242] Two new studies were included in this update.[233, 242] Six studies were conducted in Europe; two in the United Kingdom,[199, 225] two in Spain,[203, 214] one in Belgium,[210] and one in Greece;[242] one study was conducted in Taiwan.[448] All studies except one included only postmenopausal women; the one study that included both men and women was 66 percent women.[233] Five studies included participants referred for BMD testing or to an osteoporosis-related clinic.[208, 210, 214, 217, 242] One study recruited participants from the population,[233] and one study recruited participants from both primary care and specialty clinics.[203] The mean age across studies ranged from 54 to 67 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 (at one or at least one site if multiple sites measured) ranged from 18 percent to 34 percent across six of the seven included studies; one study did not report osteoporosis prevalence.[242]

Findings

The reported AUCs for women across studies ranged from 0.63 to 0.83. For determining sensitivity and specificity, most studies used a threshold of around 1; three studies used a threshold of less than or equal to 1,[210, 214, 233] one study used a threshold of less than 1,[210] and one study reported using a threshold of less than 0.5 and less than 1.5.[242] As outliers, one study used a threshold of less than 0,[208] and one used a threshold of less than or equal to -3.[203] One study did not report a threshold nor did it report sensitivity or specificity.[217] The sensitivities ranged from 58 percent to 100 percent (both restricted to studies using thresholds around 1 and unrestricted), and the specificities ranged from 6 percent to 69 percent (restricted to studies using thresholds around 1).[210, 214, 233, 242] This level of variation may be due to the underlying population or sites of reference BMD measurement.

The one study that included men reported an AUC of 0.94 for men and, using a score threshold of less than or equal to 1, reported a sensitivity of 100 percent and a specificity of 29 percent.[233]

Osteoporosis Self-Assessment Tool

Study Characteristics

Thirty studies reported on the accuracy of Osteoporosis Self-Assessment Tool (OST).[159, 195, 196, 198-201, 203, 205, 208-210, 214, 216-219, 221, 223, 225, 228, 230, 231, 234-237, 240, 242, 243] One study was good quality;[205] the rest were fair quality. Seven studies were new to this update.[234-237, 240, 242, 243] Eleven studies were conducted in the United States,[195, 198, 216, 219, 221, 230, 231, 234-237] and the rest were conducted in Canada, Australia, or various European or Asian countries. Nine studies were conducted exclusively in men,[198, 201, 216, 218, 219, 221, 235, 237, 240] one study was conducted in men and women and reported results by sex,[243] one study was conducted in men and women but did not report results separately,[196] and the 19 remaining studies were conducted exclusively in women.[159, 195, 199, 200, 203, 205, 208-210, 214, 217, 223, 225, 228, 230, 231, 234, 236, 242] The mean age across studies ranged from 51 years to 80 years; however, more than two-thirds had a mean age of 60 years or older. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 4.6 percent to 47.4 percent; the prevalence was not reported in one study.[242] The reference standard used to determine the presence of osteoporosis varied across studies: some used a single measurement at only one anatomical site, typically the FN or LS, while others used the lowest T-score from either the FN or LS or the FN, LS, or TH.

Findings

Across the 29 studies reporting AUCs, estimates ranged from 0.32 to 0.89. Nineteen (63%) studies reported an AUC of 0.70 or higher.[159, 195, 196, 200, 203, 205, 208-210, 216, 218, 219, 221, 223, 230, 236, 237, 240, 243]

All but four studies reported sensitivity and specificity.[199, 200, 217, 225] Of those studies reporting sensitivity and specificity, authors used different score thresholds, and some studies reported accuracy results for more than one threshold. The most common threshold used was a score less than 2, reported by 14 studies.[195, 201, 205, 210, 214, 216, 219, 221, 223, 228, 234-236, 243] Less than 2 is the threshold specified in the first use of the tool in a U.S. population and was selected based on it resulting in an approximately 90 percent sensitivity.[231] This is the same approach used by the original developers of the tool to establish a score threshold; however, the score threshold corresponding to a 90 percent sensitivity among the development cohorts that were in Asian populations was a score less than -1.[449]

Across the entire evidence base that used varied score thresholds, sensitivities ranged from 29 percent to 95 percent, and specificities ranged from 25 percent to 92 percent. At a threshold of less than 2, the sensitivities ranged from 53 percent to 95 percent, and the specificities ranged from 36 percent to 74 percent. When limited to the 11 studies reporting in women, the sensitivities ranged from 53 percent to 95 percent, and the specificities ranged from 37 percent to 72 percent. Seven studies reported findings among women younger than age 65 years.[159, 195, 200, 209, 228, 234, 236] The AUCs ranged from 0.63 to 0.77. Sensitivities ranged from 47 percent to 89 percent, and specificities ranged from 45 percent to 81 percent; however, studies used different thresholds which likely contributed the variability in estimates across studies.

Other reported thresholds included less than (or equal to) 1 (6 studies[159, 209, 221, 231, 235, 237]), less than (or equal to) -1 (5 studies[203, 208, 209, 230, 235]), less than (or equal to) 0 (4 studies[196, 198, 235, 450]), less than 3 (4 studies[216, 221, 235, 242]), less than -2 (1 study[218]), less than 6 (1 study[198]), and less than -1.86 (1 study[240]). The use of thresholds other than less than 2 or less than -1 appears to be authors' attempts to assess the influence of different cutoff scores on accuracy or to maximize the accuracy of the tool in the specific population under study. These data demonstrate as much variation among studies using the same score threshold as there is variation across studies that used different score thresholds. This variation is likely explained by differences in the site of BMD measurement used to determine osteoporosis, differences in reference values used for determining T-scores, and differences in the characteristics of the study populations (e.g., population-based cohorts vs. referral populations).

In the studies conducted exclusively in men or reporting separately for men, the AUCs ranged from 0.63 to 0.89, and, across the various score threshold used, sensitivities ranged from 40 percent to 93 percent, and specificities ranged from 25 percent to 85 percent.[198, 201, 216, 218, 219, 221, 235, 237, 240, 243] When limited to the studies using a score threshold of less than 2, the sensitivities ranged from 62 percent to 89 percent, and the specificities ranged from 36 percent to 74 percent.[201, 216, 218, 219, 221, 235, 243]

Osteoporosis Self-assessment Tool for Asians

Study Characteristics

Eleven fair-quality studies (total N=8,304 participants) reported on the accuracy of the Osteoporosis Self-assessment Tool for Asians (OSTA).[201, 204, 207, 212, 213, 215, 227, 229, 233, 244, 245] Three new studies were included in this update.[233, 244, 245] Nine studies were conducted in Asia; four in South Korea;[204, 215, 229, 244] two in Hong Kong;[212, 213] and one each in Singapore,[210] Taiwan,[233] and Malaysia.[245] One study was conducted in Australia (98.6% participants were White),[227] and one study was conducted among a population-based sample in Portugal (race and ethnicity not reported).[201] Six studies were conducted exclusively among women,[204, 207, 213, 215, 227, 245] four studies were conducted exclusively among men,[201, 212, 229, 244] and one study included both men (34%) and women (66%).[233] Most studies recruited participants from the local community [201, 207, 212, 213, 227, 233] or from nationally representative samples.[204, 229, 244] One study recruited participants from a primary care clinic,[245] and one study recruited participants from a menopause clinic.[215] The mean age across studies in women ranged from 59 to 71 years. The mean age across studies in men ranged from 58 to 67 years. The prevalence of osteoporosis in women as measured by DXA BMD T-score less than -2.5 at one or more sites ranged from 11 percent to 42 percent, and the prevalence in men ranged from 6 percent to 18 percent.

Findings

The reported AUCs for women across studies ranged from 0.62 to 0.87 using a reference BMD measurement at either the FN, LS, or both, except for one study that was an extreme outlier with respect to values reported for this tool (along with five other risk assessment tools), suggesting something unique about the underlying study population or study approach; results for this study are not reported further in the text.[245] One study, conducted in Australia, did not report an AUC.[227]

The reported AUCs for men across five studies ranged from 0.62 to 0.94.[201, 212, 229, 233, 244] The AUC in the one study conducted in Portuguese men[201] did not vary from the AUCs reported by the other studies that were all conducted in Asian countries.

The thresholds used for the OSTA varied widely, ranging from less than 0 to -1 and -2 in women and ranging from less than 2, to 0.5, 0, and -1, in men, with the threshold of less than or equal to -1 most used among both women and men. At this threshold, the sensitivities ranged from 41 percent to 100 percent, and the specificities ranged from 27 percent to 67 percent in women.[204, 207, 213, 215, 227, 233] For the three included studies in men that used the threshold of less than or equal to -1, the sensitivities ranged from 71 percent to 100 percent, and the specificities ranged from 58 percent to 68 percent.[212, 229, 233]

Simple Calculated Osteoporosis Risk Estimation

Study Characteristics

Twenty publications reporting on 18 unique cohorts (N=24,461 participants) reported on the accuracy of the Simple Calculated Osteoporosis Risk Estimation (SCORE).[195, 203, 206-211, 217, 222, 224, 226, 228, 230, 231, 233, 234, 236, 242, 245] We rated all analyses as fair quality. Five of the studies were new to this update.[233, 234, 236, 242, 245] One study was conducted in Canada,[206] six studies were conducted in European countries,[203, 208-211, 217, 228, 242] and three studies were conducted in Asian countries.[207, 233, 245] The rest of the studies were conducted in the United States. A third of participants (n=186) in one study were men;[233] the rest of the studies were conducted exclusively among women. The mean age across studies ranged from 51 years to 69 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 4.6 percent to 34.2 percent. This wide variation could be explained by differences in age in the enrolled populations; the study with the lowest prevalence[228] had the lowest mean age, and the study with the highest prevalence[226] had the highest mean age. Studies also varied by whether they reported prevalence based on lowest T-score at any site or based on one site.

Findings

All but one study[231] reported AUCs. The AUCs ranged from 0.58 to 0.91 (except for one outlier study that reported 0.072 and 0.161) using a reference BMD measurement at the FN, TH, LS, or lowest T-score from any of the three sites. Two studies[236, 242] did not report the site for the reference BMD measurement used. When limited to women only, the AUCs ranged from 0.58 to 0.87.

For determining sensitivity and specificity, the most common threshold reported was a score greater than or equal to 6 and was used by 11 studies.[203, 206, 211, 224, 226, 228, 230, 233, 234, 236, 245] The sensitivities for this threshold ranged from 54 to 100 percent, and the specificities ranged from 15 to 72 percent (except for outliers in two studies, which reported 8%[245] and 93%[230]); however, about half of the studies reported specificities less than 50 percent. Five studies reported a score threshold of greater than or equal to 7.[195, 209, 210, 222, 231] The sensitivities for this threshold ranged from 74 to 94 percent, and the specificities ranged from 24 to 71 percent. Five studies reported results for score thresholds of between 8 and 20.75 (**Appendix D Table 14**).[207-209, 211, 242]

At a score threshold of greater than 6 or 7, sensitivities ranged from 54 to 100 percent, and specificities ranged from 24 to 72 percent (except for one outlier that reported 93%[230]).

Six studies reported findings among women younger than age 65 years.[195, 209, 226, 228, 234, 236] In this age group, AUCs ranged from 0.58 to 0.87, sensitivities ranged from 62 percent to 100 percent, and specificities ranged from 25 percent to 71 percent; however, the same score threshold was not used by all studies.

One study, conducted in Taiwan, included 34 percent men and reported results separately for men and women.[233] For men, the AUC was 0.91 with a sensitivity of 100 percent and specificity of 45 percent at a score threshold of 6 or greater in reference to BMD measured at the FN. The AUC for women was 0.80, with a sensitivity of 100 percent and a specificity of 15 percent.

Study of Osteoporotic Fractures Research Group Study Utilizing Risk Factors

Study Characteristics

Three fair-quality studies (total N=1,720 participants) reported on the accuracy of Study of Osteoporotic Fractures Research Group Study Utilizing Risk Factors (SOFSURF);[208, 227, 231] none of the studies were new to this update. One study was conducted in the United Kingdom,[208] one was conducted in Australia,[227] and one was conducted in the United States.[231] All three studies were conducted exclusively among women. The mean age across studies ranged from 60 to 71 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 ranged from 13.8 percent to 41.5 percent. In addition to difference in mean age of the study populations, we note that these studies each used a different normative reference database to transform raw BMD values into T-scores, which may also explain differences in the prevalence of osteoporosis among these studies.

Findings

Only one study (N=208) reported an AUC, which was 0.72 (95% CI, 0.67 to 0.78), using the lowest BMD measurement at the LS or TH as the reference standard.[208]

All three studies used different score thresholds for determining sensitivity and specificity; one study used a score threshold of greater than or equal to 1,[208] another used a score threshold of greater than 1.7,[227] and another study used a score threshold of greater than or equal to 0.[231] The sensitivities ranged from 72 percent to 92 percent, and the specificities ranged from 36 percent to 67 percent.

Veterans Affairs Fracture Absolute Risk Assessment Tool

Study Characteristics

One fair-quality study (total N=463 participants) reported on the accuracy of the Veterans Affairs Fracture Absolute Risk Assessment tool (VA-FARA);[237] this study was new to this update. This study was conducted in the United States, and all participants were men (94% Caucasian). The mean age of enrolled participants was 80 years. The prevalence of osteoporosis as measured by DXA BMD T-score less than -2.5 at either the FN, TH, or LS was 24 percent. Only men older than age 70 years assigned to the bone health team at the study site were enrolled, potentially explaining the high incidence of osteoporosis among participants.

Findings

The reported AUC was 0.640 (95% CI, 0.58 to 0.70) using a reference BMD measurement for the lowest T-score at the FN, LS, or TH. For determining sensitivity and specificity, the study reported score thresholds of greater than 20 percent risk for major fracture or 3 percent for hip fracture. The sensitivity was 64 percent, and the specificity was 58 percent.

E.4 Detailed Results: Benefits and Harms of Treatment (Key Questions 4 and 5)

KQ 4 Detailed Study Characteristics: Bisphosphonates

Alendronate

We identified eight fair- to good-quality RCTs (total N=9,052) that compared alendronate with placebo and reported fracture outcomes; none reported mortality outcomes.[252-254, 256, 259, 267, 268, 277] The largest study, which was conducted in the United States, was the Fracture Intervention Trial (FIT; N=4,432).[254] The remaining seven trials had sample sizes ranging from 144 to 1,908 and consisted of four international multicenter studies[252, 256, 259] and three U.S. studies.[253, 268, 277] All studies were conducted in postmenopausal women. All studies reported the race and ethnicity of the study population, the majority of whom were White. Three studies included participants with prior fracture at baseline, making up 48.5 percent of participants in one study,[267] 21 percent of participants in the second study,[256] and 34 to 42 percent in the third study. No participants had fractures in three studies.[252-254] Two studies did not specify the proportion of participants with fractures at baseline.[259, 268] The duration of intervention ranged from 1 to 3 years. Three trials compared daily (10 mg) or weekly (70 mg) alendronate with placebo,[252, 259, 267] whereas the others compared a range of 1 mg to 40 mg of daily alendronate with placebo.[253, 254, 256, 268, 277] In the four dose-ranging studies, the groups that received daily doses above 10 mg were switched to a lower dose or placebo during the study period.

Zoledronic Acid

We identified three fair-quality[260, 262, 272] and two good-quality RCTs[251, 269] (total N=3,780) examining fracture outcomes for patients receiving zoledronic acid compared with placebo and two that reported mortality.[251, 269] Two studies were new to this update.[269, 272] The studies included sample sizes ranging from 50 to 2,000. Four of the trials were conducted in postmenopausal women,[260, 262, 269, 272] and one included only men ages 50 to 85 years.[251] Three studies reported the race and ethnicity of participants, the majority of whom were White or European,[251, 260, 269] and two studies did not report race or ethnicity.[262, 272] Four of the studies reported the proportion of participants with prior fractures at baseline ranging from 14 percent to 42 percent of the total study population,[251, 262, 269, 272] and one reported no participants with prior fractures.[260] The intervention duration ranged from 1 to 6 years. Three studies[251, 262, 269] compared 5-mg dosages of zoledronic acid intravenous (IV) with placebo, and two studies[260, 272] included dosages ranging from 0.25-mg to 5-mg IV. Two studies administered zoledronic acid as a single dosage,[262, 272] one study administered dosages at 12-month intervals,[251] and one study administered dosages at 18-month intervals.[269] Lastly, one study administered dosages at intervals ranging from 3 months to 1 year with shorter intervals being used for lower dosages and longer intervals for higher dosages.[260]

Risedronate

We identified four fair-quality RCTs (total N=10,161) examining fracture outcomes for patients receiving risedronate vs. placebo.[257, 258, 261, 267] None of these studies reported mortality data. The studies included sample sizes ranging from 111 to 9,331, with the largest being an international multicenter study.[257] All studies were conducted in postmenopausal women, nearly all of whom were White. Two studies reported the prevalence of prior fractures at baseline as 41 percent[257] and 48.5 percent,[267] one study reported no participants with fractures at baseline,[258] and one study did not report about this characteristic.[261] The intervention duration ranged from 3 months to 2 years. All studies compared 5 mg of daily risedronate with placebo. Some studies also compared 2.5 mg daily risedronate[257] and a cyclic regimen of 5 mg daily for the first 2 weeks of the month followed by placebo for the rest of the month with placebo.[258]

Ibandronate

We identified four fair-quality RCTs (total N=564) that examined fracture or mortality outcomes for patients receiving ibandronate vs. placebo, none of which reported fracture outcomes.[264-266, 275] The studies included postmenopausal women with sample sizes ranging from 144 to 240. All participants in one study were White women,[264] and the other three studies did not report the race or ethnicity of participants. Two studies had no participants with prior fractures,[264, 275] and two did not report a history of prior fractures. The intervention duration ranged from 3 months to 2 years. One study compared 150 mg of monthly ibandronate to placebo[275] and one study compared a range of daily doses of ibandronate from 0.25 mg to 5.0 mg with placebo.[264] One study compared 0.25 mg daily ibandronate and an intermittent cyclic dose of 20 mg daily for the first 24 days of every 3 months followed by 9 weeks without an active drug with placebo.[265] Another study compared monthly doses of ibandronate ranging from 50 to 150 mg with placebo, including one arm that received 50-mg ibandronate for the first month and 100 mg for the next 2 months.[266]

Sensitivity Analyses KQ 4 Bisphosphonates

Appendix E.4 Table 1. Results From Sensitivity Analyses for KQ 4 Evaluating Various Dosages of Bisphosphonates and Methods for Pooling Data With Rare Events

Outcome	RR FDA Doses	Peto OR FDA Doses	RR All Doses	Peto OR All Doses
	Measure of effect (95% CI, I^2 value)			
Vertebral fractures	0.51 (0.39 to 0.66, I^2=0%)	0.50 (0.39 to 0.64, I^2=0%)	0.50 (0.38 to 0.66, I^2=0%)	0.50 (0.38 to 0.65, I^2=0%)
Nonvertebral fractures	0.81 (0.74 to 0.88, I^2=0%)	0.78 (0.71 to 0.86, I^2=0%)	0.81 (0.75 to 0.88, I^2=0%)	0.79 (0.72 to 0.87, I^2=0%)
Hip fractures	0.67 (0.45 to 1.00, I^2=0%)	0.66 (0.44 to 0.99, I^2=0%)	0.67 (0.45 to 1.00, I^2=0%)	0.66 (0.44 to 0.99, I^2=0%)
Mortality	0.72 (0.49 to 1.05, I^2=0%)	0.70 (0.47 to 1.04, I^2=0%)	0.71 (0.48 to 1.03, I^2=0%)	0.70 (0.47 to 1.03, I^2=0%)

Abbreviations: CI=confidence interval; FDA=U.S. Food and Drug Administration; KQ=key question; OR=odds ratio; RR=relative risk.

Appendix E.4 Detailed Results: Benefits and Harms of Treatment (Key Questions 4 and 5)

Appendix E.4 Table 2. Results From Sensitivity Analyses for KQ 4 Evaluating Different Types of Vertebral Fractures

Vertebral Fracture Type	RR *Measure of Effect (95% CI, I^2 value)*	Peto OR *Measure of Effect (95% CI, I^2 value)*
All Vertebral Fractures	0.51 (0.39 to 0.66, I^2=0%)	0.50 (0.38 to 0.66, I^2=0%)
Clinical Vertebral Fractures	0.44 (0.24 to 0.79, I^2=0%)	N/A*
Radiographic Vertebral Fractures	0.51 (0.39 to 0.66, I^2=0%)	0.50 (0.39 to 0.65, I^2=0%)

* Only 1 study reported more than 0 clinical vertebral fractures in at least one study arm, therefore meta-analysis was not possible.

Abbreviations: CI=confidence interval; KQ=key question; OR=odds ratio; RR=relative risk.

KQ 4 Detailed Study Characteristics: Denosumab

We identified six fair-quality RCTs (total N=9,108) evaluating denosumab compared with placebo.[278-280, 284-286] Two studies were new to this update.[278, 286] The largest study was the phase 3 international, multicenter Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months (FREEDOM) trial (N=7,808).[280-282, 303] The other four trials had sample sizes ranging from 135 to 365. One study was an international multicenter study[278] and the rest were conducted in the United States,[284] the United States and Canada,[285] Japan,[279]or Korea.[286] Two studies reported the race and ethnicity of the study population, a majority of whom were White (86.2%, 94.2%)[278, 284] and another was conducted exclusively among Japanese individuals.[279] Most were conducted in postmenopausal women (mean age range, 59 to 72 years) with low bone mass or osteoporosis (mean T-score ranging from -3.0 to -1.6 at LS, TH, or FN). One study was conducted exclusively in men ages 30 to 85 years with low bone mass or osteoporosis (mean T-score ranging from -2.0 to -1.3 at the LS, TH, FN, or trochanter.[278] Two trials excluded women with any previous fractures.[284, 285] The other three trials in women had between 23 percent and 50 percent of participants with a prior fracture.[279, 280, 286] The trial conducted in men had 39.3 percent of participants with any prior fracture.[278] Five trials evaluated subcutaneous denosumab (60 mg every 6 months against placebo and measured outcomes at 1 to 3 years followup),[278-280, 284, 285] while one trial compared a single 60-mg intravenous dose of denosumab with placebo at 6 months followup.[286]

Sensitivity Analyses KQ 4 Denosumab

Appendix E.4 Table 3. Results From Sensitivity Analyses for KQ 4 Evaluating Various Dosages of Denosumab and Methods for Pooling Data With Rare Events

Outcome	RR FDA Doses	Peto OR FDA Doses	RR All Doses	Peto OR All Doses
	Measure of effect (95% CI, I^2 value)			
Vertebral fractures			0.33 (0.26 to 0.41, I^2=0%)	0.34 (0.28 to 0.42, I^2=0%)
Nonvertebral fractures				
Hip fractures	0.61 (0.38 to 0.99, I^2=0%)	0.61 (0.38 to 0.98, I^2=0%)		
Mortality			0.79 (0.58 to 1.07, I^2=0%)	0.79 (0.58 to 1.08, I^2=0%)

Abbreviations: CI=confidence interval; FDA=U.S. Food and Drug Administration; KQ=key question; OR=odds ratio; RR=relative risk.

Appendix E.4 Table 4. Results From Sensitivity Analyses for KQ 4 Evaluating Effect of Denosumab on Different Types of Vertebral Fractures

Vertebral Fracture Type	RR *Measure of Effect (95% CI, I^2 value)*	Peto OR *Measure of Effect (95% CI, I^2 value)*
All vertebral fractures	0.33 (0.26 to 0.41, I^2=0%)	0.34 (0.28 to 0.42, I^2=0%)
Clinical vertebral fractures	0.31 (0.21 to 0.47, I^2=0%)	0.35 (0.24 to 0.49, I^2=0%)
Radiographic vertebral fractures*	0.33 (0.01 to 8.12, I^2=0%)	0.37 (0.02 to 5.89, I^2=0%)

* Only one study included in this stratum.

Abbreviations: CI=confidence interval; KQ=key question; OR=odds ratio; RR=relative risk.

KQ 5 Detailed Study Characteristics: Bisphosphonates

Alendronate

We identified 16 RCTs (total N=11,941) reporting on harms. Three were good quality;[254, 298, 300] the rest were fair quality. Twelve RCTs were conducted exclusively in postmenopausal women,[252-254, 256, 259, 267, 268, 276, 294, 296, 297, 300, 302] and two RCTs were conducted in combined populations of women and men (however, women made up over 90% of the study population in these two RCTs).[298, 299] Four RCTs included participants with prior fracture at baseline,[256, 267, 277, 296] and the proportion with a prior fracture ranged from 6.8 percent to 48.5 percent. Four RCTs excluded participants with prior fractures,[252-254, 300] and eight did not specify the proportion enrolled with prior fractures.[259, 268, 276, 294, 297-299, 302] Fourteen RCTs reported the race or ethnicity of participants and one was conducted among exclusively African American women.[276] The other studies reporting race or ethnicity were conducted among mostly White participants.[252-254, 256, 259, 267, 268, 277, 294, 296-299] Two RCTs did not specify the race or ethnicity of participants.[300, 302]

The largest trial was the Fracture Intervention Trial (FIT, N=4,432).[254] The other trials had sample sizes ranging from 144 to 1,908, seven of which were international multicenter RCTs.[252, 256, 259, 267, 294, 298, 302] Eight RCTs, including FIT, were conducted in the United States,[253, 254, 268, 276, 277, 297, 299, 300] and one was conducted in Canada and Colombia.[296] All RCTs compared daily or weekly oral alendronate with placebo for durations ranging from 3 months to 3 years. Six RCTs administered doses of 10 mg daily,[252, 259, 276, 294, 296, 300] four RCTs administered 70 mg weekly,[267, 297-299] and six administered doses ranging from 1 mg to 40 mg daily.[253, 254, 256, 268, 277, 302] In the dose-ranging RCTs, the groups that received daily doses above 10 mg were switched to a lower dose or placebo during the study period.[254, 256, 268]

Zoledronic Acid

We identified six RCTs (total N=4,361) reporting on harms,[251, 260, 262, 269, 272, 287] two of which were new to this update.[269, 272] Two RCTs were good quality,[251, 269] and the rest were fair quality. Study sample sizes ranged from 50 to 2,000. Five of these RCTs were conducted in postmenopausal women,[260, 262, 269, 272, 287] and one was conducted in men ages 50 to 85 years.[251]

Four of the RCTs reported the prevalence of prior fractures at baseline, ranging from 14 percent to 42 percent of the total study population,[251, 262, 269, 272]and two excluded participants with prior fractures.[260, 269] Three RCTs reported the race and ethnicity of participants, a majority of whom were White or European,[251, 260, 269] and three RCTs did not report this information.[262, 272, 287] The duration of the RCTs ranged from 1 to 6 years. Five RCTs compared doses of 5-mg zoledronic acid IV with placebo as either a single dose[262, 272, 287] or with a repeat dose every 1[251, 272] to 1.5 years.[269] One trial administered doses between 0.25 mg and 4 mg at intervals ranging from 3 months to 1 year with shorter intervals being used for lower doses and longer intervals for higher doses.[260]

Risedronate

We identified five fair-quality RCTs (total N=10,372) reporting on harms.[257, 258, 261, 267, 295] The RCTs included sample sizes ranging from 111 to 9,331, with the largest being an international multicenter study.[257] All RCTs were conducted in postmenopausal women. Four RCTs included nearly all White participants,[257, 258, 261, 267] and one RCT was conducted in Japanese women.[295] The duration of the RCTs ranged from 3 months to 2 years. All RCTs compared 5 mg of daily risedronate with placebo. Some RCTs also compared placebo with 1 mg of daily risedronate,[295] 2.5 mg of daily risedronate,[257, 295] or a cyclic regimen of 5 mg daily for the first 2 weeks of the month followed by placebo for the rest of the month.[258]

Ibandronate

We identified eight fair-quality RCTs (N=2,281) reporting on harms.[264-266, 275, 288, 291-293] The RCTs included sample sizes ranging from 126 to 653, all of whom were postmenopausal women. One trial in Denmark included all White participants,[264] and the other trials took place at various sites across North America and Europe but did not report specific race and ethnicity of their participants. The duration of the trials ranged from 3 months to 2 years. Two trials compared 150-mg oral ibandronate monthly with placebo.[275, 288] Four trials compared oral doses ranging from 0.25 to 5 mg daily,[264] 0.5 to 2.5 mg daily,[291] 5 to 20 mg weekly,[292] or 50 to 150 mg monthly with placebo.[266] Two trials compared placebo with cyclic oral regimens including 50 mg for 1 month followed by 100 mg for 2 months[266] and 20 mg daily for the first 24 days of every 3 months followed by 9 weeks without active treatment.[265] One trial compared IV doses of ibandronate ranging from 0.25 to 2 mg every 3 months paired with 1,000 mg daily calcium with placebo.[293]

K5 Sensitivity Analyses: Bisphosphonates

Appendix E.4 Table 5. Results From Sensitivity Analyses for KQ 5 Evaluating Various Dosages of Bisphosphonates and Methods for Pooling Data With Rare Events

Outcome	RR FDA Doses	Peto OR FDA Doses	RR All Doses	Peto OR All Doses
	Measure of effect (95% CI, I^2 value)			
Discontinuations due to AEs	1.00 (0.92 to 1.08); I^2=0%	1.00 (0.91 to 1.10); I^2=0%	0.99 (0.92 to 1.07); I^2=0%	1.00 (0.91 to 1.09); I^2=0%
Serious AEs	0.97 (0.91 to 1.04); I^2=0%	0.96 (0.88 to 1.05); I^2=0%	0.97 (0.91 to 1.03); I^2=0%	0.96 (0.88 to1.04); I^2=0%

Outcome	RR FDA Doses	Peto OR FDA Doses	RR All Doses	Peto OR All Doses
Upper GI AEs	1.02 (0.98 to 1.06); I^2=0%	NA	1.01 (0.97 to 1.05); I^2=0%	NA

Abbreviations: AE=adverse event; CI=confidence interval; FDA=U.S. Food and Drug Administration; GI=gastrointestinal; KQ=key question; OR=odds ratio; RR=relative risk.

KQ 5 Detailed Study Characteristics: Denosumab

The studies included for KQ 5 were the same as the studies included for KQ 4. Please refer to the earlier section for a detailed description.

K5 Sensitivity Analyses: Denosumab

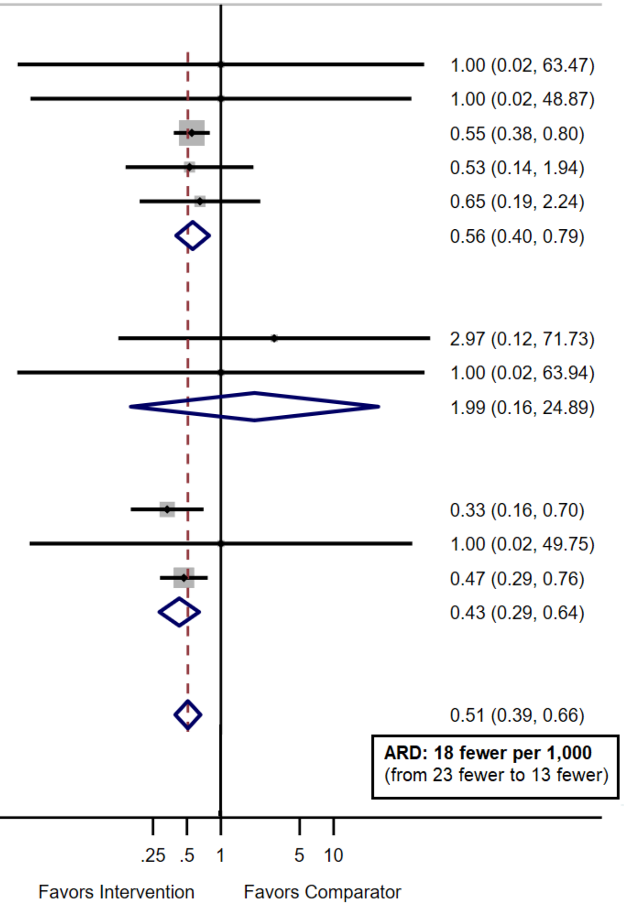
Appendix E.4 Table 6. Results From Sensitivity Analyses for KQ 5 Evaluating Various Dosages of Denosumab and Methods for Pooling Data With Rare Events

Outcome	RR FDA Doses	Peto OR FDA Doses	RR All Doses	Peto OR All Doses
	Measure of effect (95% CI, I^2 value)			
Discontinuations due to AEs	1.16 (0.87 to 1.54); I^2=0%	1.16 (0.85 to 1.56); I^2=0%		
Serious AEs	1.04 (0.97 to 1.12); I^2=0%		1.04 (0.97 to 1.12); I^2=0%	
Upper GI AEs	2.18 (0.74 to 6.16); I^2=0%	2.42 (0.84 to 7.00); I^2=0%	2.13 (0.85 to 5.34), I^2=0%	2.52 (0.96 to 6.66); I^2=0%

Abbreviations: AE=adverse event; CI=confidence interval; FDA=U.S. Food and Drug Administration; GI=gastrointestinal; KQ=key question; OR=odds ratio; RR=relative risk.

Appendix E.4 Figure 1. Key Question 4 Bisphosphonates vs. Placebo Vertebral Fractures

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
Ascott-Evans, 2003	10 mg/d	1 y	0/95 (0.0)	0/47 (0.0)	1.00 (0.02, 63.47)
Chesnut, 1995	10 mg/d	2 y	0/30 (0.0)	0/31 (0.0)	1.00 (0.02, 48.87)
FIT (Cummings, 1998)	Varied*	3 y	43/2214 (1.9)	78/2218 (3.5)	0.55 (0.38, 0.80)
Liberman, 1995	Varied†	3 y	4/384 (1.0)	5/253 (2.0)	0.53 (0.14, 1.94)
Bone, 1997	5 mg/d	2 y	4/93 (4.3)	6/91 (6.6)	0.65 (0.19, 2.24)
Subgroup, DL (I^2 = 0.0%, p = 0.994)					0.56 (0.40, 0.79)
Risedronate					
Mortensen, 1998	5 mg/d	2 y	1/37 (2.7)	0/36 (0.0)	2.97 (0.12, 71.73)
Valimaki, 2007	5 mg/d	2 y	0/114 (0.0)	0/56 (0.0)	1.00 (0.02, 63.94)
Subgroup, DL (I^2 = 0.0%, p = 0.683)					1.99 (0.16, 24.89)
Zoledronic acid					
Boonen, 2012	5 mg/y	2 y	9/588 (1.5)	28/611 (4.6)	0.33 (0.16, 0.70)
Reid, 2002	4 mg/y	1 y	0/61 (0.0)	0/56 (0.0)	1.00 (0.02, 49.75)
Reid, 2018	5 mg/18 m	6 y	23/1000 (2.3)	49/1000 (4.9)	0.47 (0.29, 0.76)
Subgroup, DL (I^2 = 0.0%, p = 0.688)					0.43 (0.29, 0.64)
Heterogeneity between groups: p = 0.340					
Overall, DL (I^2 = 0.0%, p = 0.951)					0.51 (0.39, 0.66)



* Varied dose regimen of 5 mg/d for 2 years then 10 mg/d from 1 year for those without existing vertebral fractures and for 2 to 2.6 years for those with vertebral fractures.

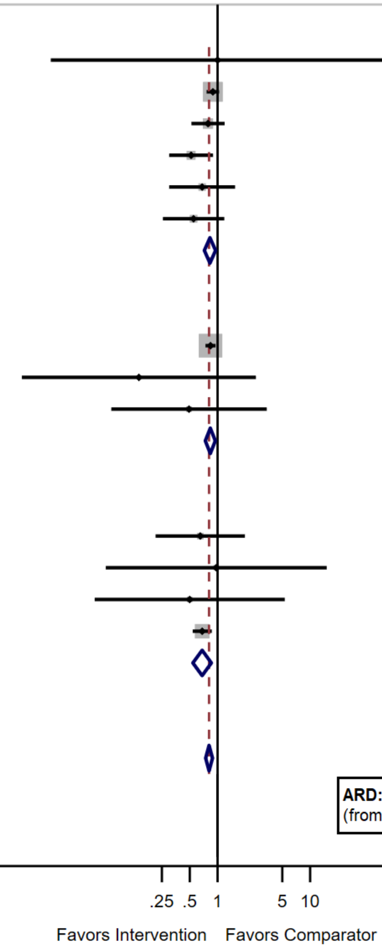
† Varied dose regimen of 5 or 10 mg/d for 3 years or 20 mg/d for 2 years followed by 5 mg/d for 1 year.

Appendix E.4 Figure 1. Key Question 4 Bisphosphonates vs. Placebo Vertebral Fractures

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FIT=Fracture Intervention Trial; vs.=versus; y=year.

Appendix E.4 Figure 2. Key Question 4 Bisphosphonates vs. Placebo Nonvertebral Fractures

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
Ascott-Evans, 2003	10 mg/d	1 y	0/95 (0.0)	0/47 (0.0)	1.00 (0.02, 63.47)
FIT (Cummings, 1998)	Varied*	3 y	261/2214 (11.8)	294/2218 (13.3)	0.89 (0.76, 1.04)
Liberman, 1995	Varied†	3 y	45/597 (7.5)	38/397 (9.6)	0.79 (0.52, 1.19)
FOSIT (Pols, 1999)	10 mg/d	1 y	19/950 (2.0)	37/958 (3.9)	0.52 (0.30, 0.89)
Tucci, 1996	10 mg/d	3 y	7/94 (7.4)	21/192 (10.9)	0.68 (0.30, 1.54)
Bone, 1997	5 mg/d	2 y	9/93 (9.7)	16/91 (17.6)	0.55 (0.26, 1.18)
Subgroup, DL (I^2 = 0.6%, p = 0.412)					0.83 (0.72, 0.95)
Risedronate					
McClung, 2001	2.5 or 5 mg/d	2 y	582/6197 (9.4)	351/3134 (11.2)	0.84 (0.74, 0.95)
Mortensen, 1998	5 mg/d	2 y	0/37 (0.0)	3/36 (8.3)	0.14 (0.01, 2.59)
Valimaki, 2007	5 mg/d	2 y	2/114 (1.8)	2/56 (3.6)	0.49 (0.07, 3.40)
Subgroup, DL (I^2 = 0.0%, p = 0.422)					0.83 (0.74, 0.94)
Zoledronic acid					
Boonen, 2012	5 mg/y	2 y	5/588 (0.9)	8/611 (1.3)	0.65 (0.21, 1.97)
Reid, 2002	4 mg/1 y	1 y	1/61 (1.6)	1/59 (1.7)	0.97 (0.06, 15.11)
Grey, 2012	5 mg single dose	1 y	1/45 (2.2)	2/45 (4.4)	0.50 (0.05, 5.32)
Reid, 2018	5 mg/18 m	6 y	101/1000 (10.1)	148/1000 (14.8)	0.68 (0.54, 0.87)
Subgroup, DL (I^2 = 0.0%, p = 0.987)					0.68 (0.54, 0.86)
Heterogeneity between groups: p = 0.288					
Overall, DL (I^2 = 0.0%, p = 0.668)					0.81 (0.74, 0.88)



ARD: 28 fewer per 1,000
(from 38 fewer to 18 fewer)

* Varied dose regimen of 5 mg/d for 2 years then 10 mg/d from 1 year for those without existing vertebral fractures and for 2 to 2.6 years for those with vertebral fractures.

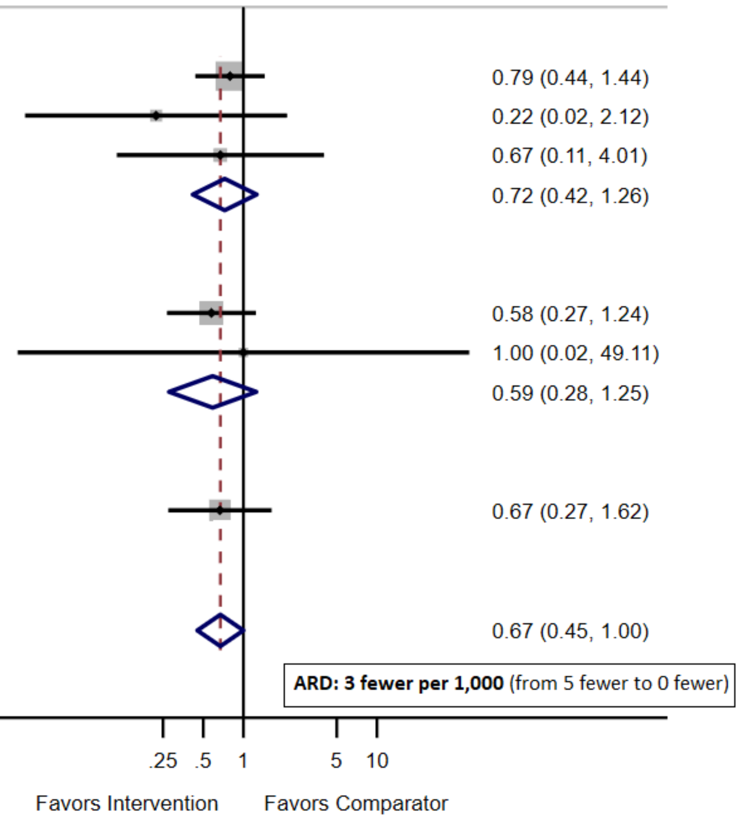
† Varied dose regimen of 5 or 10 mg/d for 3 years or 20 mg/d for 2 years followed by 5 mg/d for 1 year.

Appendix E.4 Figure 2. Key Question 4 Bisphosphonates vs. Placebo Nonvertebral Fractures

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FIT=Fracture Intervention Trial; FOSIT=Fosamax International Trial; vs.=versus; y=year.

Appendix E.4 Figure 3. Key Question 4 Bisphosphonates vs. Placebo Hip Fractures

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
FIT (Cummings, 1998)	Varied*	3 y	19/2214 (0.9)	24/2218 (1.1)	0.79 (0.44, 1.44)
Liberman, 1995	Varied†	3 y	1/597 (0.2)	3/397 (0.8)	0.22 (0.02, 2.12)
FOSIT (Pols, 1999)	10 mg/d	1 y	2/950 (0.2)	3/958 (0.3)	0.67 (0.11, 4.01)
Subgroup, DL (I^2 = 0.0%, p = 0.563)					0.72 (0.42, 1.26)
Risedronate					
McClung, 2001‡	2.5 or 5 mg/d	2 y	14/1773 (0.8)	12/875 (1.4)	0.58 (0.27, 1.24)
Mortensen, 1998	5 mg/d	2 y	0/37 (0.0)	0/36 (0.0)	1.00 (0.02, 49.11)
Subgroup, DL (I^2 = 0.0%, p = 0.785)					0.59 (0.28, 1.25)
Zoledronic acid					
Reid, 2018	5 mg/18 m	6 y	8/1000 (0.8)	12/1000 (1.2)	0.67 (0.27, 1.62)
Heterogeneity between groups: p = 0.909					
Overall, DL (I^2 = 0.0%, p = 0.923)					0.67 (0.45, 1.00)



* Varied dose regimen of 5 mg/d for 2 years then 10 mg/d from 1 year for those without existing vertebral fractures and for 2 to 2.6 years for those with vertebral fractures.

† Varied dose regimen of 5 or 10 mg/d for 3 years or 20 mg/d for 2 years followed by 5 mg/d for 1 year.

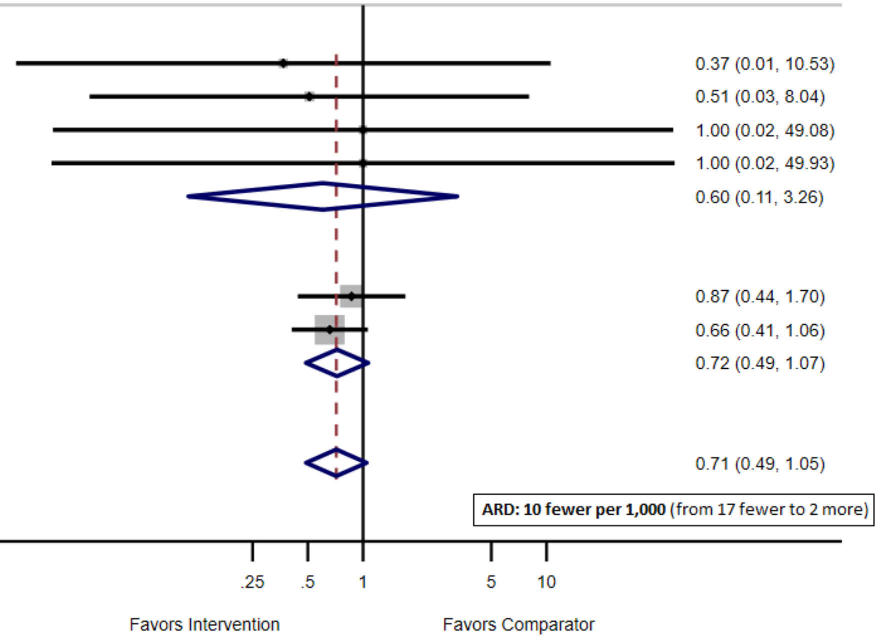
‡ Data included for this analysis are a subgroup without vertebral fracture at baseline. The overall risk ratio when including the entire study population is 0.72 (95% CI, 0.58 to 0.91); Peto odds ratio is 0.71 (95% CI, 0.56 to 0.90).

Appendix E.4 Figure 3. Key Question 4 Bisphosphonates vs. Placebo Hip Fractures

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FIT=Fracture Intervention Trial; FOSIT=Fosamax International Trial; vs.=versus; y=year.

Appendix E.4 Figure 4. Key Question 4 Bisphosphonates vs. Placebo Mortality

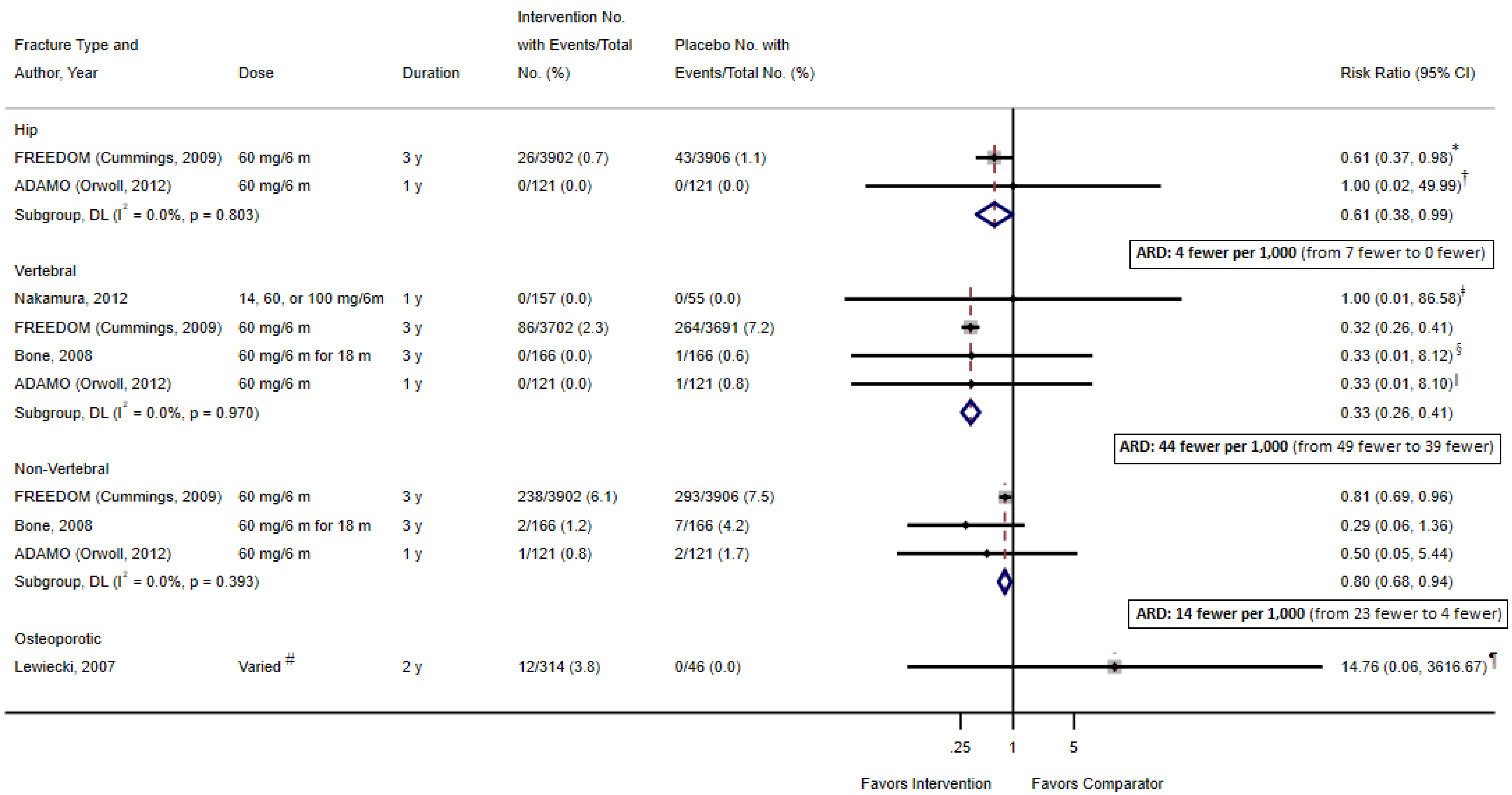
Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Ibandronate					
Ravn, 1996	5.0 mg/d	1 y	0/18 (0.0)	1/25 (4.0)	0.37 (0.01, 10.53)
Riis, 2001	Varied*	2 y	1/159 (0.6)	1/81 (1.2)	0.51 (0.03, 8.04)
MOPS (Reginster, 2005)	150 mg/m	3 m	0/36 (0.0)	0/36 (0.0)	1.00 (0.02, 49.08)
McClung, 2009	150 mg/m	1 y	0/77 (0)	0/83 (0)	1.00 (0.02, 49.93)
Subgroup, DL (I^2 = 0.0%, p = 0.973)					0.60 (0.11, 3.26)
Zoledronic acid					
Boonen, 2012	5 mg/y	2 y	15/588 (2.6)	18/611 (2.9)	0.87 (0.44, 1.70)
Reid, 2018	5 mg/18 m	6 y	27/1000 (2.7)	41/1000 (4.1)	0.66 (0.41, 1.06)
Subgroup, DL (I^2 = 0.0%, p = 0.517)					0.72 (0.49, 1.07)
Heterogeneity between groups: p = 0.840					
Overall, DL (I^2 = 0.0%, p = 0.984)					0.71 (0.49, 1.05)



* 20 mg/2 d for the first 24 days out of every 3 months, followed by a 9-week period without active drug.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; m=month; MOPS=Monthly Oral Pilot Study; vs.=versus; y=year.

Appendix E.4 Figure 5. Key Question 4 Denosumab vs. Placebo Fractures



Note: RRs listed here may differ slightly from the RRs reported by study authors because of differences in statistical packages used. Vertebral fractures reported were radiographic in FREEDOM.

* RR reported by study authors was 0.60 (95% CI, 0.37 to 0.97).

† Peto odds ratio estimate, 1.00 (95% CI, 0.38 to 0.98).

‡ Peto odds ratio estimate, 1.00 (95% CI, 0.01 to 87.49).

§ Peto odds ratio estimate, 0.37 (95% CI, 0.02 to 5.89).

Appendix E.4 Figure 5. Key Question 4 Denosumab vs. Placebo Fractures

[‖] Peto odds ratio estimate, 0.14 (95% CI, 0.00 to 6.62).

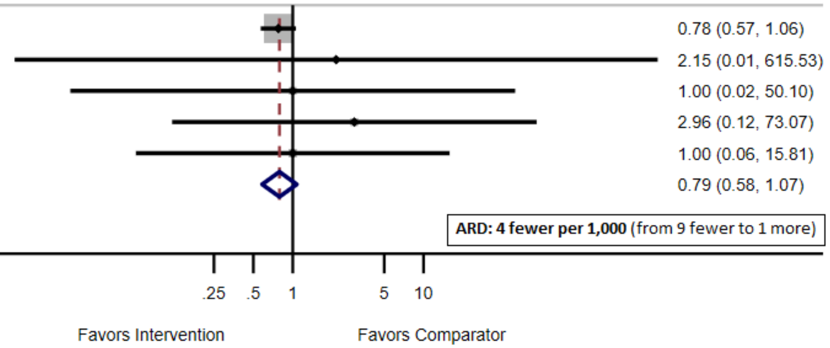
[¶] Peto odds ratio estimate, 2.99 (95% CI, 0.57 to 15.65)

[#] Varied dose regimen of 6, 14, or 30 mg/3 mo or 14, 60, 100, or 210 mg/6 mo

Abbreviations: ARD=absolute risk difference; CI=confidence interval; FREEDOM=Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months; m=month; RR=relative risk; vs.=versus; y=year.

Appendix E.4 Figure 6. Key Question 4 Denosumab vs. Placebo Mortality

Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
FREEDOM (Cummings, 2009)	60 mg/6 m	3 y	70/3886 (1.8)	90/3876 (2.3)	0.78 (0.57, 1.06)
Lewiecki, 2007	Varied*	2 y	1/314 (0.3)	0/46 (0.0)	2.15 (0.01, 615.53)
Bone, 2008	60 mg/6 m for 18 m	3 y	0/164 (0.0)	0/165 (0.0)	1.00 (0.02, 50.10)
Koh, 2016	60 mg single dose	6 m	1/69 (1.4)	0/66 (0.0)	2.96 (0.12, 73.07)
ADAMO (Orwoll, 2012)	60 mg /6 m	1 y	1/121 (0.8)	1/121 (0.8)	1.00 (0.06, 15.81)
Overall, DL (I^2 = 0.0%, p = 0.935)					0.79 (0.58, 1.07)

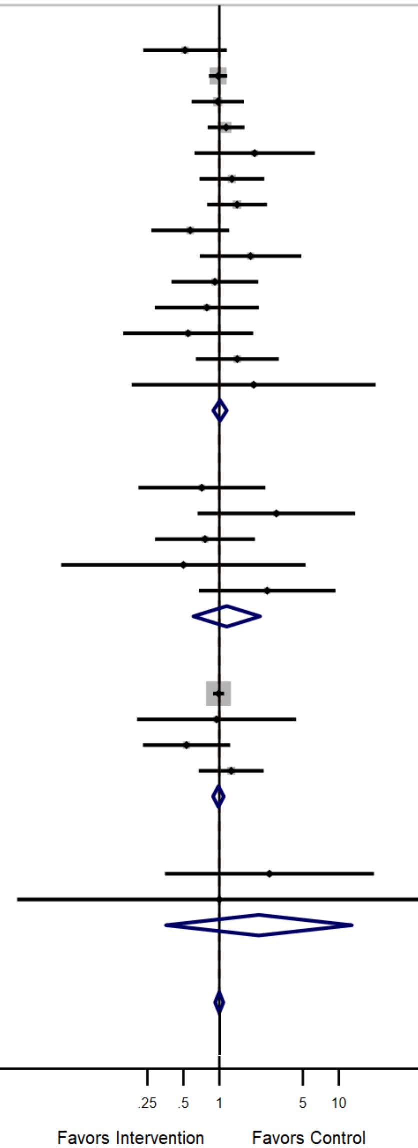


* Varied dose regimen of 6, 14, or 30 mg every 3 months or 14, 60, 100, or 210 mg every 6 months.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; FREEDOM=Fracture Reduction Evaluation of Denosumab in Osteoporosis Every 6 Months; m=month; vs.=versus; y=year.

Appendix E.4 Figure 7. Key Question 5 Bisphosphonates vs. Placebo Discontinuation due to Adverse Events

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
Ascott-Evans, 2003	10 mg/d	1 y	10/95 (10.5)	10/49 (20.4)	0.52 (0.23, 1.15)
FIT (Cummings, 1998)	Varied*	3 y or 4.6 y	221/2214 (10)	227/2218 (10.2)	0.98 (0.82, 1.16)
Liberman, 1995	Varied†	3 y	35/597 (5.9)	24/397 (6)	0.97 (0.59, 1.60)
FOSIT (Pols, 1999)	10 mg/d	1 y	61/950 (6.4)	54/958 (5.6)	1.14 (0.80, 1.62)
Johnell, 2002	10 mg/d	1y	8/83 (9.6)	4/82 (4.9)	1.98 (0.62, 6.31)
Hosking, 2003‡	70 mg/w	1y	31/219 (14.2)	12/108 (11.1)	1.27 (0.68, 2.38)
Adachi, 2009	10 mg/d	3 m	39/291 (13.4)	14/147 (9.5)	1.41 (0.79, 2.51)
Cryer, 2005	70 mg/w	6 m	10/224 (4.5)	18/230 (7.8)	0.57 (0.27, 1.21)
Eisman, 2004	70 mg/w	12 w	11/225 (4.9)	6/224 (2.7)	1.83 (0.69, 4.85)
Greenspan, 2002	70 mg/w	12 w	10/224 (4.5)	11/226 (4.9)	0.92 (0.40, 2.12)
Tucci, 1996	10 mg/d	3 y	5/94 (5.3)	13/192 (6.8)	0.79 (0.29, 2.14)
Devogelaer, 1996	10 mg/d	3 y	3/102 (2.9)	11/205 (5.4)	0.55 (0.16, 1.92)
Bone, 1997	5 mg/d	2 y	13/93 (14.0)	9/91 (9.9)	1.41 (0.64, 3.14)
Bell, 2002	10 mg/d	2 y	2/33 (6.1)	1/32 (3.1)	1.94 (0.18, 20.35)
Subgroup, DL (I^2 = 0.0%, p = 0.513)					1.01 (0.89, 1.15)
Ibandronate					
Chapurlat, 2013	150 mg/m	2 y	4/71 (5.6)	6/76 (7.9)	0.71 (0.21, 2.42)
Ravn, 1996	5.0 mg/d	1 y	6/30 (20.0)	2/30 (6.7)	3.00 (0.66, 13.69)
McClung, 2004	2.5 mg/d	2 y	7/163 (4.3)	9/159 (5.7)	0.76 (0.29, 1.99)
MOPS (Reginster, 2005)	150 mg/m	3 m	1/36 (2.8)	2/36 (5.6)	0.50 (0.05, 5.27)
McClung, 2009	150 mg/m	1 y	7/77 (9.1)	3/83 (3.6)	2.52 (0.67, 9.38)
Subgroup, DL (I^2 = 14.3%, p = 0.323)					1.15 (0.61, 2.20)
Risedronate					
McClung, 2001	2.5 or 5 mg/d	2 y	550/3104 (17.7)	564/3134 (18.0)	0.98 (0.89, 1.10)
Mortensen, 1998	5 mg/d	2 y	3/38 (7.9)	3/36 (8.3)	0.95 (0.20, 4.39)
Valimaki, 2007	5 mg/d	2 y	10/115 (8.7)	9/55 (16.4)	0.53 (0.23, 1.23)
Hosking, 2003‡	5mg/day	3 m	31/222 (14.0)	12/108 (11.1)	1.26 (0.67, 2.35)
Subgroup, DL (I^2 = 0.0%, p = 0.449)					0.98 (0.88, 1.09)
Zoledronic acid					
Reid, 2002	Varied§	1 y	13/292 (4.5)	1/59 (1.7)	2.63 (0.35, 19.69)
Grey, 2012	5 mg single dose	1 y	0/45 (0.0)	0/45 (0.0)	1.00 (0.02, 49.33)
Subgroup, DL (I^2 = 0.0%, p = 0.666)					2.14 (0.36, 12.83)
Heterogeneity between groups: p = 0.791					
Overall, DL (I^2 = 0.0%, p = 0.656)					1.00 (0.92, 1.08)



ARD: 0 fewer per 1,000
(from 9 fewer to 9 more)

Appendix E.4 Figure 7. Key Question 5 Bisphosphonates vs. Placebo Discontinuation due to Adverse Events

[*] Varied dose regimen of 5 mg/d for 2 years then 10 mg/d for 1 year for those without existing vertebral fractures and for 2 to 2.6 years for those with vertebral fractures.

[†] Varied dose regimen of 5 or 10 mg/d for 3 years or 20 mg/d for 2 years followed by 5 mg/d for 1 year.

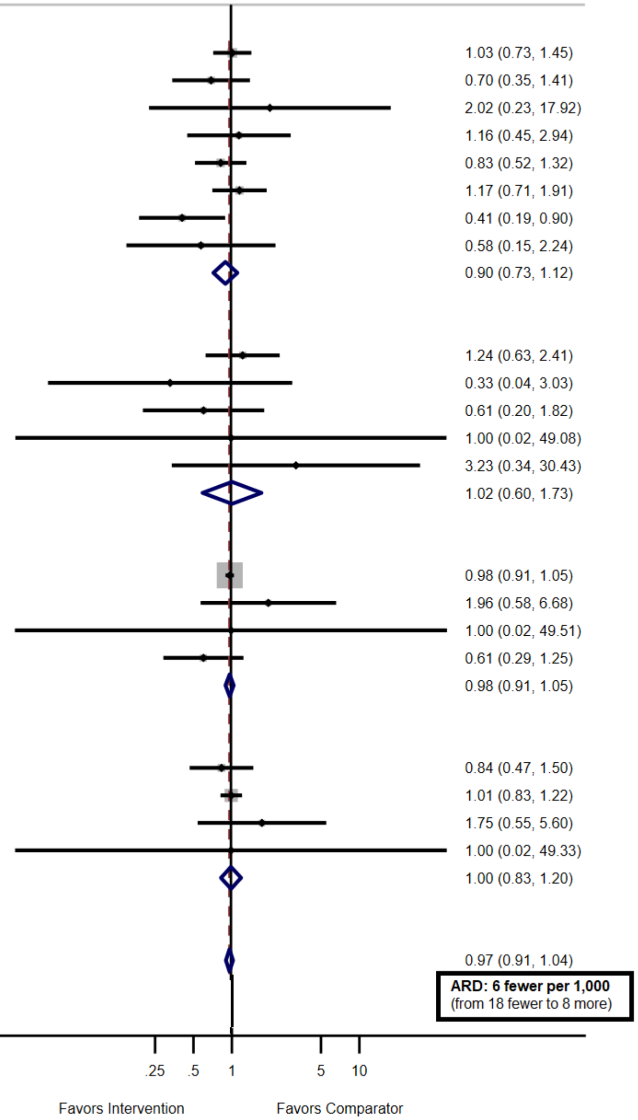
[‡] This study included three study arms: alendronate, risedronate, and placebo. The same placebo group was used in each comparison to the active drug.

[§] Varied dose regimen of 0.25, 0.5, or 1 mg every 3 months; 2 mg every 6 months; or 4 mg every year.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; d=day; DL=DerSimonian & Laird estimator for pooling estimates; FIT=Fracture Intervention Trial; FOSIT= Fosamax International Trial; m=month; MOPS=Monthly Oral Pilot Study; vs.=versus; y=year.

Appendix E.4 Figure 8. Key Question 5 Bisphosphonates vs. Placebo Serious Adverse Events

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
FOSIT (Pols, 1999)	10 mg/d	1 y	61/950 (6.4)	60/958 (6.3)	1.03 (0.73, 1.45)
Hosking, 2003*	70 mg/w	1y	17/219 (7.8)	12/108 (11.1)	0.70 (0.35, 1.41)
Adachi, 2009	10 mg/d	3 m	4/291 (1.4)	1/147 (0.7)	2.02 (0.23, 17.92)
Cryer, 2005	70 mg/w	6 m	9/224 (4.0)	8/230 (3.5)	1.16 (0.45, 2.94)
Greenspan, 2002	70 mg/w	12 w	28/224 (12.5)	34/226 (15)	0.83 (0.52, 1.32)
Tucci, 1996	10 mg/d	3 y	20/94 (21.3)	35/192 (18.2)	1.17 (0.71, 1.91)
Devogelaer, 1996	10 mg/d	3 y	7/102 (6.9)	34/205 (16.6)	0.41 (0.19, 0.90)
Bell, 2002	10 mg/d	2 y	3/33 (9.1)	5/32 (15.6)	0.58 (0.15, 2.24)
Subgroup, DL (I^2 = 3.7%, p = 0.402)					0.90 (0.73, 1.12)
Ibandronate					
Chapurlat, 2013	150 mg/m	2 y	15/71 (21.1)	13/76 (17.1)	1.24 (0.63, 2.41)
Ravn, 1996	5.0 mg/d	1 y	1/30 (3.33)	3/30 (10.0)	0.33 (0.04, 3.03)
McClung, 2004	2.5 mg/d	2 y	5/163 (3.1)	8/159 (5)	0.61 (0.20, 1.82)
MOPS (Reginster, 2005)	150 mg/m	3 m	0/36 (0.0)	0/36 (0.0)	1.00 (0.02, 49.08)
McClung, 2009	150 mg/m	1 y	3/77 (3.9)	1/83 (1.2)	3.23 (0.34, 30.43)
Subgroup, DL (I^2 = 0.0%, p = 0.530)					1.02 (0.60, 1.73)
Risedronate					
McClung, 2001	2.5 or 5 mg/d	2 y	943/3104 (30.4)	973/3134 (31.0)	0.98 (0.91, 1.05)
Valimaki, 2007	5 mg/d	2 y	12/114 (10.5)	3/56 (5.4)	1.96 (0.58, 6.68)
Shiraki, 2003	5 mg/d	8 m	0/53 (0.0)	0/51 (0.0)	1.00 (0.02, 49.51)
Hosking, 2003*	5mg/d	3 m	15/222 (6.8)	12/108 (11.1)	0.61 (0.29, 1.25)
Subgroup, DL (I^2 = 0.0%, p = 0.407)					0.98 (0.91, 1.05)
Zoledronic acid					
McClung, 2009	5 mg/y	2 y	19/198 (9.6)	23/202 (11.4)	0.84 (0.47, 1.50)
Boonen, 2012	5 mg/y	2 y	149/588 (25.3)	154/611 (25.2)	1.01 (0.83, 1.22)
Reid, 2002	Varied†	1 y	26/292 (8.9)	3/59 (5.1)	1.75 (0.55, 5.60)
Grey, 2012	5 mg single dose	1 y	0/45 (0.0)	0/45 (0.0)	1.00 (0.02, 49.33)
Subgroup, DL (I^2 = 0.0%, p = 0.744)					1.00 (0.83, 1.20)
Heterogeneity between groups: p = 0.904					
Overall, DL (I^2 = 0.0%, p = 0.770)					0.97 (0.91, 1.04)



Appendix E.4 Figure 8. Key Question 5 Bisphosphonates vs. Placebo Serious Adverse Events

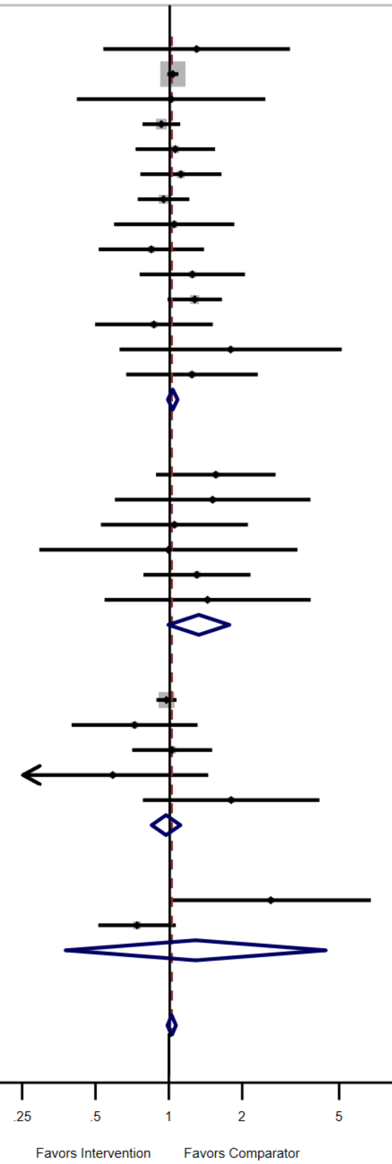
[*] This study included three study arms: alendronate, risedronate, and placebo. The same placebo group was used in each comparison to the active drug.

[†] Varied dose regimen of 0.25, 0.5, or 1 mg every 3 months; 2 mg every 6 months; 4 mg every year.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; d=day; DL=DerSimonian & Laird estimator for pooling estimates; FOSIT=Fosamax International Trial; m=month; MOPS=Monthly Oral Pilot Study; vs.=versus; y=year.

Appendix E.4 Figure 9. Key Question 5 Bisphosphonates vs. Placebo Gastrointestinal Adverse Events

Drug and Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Alendronate					
Ascott-Evans, 2003	10 mg/d	1 y	15/95 (15.8)	6/49 (12.2)	1.29 (0.53, 3.11)
FIT (Bauer, 2000)	Varied*	3 y or 4.6 y	1536/3226 (47.6)	1490/3223 (46.2)	1.03 (0.98, 1.08)
Liberman, 1995	Varied†	3 y	7/196 (3.6)	14/397 (3.5)	1.01 (0.42, 2.47)
FOSIT (Pols, 1999)	10 mg/d	1 y	185/950 (19.5)	202/958 (21.1)	0.92 (0.77, 1.10)
Hosking, 2003‡	70 mg/w	1y	62/219 (28.3)	29/108 (26.9)	1.05 (0.72, 1.54)
Adachi, 2009	10 mg/d	3 m	66/291 (22.7)	30/147 (20.4)	1.11 (0.76, 1.63)
Cryer, 2005	70 mg/w	6 m	79/224 (35.3)	86/230 (37.4)	0.94 (0.74, 1.20)
Eisman, 2004	70 mg/w	12 w	22/225 (9.8)	21/224 (9.4)	1.04 (0.59, 1.84)
Greenspan, 2002	70 mg/w	12 w	25/224 (11.2)	30/226 (13.3)	0.84 (0.51, 1.38)
Greenspan, 2003	10 mg/d	3 y	26/93 (28.0)	21/93 (22.6)	1.24 (0.75, 2.04)
Tucci, 1996	10 mg/d	3 y	49/94 (52.1)	79/192 (41.1)	1.27 (0.98, 1.64)
Devogelaer, 1996	10 mg/d	3 y	52/311 (16.7)	35/205 (17.1)	0.86 (0.49, 1.50)
Johnell, 2002	10 mg/d	1 y	9/83 (10.8)	5/82 (6.1)	1.78 (0.62, 5.08)
Bell, 1997	10 mg/d	2 y	14/33 (42.4)	11/32 (34.3)	1.23 (0.66, 2.30)
Subgroup, DL (I^2 = 0.0%, p = 0.858)					1.03 (0.98, 1.08)
Ibandronate					
Ravn, 1996	5.0 mg/d	1 y	17/30 (56.7)	11/30 (36.7)	1.55 (0.88, 2.72)
MOPS (Reginster, 2005)	150 mg/m	3 m	9/36 (25.0)	6/36 (16.7)	1.50 (0.60, 3.78)
McClung, 2004	2.5 mg/d	2 y	15/163 (9.2)	14/159 (8.8)	1.05 (0.52, 2.09)
Tanko, 2003	20 mg/w	2 y	5/158 (3.2)	5/156 (151)	0.99 (0.29, 3.34)
McClung, 2009	150 mg/m	1 y	24/77 (31.2)	20/83 (24.1)	1.29 (0.78, 2.15)
Thiebaud, 1997	Varied§	1 y	22/100 (22.0)	4/26 (15.4)	1.43 (0.54, 3.79)
Subgroup, DL (I^2 = 0.0%, p = 0.958)					1.32 (0.99, 1.76)
Risedronate					
McClung, 2001	2.5 or 5 mg/d	2 y	657/3104 (21.2)	684/3134 (21.8)	0.97 (0.88, 1.07)
Valimaki, 2007	5 mg/d	2 y	21/115 (18.3)	14/55 (25.5)	0.72 (0.40, 1.30)
Hosking, 2003‡	5mg/day	3 m	61/222 (27.5)	29/108 (26.9)	1.02 (0.70, 1.49)
Mortensen, 1998	5 mg/d	2 y	6/37 (16.2)	10/36 (27.8)	0.58 (0.24, 1.44)
Shiraki, 2003	5 mg/d	8 m	13/53 (24.5)	7/51 (13.7)	1.79 (0.78, 4.12)
Subgroup, DL (I^2 = 7.8%, p = 0.362)					0.97 (0.84, 1.11)
Zoledronic acid					
Grey, 2012	5 mg single dose	1 w	13/45 (28.9)	5/45 (11.1)	2.60 (1.01, 6.69)
Reid, 2018	5mg/18 m	6 y	47/1000 (4.7)	64/1000 (6.4)	0.73 (0.51, 1.06)
Subgroup, DL (I^2 = 83.3%, p = 0.014)					1.28 (0.37, 4.37)
Heterogeneity between groups: p = 0.289					
Overall, DL (I^2 = 0.0%, p = 0.550)					1.02 (0.98, 1.06)



ARD: 5 more per 1,000
(from 5 fewer to 16 more)

Appendix E.4 Figure 9. Key Question 5 Bisphosphonates vs. Placebo Gastrointestinal Adverse Events

[*] Varied dose regimen of 5 mg/d for 2 years then 10 mg/d for 1 year for those without existing vertebral fractures and for 2 to 2.6 years for those with vertebral fractures.

[†] Varied dose regimen of 5 or 10 mg/d for 3 y or 20 mg/d for 2 y followed by 5 mg/d for 1 year.

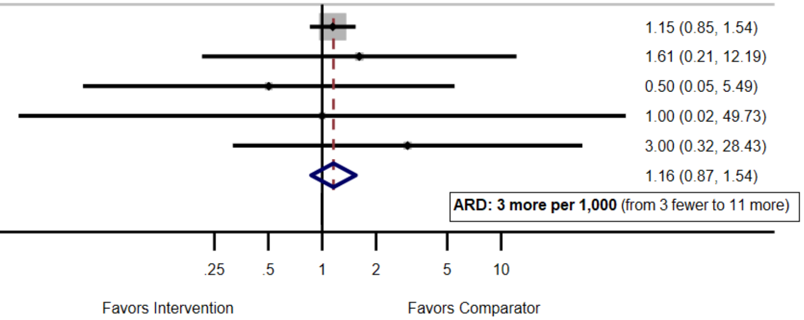
[‡] This study included three study arms: alendronate, risedronate, and placebo. The same placebo group was used in each comparison to the active drug.

[§] Varied dose regimen of 0.25 mg, 0.5mg, 1.0 mg, or 2.0 mg every 3 months.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; d=day; DL=DerSimonian & Laird estimator for pooling estimates; FIT=Fracture Intervention Trial; FOSIT=Fosamax International Trial; MOPS=Monthly Oral Pilot Study; vs.=versus; w=week; y=year.

Appendix E.4 Figure 10. Key Question 5 Denosumab vs. Placebo Discontinuation due to Adverse Events

Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
FREEDOM (Cummings, 2009)	60 mg/6 m	3 y	93/3886 (2.4)	81/3876 (2.1)	1.15 (0.85, 1.54)
Lewiecki, 2007	Varied*	2 y	11/314 (3.5)	1/46 (2.2)	1.61 (0.21, 12.19)
Bone, 2008	60 mg/6 m for 18 m	3y	1/164 (0.6)	2/165 (1.2)	0.50 (0.05, 5.49)
Koh, 2016	60 mg single dose	6 m	0/69 (0.0)	0/66 (0.0)	1.00 (0.02, 49.73)
ADAMO (Orwoll, 2012)	60 mg /6 m	1 y	3/120 (2.5)	1/120 (0.8)	3.00 (0.32, 28.43)
Overall, DL (I^2 = 0.0%, p = 0.867)					1.16 (0.87, 1.54)

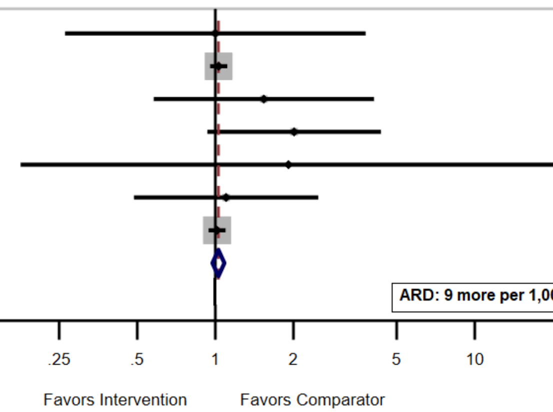


* Varied dose regimen of 6, 14, or 30 mg every 3 months or 14, 60, 100, or 210 mg every 6 months.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; m=month; vs.=versus; y=year.

Appendix E.4 Figure 11. Key Question 5: Denosumab vs. Placebo Serious Adverse Events

Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Nakamura, 2012	60 mg/6 m	1 y	4/54 (7.4)	4/54 (7.4)	1.00 (0.26, 3.79)
FREEDOM (Cummings, 2009)	60 mg/6 m	3 y	1004/3886 (25.8)	972/3876 (25.1)	1.03 (0.95, 1.11)
Lewiecki, 2007	Varied*	2 y	42/314 (13.4)	4/46 (8.7)	1.54 (0.58, 4.09)
Bone, 2008	60 mg/6 m for 18 m	3y	18/164 (11)	9/165 (5.5)	2.01 (0.93, 4.35)
Koh, 2016	60 mg single dose	6 m	2/69 (2.9)	1/66 (1.5)	1.91 (0.18, 20.60)
ADAMO (Orwoll, 2012)	60 mg/6 m	1 y	11/120 (9.2)	10/120 (8.3)	1.10 (0.49, 2.49)
.					1.02 (0.94, 1.09)
Overall, DL (I^2 = 0.0%, p = 0.681)					1.03 (0.97, 1.08)



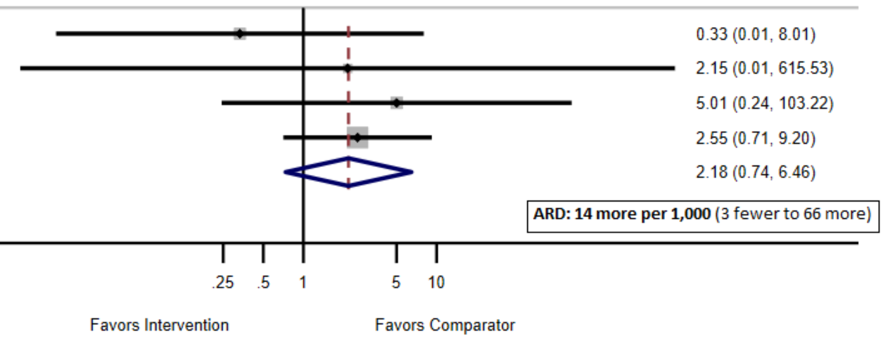
ARD: 9 more per 1,000 (from 7 fewer to 28 more)

* Varied dose regimen of 6, 14, or 30 mg every 3 months or 14, 60, 100, or 210 mg every 6 months.

Abbreviations: ARD=absolute risk difference; CI=confidence interval; DL=DerSimonian & Laird estimator for pooling estimates; m=month; vs.=versus; y=year.

Appendix E.4 Figure 12. Key Question 5: Denosumab vs. Placebo Gastrointestinal Adverse Events

Author, Year	Dose	Duration	Intervention No. with Events/Total No. (%)	Placebo No. with Events/Total No. (%)	Risk Ratio (95% CI)
Nakamura, 2012	60 mg/6 m	1 y	0/54 (0.0)	1/54 (1.9)	0.33 (0.01, 8.01)
Lewiecki, 2007	Varied*	2 y	1/314 (0.3)	0/46 (0.0)	2.15 (0.01, 615.53)
Bone, 2008	60 mg/6 m for 18 m	3y	2/164 (1.2)	0/165 (0.0)	5.01 (0.24, 103.22)
Koh, 2016†	60 mg single dose	6 m	8/69 (11.6)	3/66 (4.5)	2.55 (0.71, 9.20)
Overall, DL (I^2 = 0.0%, p = 0.639)					2.18 (0.74, 6.46)



* Varied dose regimen of 6, 14, or 30 mg every 3 months or 14, 60, 100, or 210 mg every 6 months.

† GI AEs include constipation (Tx 5/69, Ctl 2/66) and Gastritis (Tx 3/69, Ctl 1/66).

Abbreviations: AE=adverse event; ARD=absolute risk difference; CI=confidence interval; Ctl=control; DL=DerSimonian & Laird estimator for pooling estimates; GI=gastrointestinal; m=month; Tx=treatment; vs.=versus; y=year.

F.1. Contextual Question 1

What is the evidence from modeling studies about the effectiveness of risk screening strategies that use different ages at which to start and stop screening and different screening intervals?

Contextual evidence comes from a small number of publications that have attempted to identify appropriate screening intervals based on the time in which it takes individuals to transition to osteoporosis or a certain fracture risk threshold. This range varied across studies.

A publication using healthy postmenopausal women age 65 years or older from the Study of Osteoporotic Fractures evaluated the time for 10 percent of women to develop osteoporosis across the various BMD categories;[317] it found that baseline T-score is the most important determinant of BMD testing intervals, with results suggesting that the times for 10 percent of women to develop osteoporosis are as follows: 16.8 years (95% CI, 11.5 to 24.6) for women with normal BMD (T-score, -1.00 or higher), 17.3 years (95% CI, 13.9 to 21.5) for women with mild osteopenia (T-score, -1.01 to -1.49), 4.7 years (95% CI, 4.2 to 5.2) for women with moderate osteopenia (T-score, -1.50 to -1.99), and 1.1 years (95% CI, 1.0 to 1.3) for women with advanced osteopenia (T-score, -2.00 to -2.49).[317] Within a given T-score range, the estimated time for 10 percent of women to transition from osteopenia to osteoporosis was longer for women with younger age and for those taking estrogen at baseline. For women with moderate osteopenia at baseline, the estimated BMD testing interval was 5.6 years (95% CI, 4.9 to 6.4) for women age 67 years compared with 3.2 years (95% CI, 2.6 to 3.9) for women age 85 years. Also for women with moderate osteopenia, the estimated BMD testing interval for past or never-users of estrogen was shorter, 4.3 years (95% CI, 3.9 to 4.8), than for women with current estrogen use, 6.9 years (95% CI, 5.7 to 8.4).[317]

Using an absolute risk-based prognostic model with a sample of nonosteoporotic women and men older than age 60 years from the Dubbo Osteoporosis Epidemiology study, authors found that current age and BMD T-score could be used to estimate the optimal time to repeat BMD testing for both men and women.[451] For example, the time for women age 60 years with a normal BMD to reach a 10 percent risk of sustaining a fracture or developing osteoporosis was 8.9 years (90% CI, 6.7 to 10.6); it was 2.7 years (90% CI, 2.3 to 3.1) for women age 80 years.

A third study provided contextual evidence for identifying the time to transition to fracture (rather than osteoporosis) in younger postmenopausal women ages 50 to 64 years.[452] In a study of women from the Women's Health Initiative with a baseline BMD, investigators estimated the time for 1 percent of women to sustain a hip or clinical vertebral fracture and for 3 percent of women to sustain a MOF.[452] Women were followed for up to 11 years after the initial BMD. Similar to findings of studies estimating time to transition to osteoporosis, the study found that age and baseline T-score were associated with the estimated time for 1 percent of women to transition to fracture. For women without osteoporosis at baseline (T-score >-2.50), the estimated times for 1 percent of women to transition to hip or clinical vertebral fracture were 12.8 years (95% CI, 8 to 20.4) for ages 50 to 54 years, 11.7 years (95% CI, 6.9 to 20) for ages 55 to 59 years, and 7.6 years (95% CI, 4.8 to 12.1) for ages 60 to 64 years. For all women with osteoporosis at baseline (T-score ≤-2.50), the time interval for 1 percent of women ages 50 to 64 years to transition to hip or clinical vertebral fracture was 3.0 years (95% CI, 1.3 to 7.1). There were similar findings for MOF.

F.2. Contextual Question 5

What are the implications of using fixed-fracture risk thresholds for decisions regarding stepwise screening or treatment?

The predictive and diagnostic accuracy of risk assessment instruments are described in detail in KQs 2a and 2c of this update evidence report. The most reported accuracy outcome was AUC, which represents the average value of sensitivity and specificity over all possible values. However, for risk assessments to be usable in clinical practice to inform shared decision making about who to screen with DXA or who to treat with pharmacotherapy, risk thresholds must be established. Although many studies included for KQ 2a and KQ 2c reported AUC, fewer studies reported the sensitivity or specificity of specific risk thresholds. In considering the role of risk assessments in clinical practice, an understanding of the origin of commonly cited thresholds and advantages and disadvantages of fixed or variable risk thresholds is warranted. Nearly all articles that discussed intervention thresholds focused on FRAX because it is the most ubiquitous and widely studied risk assessment tool. Thus, this CQ will focus exclusively on the impact of using fixed-fracture risk thresholds with FRAX and challenges related to a mechanistic application of thresholds versus their use as part of shared decision making.

Origins of FRAX Fixed Threshold for Intervention

For primary fracture prevention in the United States, the Bone Health and Osteoporosis Foundation (formerly known as the National Osteoporosis Foundation [NOF]) recommends treatment for individuals with osteoporosis, prior fragility fracture, or in persons with low bone mass (i.e., formerly called osteopenia) who have a 10-year hip fracture risk of at least 3 percent or a 10-year MOF risk of at least 20 percent based on FRAX.[84] The hip fracture risk threshold was selected based on a U.S.-specific economic analysis of cost-effectiveness from a societal perspective sponsored by the NOF and that assumed one-step BMD screening, use of generic bisphosphonates, a relative risk reduction of 35 percent for all fracture types, and a willingness-to-pay threshold of $60,000 per quality-adjusted life-year gained.[80, 85] The MOF threshold was derived from the hip fracture threshold through a complex transformation.[88] These thresholds (3% hip, 20% MOF) are pervasively cited in the literature and have formed the basis of intervention thresholds used in many countries other than the United States but have never been evaluated in trials. Some countries have used a similar methodology to derive their own country-specific intervention thresholds for considering treatment. Such studies have factored in reimbursement considerations, access to DXA, local health economic assessments, and willingness to pay for osteoporosis-related care.[88] Thus, intervention thresholds that are often recommended for use in clinical practice are based on a variety of factors beyond clinical benefit or harms, including economic considerations.

The prevalence of estimated FRAX risks above the 3 percent hip/20 percent MOF risks based on 2013–2014 U.S. NHANES data are summarized in **Table F.2-1**.[453] Across all adults age 50 years or older, 81 percent will have estimated fracture risks below these thresholds for both fracture types, 8 percent will have an estimated risk above both the hip fracture and MOF thresholds, 11 percent will have an estimated hip risk above the threshold alone, and less than 1 percent will have an estimated MOF risk above the threshold alone.[453] If these prevalences are applied to the entire U.S. population based on 2020 Census data, the absolute number of persons with estimated fracture risk at or above these thresholds can be estimated (**Table F.2-1**). If a lower risk threshold is used, more people would be above the threshold, and even small changes have the potential to affect a large absolute number of persons.[454] Similarly, the use of a higher threshold (such as what might result from an increase in the price of medications or willingness-to-pay assumption) would result in fewer persons above the threshold.

As the use of fracture risk assessments has become more common, experts continue to emphasize that decisions about treatment should not be based solely on fracture risk and that clinical judgment and shared decision making should continue to play a key role in decision making.[20, 80] An overlay mechanistic application of thresholds can lead to clinically illogical scenarios, for example, offering treatment to someone just above the threshold but not to someone with the same clinical risks who might fall just below the threshold because they are a few years younger. Further, the sensitivity of the currently established thresholds to the price of medication may be of concern for implementing fixed thresholds for individual clinical decision making. Although much has been published on establishing treatment thresholds, relatively less has been published concerning thresholds for screening with DXA.

Table F.2-1. Prevalence of High Fracture Risk Among Adults Age 50 or Older Based on NHANES Data (NHANES, 2013–2014)[453] Extrapolated to the Size of the U.S. Population Based on 2020 Census Data[455]

Characteristics	Proportion With 10-Year Hip Fracture Risk 3% or Higher	Number of Persons	Proportion With 10-Year MOF Risk 20% or Higher	Number of Persons
All persons	22.6%*	26,088,315	9.6%*	11,081,762
Men	16.6%*	8,948,369	2.3%*	1,239,834
Women	27.4%*	16,858,997	15.5%*	9,537,024
Ages 50 to 59	6.7%	2,864,685	2.9%	1,239,938
Ages 60 to 69	11.3%	4,250,776	6.4%	2,407,519
Ages 70 to 79	38.6%	8,630,918	16.1%	3,599,943
Age 80 or older	71.6%	9,094,026	27.4%	3,480,116
Non-Hispanic White	25.5%*	20,970,366	11.6%*	9,539,461
Non-Hispanic Black	4.8%*	589,696	Unstable estimate	NA
Hispanic	10.7%*	1,530,225	1.8%*	257,421
Non-Hispanic Asian	16.0%*	991,740	Unstable estimate	NA

* Age-adjusted estimate.

Note: We calculated the number of persons by multiplying the number of persons in the age/sex/race category by the proportion with a 10-year FRAX fracture risk equal to or more than the 3% (hip) or 20% (MOF) risk. The number of persons represent the ceiling of potential persons who would be candidates for screening or treatment as some will not be eligible for various clinical or other reasons.

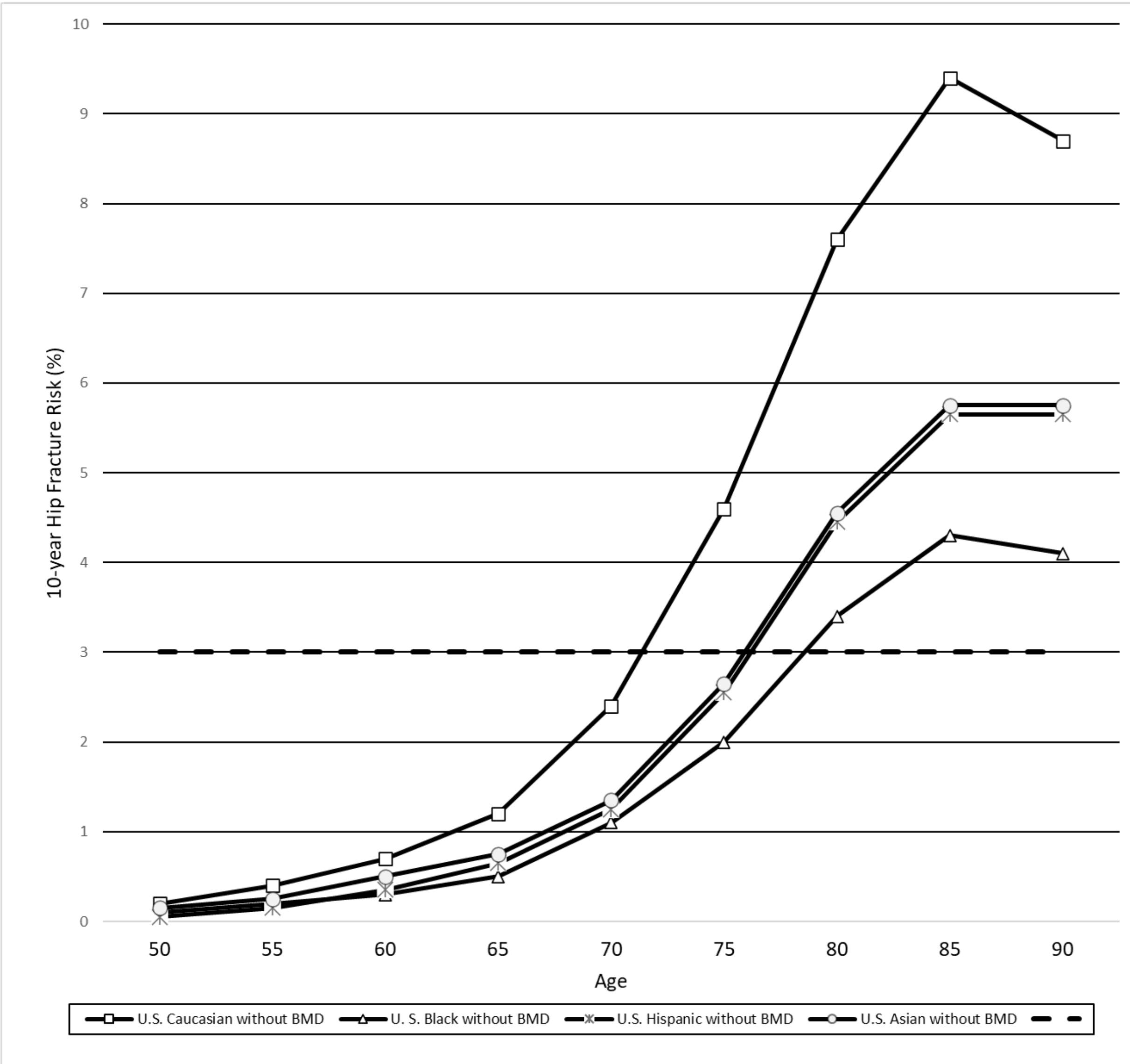
Abbreviations: MOF=major osteoporotic fracture; NA=not available; NHANES=National Health and Nutrition Examination Survey; U.S.=United States.

Types of Intervention Thresholds

The intervention threshold described in the previous section is considered a fixed threshold because it is applied to men and women irrespective of age. Fixed thresholds are the easiest to implement in clinical practice. However, if one considers that persons with prior fragility fracture should be treated (regardless of BMD), then a fixed threshold creates a problem for younger persons (i.e., women younger than age 65 years and men younger than age 70 years), who will seldom have a risk above established thresholds even with a prior fracture. Yet, lowering established thresholds means a sizable proportion of the population would suddenly become eligible, and nearly all persons at older ages would be eligible. Further, because hip fracture incidence in the United States is lower in most non-White racial and ethnic groups, predicted fracture risk estimates for persons in non-White racial and ethnic groups will always be lower than risk estimates for White persons of the same age, sex, weight, BMD, and clinical risks used in the FRAX model.[80] **Figure F.2-1** illustrates the predicted FRAX 10-year hip fracture risk (without BMD input) for women with BMI of 25 kg/m^2 without any clinical risk factors for women across ages 50 to 90 years. Estimates at the oldest ages decline because of competing mortality. The pattern is similar for men except that the steep increase in predicted fracture starts a decade later in men compared with women. White women cross the hip fracture risk intervention threshold of 3 percent just after age 70 years, while Black women do not cross for 6 to 7 years later. A systematic review published in 2016 reported 82 guidelines recommending the use of FRAX; 58 recommended fixed thresholds and 24 recommended age-dependent thresholds.[88] In almost all cases, these guidelines were recommending thresholds for treatment intervention, and the role of these thresholds for informing decisions about DXA testing varied across guidelines.

Age-specific thresholds vary the threshold for intervention by age. The most common way this is done is by setting the intervention threshold at the risk equivalent of a person of the same age with a prior fracture. The rationale for this approach is that if a person at a certain age with prior fracture is eligible for treatment, then a person without fracture but at the same risk (presumably because of other risk factors) should also be eligible for treatment. Under this model, the intervention thresholds are generally lower at younger ages and increase with age, but then plateau or even decrease to account for competing mortality at the oldest of ages.[88] This allows for younger persons at elevated relative risk to be identified without having to lower the threshold for all ages, which would result in most older persons being above the threshold. Age-dependent thresholds are more complicated to implement in practice but may be better at efficiently identifying the persons at highest risk.[88] Age-dependent thresholds also have the advantage of not being dependent on cost-effectiveness findings, which become outdated as costs of drugs or willingness-to-pay thresholds change.[20] However, some have suggested that the use of a threshold equivalent to someone with a prior fracture sets the risk threshold too high, and empiric evaluations of this approach suggests it misses many persons who end up having fractures who may have benefited from treatment.[456] Further, in one application of age-dependent thresholds in the United Kingdom, analyses suggested the creation of a disparity in access to treatment for some women age 70 years or older without prior fracture, as these women had higher estimated fracture risks than women of same age with a prior fracture, yet were not getting offered treatment.[454, 457] As a result, hybrid thresholds were implemented that included age-dependent thresholds through age 70 years and then applied fixed thresholds after age 70 years.[88]

Figure F.2-1. Ten-Year Hip Fracture Risk According to the Fracture Risk Assessment Tool (FRAX) for Women Ages 50 to 90 Years



Note: Fracture risk based on woman with BMI of 25.0 (height 64 in, weight 141 lb) and no other clinical risks. The horizontal dashed line at 3% percent 10-year hip fracture risk represents a common threshold for treatment intervention promoted in the United States.

Abbreviations: BMD=bone mineral density; BMI=body mass index; FRAX= Fracture Risk Assessment Tool; U.S.=United States.

The U.K. National Osteoporosis Guideline Group recommends a hybrid-threshold and direct treatment (without BMD testing) for those above the threshold considered high risk and reassurance and no BMD testing for those below the threshold considered low risk.[457, 458] In this approach, BMD measurement is reserved for those considered at intermediate risk based on initial fracture risk assessment. The fracture risk is then recomputed with BMD, and the patient is reclassified as high risk or low risk. At least one study has demonstrated that the use of a fixed threshold in the oldest age groups reduced the need for BMD in older age groups compared with an age-dependent threshold.[457] Opponents of an approach that recommends direct treatment for high fracture risk cited the lack of trial evidence in persons without BMD testing.[20] Proponents argued that because many (if not the majority of) fragility fractures occur in community-dwelling people with T-scores greater than -2.5, requiring a BMD assessment in the osteoporosis range for treatment is not useful.[20] Post hoc analyses of some treatment trials demonstrated no treatment heterogeneity based on baseline BMD level, and larger fracture reductions in persons at higher baseline FRAX risk compared with lower baseline risk seemed to support this position.[454] Proponents also suggested this approach may be most useful in low-resource settings where DXA resources are limited.[454]

F.3 Contextual Question 6

What is the evidence for rare harms of bisphosphonate treatment (i.e., osteonecrosis of the jaw, atypical femur fractures) from observational studies that use noneligible control groups or are uncontrolled?

Summary

In addition to studies eligible for inclusion in the SR portion of this update (KQ 5: Harms of Treatment), we sought recent seminal reviews and reports supplemented with new observational studies with large sample sizes (≥1,000) to address this CQ. The studies we identified for this CQ consistently suggest increased risk of atypical femur fractures (AFF) or osteonecrosis of the jaw (ONJ) with bisphosphonate (BP) use and increases in risk with longer duration of therapy, though risk estimates vary widely given differences in comparator arms, definitions and method of outcome ascertainment, and followup duration. In addition, estimates related to long-term use may be subject to confounding by indication as longer-term users may also have lower initial BMD or elevated fracture risk factors. However, the absolute risk of these outcomes is rare.[459] Risk for these harms typically declined with cessation of BP treatment. Few studies included for this CQ reported on BPs other than alendronate. Studies frequently also considered BPs as a class in analyses. Studies typically included primarily postmenopausal women.

Detailed Findings

Atypical Femur Fracture

Definition. The American Society for Bone Mineral Research revised its definition of AFF in 2013 to include fractures located along the femoral diaphysis from just distal to the lesser trochanter to just proximal to the supracondylar flare, with four of five major features present. Major features include:

- Fracture associated with minimal or no trauma
- Fracture line originating at lateral cortex with substantially transverse orientation
- Complete fractures extend through both cortices, incomplete involve only lateral cortex
- Noncomminuted or minimally comminuted fractures
- Localized periosteal or endosteal thickening of the lateral cortex at fracture site.[459]

The definition also outlines minor features (increased cortical thickness of the femoral diaphysis, prodromal groin or thigh pain, bilateral incomplete or complete femoral diaphysis fractures, delayed fracture healing) that are not required to be present but have been associated with AFF.

Evidence From Systematic Reviews and Seminal Reports

Bisphosphonates. ASBMR reports noted variable risk ratios (RRs) for AFF but consistently low absolute risk (3.2 to 50 cases/100,000 person-years [**Appendix F.3 Table 1**]).[459, 460] Risk increased with prolonged BP use (>3 years) and declined with discontinuation. While the optimal duration of use is not clear and likely varies based on patient risk factors, ASBMR suggests that for up to 5 years of BP use among 100,000 users, 175 hip fractures, 1,470 vertebral fractures, and 945 wrist fractures would be averted (2,590 total) and 16 AFFs would occur, for a total of 162 fractures of the spine, hip, or forearm prevented per AFF caused.[460] It should be noted that AFF also occurs in individuals who did not receive antiresorptive therapies.[460]

Three SRs[461-463] reported consistently increased risk of AFF with BP use, though the magnitude of risk varied by agent and study design (**Appendix F.3 Table 1**). Absolute numbers of AFF cases, when reported, varied from 0 to 412 in populations ranging from 2,000 to over 2 million. One review of SRs reported an adjusted odds ratio (OR) for AFF of 1.99 to 2.08 in studies of BP users vs. control or no exposure and RR estimates ranging from 1.52 to 11.12 depending on the type of studies (RCTs, observational studies) included and the duration of use (variably defined as >1 year to >6 years).[461] Strength of evidence reported for these findings in this SR ranged from very low to moderate. Overlap among the studies included in these SRs was not described, and the authors reported variable methodologic quality (median 7.5 rating on 11-point AMSTAR scale).[461]

A second SR examined fracture risk with long-term use of BPs in postmenopausal women with at least 12 months of exposure and similarly reported higher AF risk in women taking alendronate vs. placebo.[462] Observational studies included in this SR reported increased risk of AFF with longer treatment duration. In one Kaiser Permanente cohort, the incidence rate of AFF after 2 years of BP exposure was 2 per 100,000 person-years and 78 per 100,000 person-years after 8 years.

A third SR assessed long-term (>3 years) use of BPs and reported wide-ranging risk estimates for AFF depending on study design: in an RCT and observational studies of alendronate vs. placebo or no treatment, HRs for AFF (with or without radiologic confirmation) ranged from 1.03 to 2.90, while in observational studies of BPs vs. no BPs, ORs ranged from 9.46 to 116.[463] Overall, this SR reported increased AFF risk with BP use with low strength of evidence.

Appendix F.3 Table 1. Reviews and Seminal Reports Addressing Atypical Femoral Fracture and Bisphosphonate Use

Author, Year (Search Dates)	Focus	Agents Included	Included Studies Addressing AFF (N Cases/N Participants)	AFF Outcomes
ASBMR Reports				
Shane et al 2013[459] (NA)	Epidemiology and definition of AFF	BPs as a class	12 observational studies with radiographic adjudication (458/NR)	Proportion ST/FS fractures with AFF features ranged from 1% to 48% Number of AFF in each study ranged from 6 to 142, proportion occurring in BP users ranged from 12% to 97% aORs for AFF in BP users ranged from 2.1 to 69.1 Absolute risk for AFF ranged from 3.2 to 50 cases/100,000 patient-years. For up to 5 years of use among 100,000 BP users, 2,590 fractures would be averted and 16 AFF would occur (162 fractures averted per 1 AFF caused).
Adler et al 2016[460] (NR)	Safety of long-term BP use	BPs as a class	NA	AFF risk increases with duration of BP exposure; in one study aIR increased from 1.8/100,000 per year with 2 years exposure to 113/100,000 per year with 8 to 9.9 years of exposure
Reviews				
Lu et al 2020[461] (NR-Dec. 2018)	Review of SRs and meta-analyses of RCTs or observational studies reporting rare harms associated with long-term (>1 year) BP use	Alendronate, ibandronate, etidronate, zoledronic acid, or risedronate	3 SRs addressing alendronate or BPs as a drug class (NR)	Pooled risk measures (95% CI) for AFF in included SRs Lee 2015 (BP vs. control or no exposure, N=658,497, 1 RCT+9 obs studies); aOR=1.99 (1.28 to 3.10); GRADE: Very low Lee 2015 (BP vs. control or no exposure, N=643,174, 9 obs studies); aOR=2.08 (1.29 to 3.35); GRADE: Low Lydia 2013 (BP vs. control, N pts=NR, 11 obs studies); RR=1.70 (1.22 to 2.37); GRADE: Very low Lydia 2013 (BP vs. control, N=686,929, 6 obs studies); RR=1.52 (1.08 to 2.15); GRADE: Moderate Lydia 2013 (BP vs. control, N=NR, 5 case-control studies); RR=11.12 (2.68 to 46.18); GRADE: Low Lu 2013 (BP vs. no exposure, N=686,929, 6 obs studies); RR=1.55 (0.94 to 2.16); GRADE: Very Low Lu 2013 (BP vs. no exposure >5 years' use, N=247,211, 3 obs studies); RR=1.54 (1.16 to 1.92); GRADE: Moderate
Dennison et al 2019[462] (NR)	Fracture risk with long-term use of BP in postmenopausal	Alendronate, zoledronic acid, risedronate	1 secondary analysis of FIT and FLEX and Horizon RCTs* (2/14,195)	Overall IR=2.3 per 10,000 patient-years HRs ranged from 1.03 to 1.33 in 1 RCT and extension of alendronate vs. placebo HR of 1.50 (0.25 to 9.00) in RCT of zoledronic acid vs. placebo

Author, Year (Search Dates)	Focus	Agents Included	Included Studies Addressing AFF (N Cases/N Participants)	AFF Outcomes
Dennison et al 2019[462] (continued)	women with ≥1 year of treatment exposure		4 obs studies (NR)	Swedish registry study, 59 AFF: OR for AFF, BP vs. no BP ranged from=0.28 to 33.3 depending on duration of use and time since last use (decline in risk post-discontinuation) Swedish registry study, 172 AFF: aRR for AFF in female BP users=55 and 54 in males. RR after ≥4 years' use=126 (95% CI, 55 to 288), AR=11 AFF per 10,000 person-years (95% CI, 7 to 14); declines in risk after discontinuation Southern California OP cohort: HR for AFF in women stopping BP vs. continuing=0.56 (95% CI, 0.38 to 0.82); IR in current users=46 per 100,000 patient-years Kaiser Permanente cohort: Unadjusted IR after 2 years BP exposure=2 per 100,000 person-years, after 8 years=78 per 100,000 person-years
			1 SR (23 studies)	IR ranged from 3.0 to 9.8 per 100,000 patient-years
Fink et al 2019[463] (Jan. 1995-Oct. 2018)	SR of RCTs or obs studies of long-term (>3 years) use of OP drug treatment vs. control in men or women age ≥50	Alendronate, zoledronic acid, any BP	1 secondary analysis of FIT AND FLEX and Horizon RCTs*, 2 obs studies of alendronate vs. no treatment (316/227,353)	RCT analysis, HR for ST or FS fracture with alendronate=1.03 (95% CI, 0.06 to 16.46), ARR=0 (95% CI, -0.09 to 0.09); SOE: Insufficient Obs studies—HR for ST or FS ranged from 1.37 to 2.90, ARRs 0.09 to 0.20; SOE: Low
			3 obs studies, BP vs. no BP (412/~2,808,032)	OR for AFF or ST/FS without X-ray confirmation ranged from 9.46 to 116; SOE: Low
			2 obs studies, current BP use vs. past BP use (368/2,027)	Higher risk with current use, OR ranged from 1.59 to 5.17; SOE: Low
			1 RCT†, zoledronic acid vs. placebo (0/2,000)	0 AFF, SOE: Not rated

* One FIT publication (Cummings et al, 1998[254]) included to address KQs.

† RCT included to address KQ—Reid et al, 2018.[269]

Abbreviations: AFF=atypical femoral fracture; aIR=adjusted incidence rate; aOR=adjusted odds ratio; AR=absolute risk; aRR=adjusted risk ratio; ARR=absolute risk reduction; ASBMR=American Society for Bone and Mineral Research; BP=bisphosphonates; CI=confidence interval; FIT=Fracture Intervention Trial; FLEX=Fracture Intervention Trial Long-term Extension; GRADE=Grading of Recommendations, Assessment, Development, and Evaluation; HR=hazard ratio; IR=incidence rate; KQ=key question; N=number; NA=not applicable; NR=not reported; obs=observational; OP=osteoporosis; OR=odds ratio; RCT=randomized, controlled trial; RR=risk ratio; SOE=strength of the evidence; SR=systematic review; ST/FS=subtrochanteric and femoral shaft; vs.=versus.

Evidence From Observational and Long-Term Extension Studies

Bisphosphonates. Recent observational studies from the Republic of Korea (k=1) and the United States (k=2) included primarily postmenopausal women and reported cases of AFF ranging from 46 to 113 in populations ranging from 6,000 to more than 94,000 individuals receiving BPs (**Appendix F.3 Table 2**).[464-466] Two studies included radiographic evaluation of fractures using blinded or dual assessment and used 2013 ASBMR criteria for AFF. Duration of BP use ranged from less than 1 year to 10 years.

The study conducted in the Republic of Korea reported that the incidence of AFF increased with duration of use from 31.2 per 100,000 person-years for short-term users to 67.1 per 100,000 person-years in long-term users ($p<0.001$).[464] The two studies conducted in the United States reported a similarly increased incidence of AF with duration of use.[465, 466] In these two studies reporting on an overlapping cohort of more than 80,000 Kaiser Permanente health plan users, for postmenopausal women incidence increased from 9 per 100,000 person-years with between 2 and 4 years of BP exposure to 112 per 100,000 person-years with 8 or more years of BP exposure in one of the studies.[465] In the second study from the United States, the adjusted cumulative AFF incidence in short-term (<3 years) BP users was 27 per 100,000 patient-years compared with 363 per 100,000 person-years in long-term (≥3 years) users.[466]

Denosumab. Few AFF were reported with denosumab use in studies reviewed for this CQ. In the multinational FREEDOM RCT and long-term extension (up to 7 years of denosumab after the 3-year RCT), 2 AFF occurred (0.8 per 10,000 participant-years): one in a participant receiving 7 years of denosumab and one in a crossover participant who had received denosumab for 3 years.[467]

Appendix F.3 Table 2. Recent Observational Studies Addressing Atypical Femoral Fracture and Bisphosphonate Use

Author, Year Country Study Design	Population and Number of AFF Cases	Assessment Method (AFF Criteria)	AFF Outcomes
Won et al 2020[464] Republic of Korea Controlled cohort study	36,529 women age 50 years or older initiating oral or IV BP between 2003 and 2011 (Korean National Health Insurance database) 61 AFF in long-term users (≥1 year, n=14,689) and 36 AFF in short-term users (<1 year, n=21,840)	ICD fracture codes and codes for fracture repair surgical procedures	Unadjusted IR in long-term users=67.1 per 100,000 person-years (95% CI, 50.3 to 83.9) and 31.2 per 100,000 person-years (95% CI, 21.0 to 41.4) in short-term users (p<0.001) aHR 2.34 (95% CI, 1.54 to 3.57) for long-term vs. short-term use NNH=400 (1 AFF per 400 women with long-term treatment)
*Lo et al 2020[466] U.S. Controlled cohort study	87,820 women between ages 45 and 84 years starting oral BP between 2002–2014 from a large health system; mean age=68.6 (SD 9.1) 46 AFF in 86,204 short-term (<3 years) users and 82,239 long-term (≥3 years) users; 32 of these AFF occurred after 3 years of use	Identification of fractures using ICD codes, blinded radiologic review (ASBMR 2013 criteria)	Cumulative AFF incidence remained stable in short-term users and increased in long-term users At 10 years, adjusted cumulative AFF incidence in short-term group=27 per 100,000 person-years (95% CI, 8 to 46) vs. 363 per 100,000 person-years (95% CI, 132 to 593) in long-term users Adjusted 10-year absolute risk difference=336 per 100,000 person-years (95% CI, 110 to 570)
*Lo et al 2019[465] U.S. Cohort study Note: Population overlaps with Lo 2020	94,542 women between ages 45 and 89 years initiating oral BP between 2002–2014 from a large health system; mean age=69.9 (SD 10) 113 AFF (107 occurring during BP exposure or <12 months after cessation); median BP exposure=2.2 years (0.5 to 5.0)	Identification of fractures using ICD codes, blinded radiologic review (ASBMR 2013 criteria)	22 AFF cases occurred with exposure <4 years vs. 85 cases with ≥4 years' exposure; 6 cases occurred 1 to 3.5 years after cessation Age-adjusted incidence at 2 to <4 years exposure=9 per 100,000 person-years, incidence at ≥8 years exposure=112 per 100,000 person-years Majority of AFF occurred in Asian women (62.8%) and non-Hispanic Whites (26.6%)

Author, Year Country Study Design	Population and Number of AFF Cases	Assessment Method (AFF Criteria)	AFF Outcomes
Bone et al 2017[467] FREEDOM RCT + open-label extension	2,343 women receiving denosumab in RCT and extension study and 2,207 women receiving placebo in RCT and denosumab in extension, mean age overall at enrollment=72.3 (SD 5.2) 2 AFF	Radiologic review (ASBMR 2010 or 2013 criteria, depending on study year)	1 AFF occurred in participant receiving denosumab in RCT and extension after 7 years of treatment, 1 in a crossover participant after 3 years of denosumab (exposure-adjusted incidence=0.8 AFF per 10,000 participant-years) No AFF reported in years 5–7 of extension study in either group

* Lo et al (2020) and Lo et al (2019) included an overlapping population of women from the Kaiser Permanente Northern California System.

Abbreviations: AFF=atypical femoral fracture; aHR=adjusted hazard ratio; ASBMR=American Society for Bone and Mineral Research; BP=bisphosphonates; CI=confidence interval; ICD=International Statistical Classification of Diseases and Related Health Problems; IR=incidence rate; IV=intravenous; n=number; NNH=number needed to harm; RCT=randomized, controlled trial; SD=standard deviation; vs.=versus.

Osteonecrosis of the Jaw

Definition. ONJ nomenclature has changed over time to reflect the agents with which ONJ has been associated (e.g., BPs, denosumab, tyrosine kinase inhibitors), which may complicate understanding of risk and incidence. In 2014, the American Association of Oral and Maxillofacial Surgeons (AAOMS) revised characteristics used to define medication-related ONJ[468]:

- Current or previous treatment with antiresorptive or antiangiogenic agents, and
- Exposed bone or bone in the maxillofacial region for longer than 8 weeks, and
- No history of radiation therapy to or obvious metastatic disease of the jaws.

ONJ pathogenesis is likely related to multiple factors including infection, immune system dysfunction, tooth extraction, smoking, poor oral hygiene, and use of antiresorptive or antiangiogenic medications.[460, 468, 469]

Evidence From Systematic Reviews and Seminal Reports

Bisphosphonates. An ASBMR report noted that the incidence of ONJ is rare: approximately 1 case per 10,000 to 100,000 person-years, with a largely self-limiting clinical course in patients with osteoporosis treated with BPs (**Appendix F.3 Table 3**).[460] Three SRs addressed the association between ONJ and BP use.[461, 463, 469] One SR conducted for the European Calcified Tissue Society noted incidence estimates in individuals using BPs ranging from 0.01 percent to 0.06 percent, with higher incidence among persons in Asian countries.[469] A review of SRs reported increased risk of ONJ with BPs vs. control in observational studies (ORs ranging from 2.57 to 3.29, low strength of evidence).[461] Another SR similarly reported increased risk for ONJ (with and without radiographic review) with alendronate vs. no treatment or raloxifene (HRs from 0.86 to 7.42); the review authors recalculated the reported HR estimate of 0.86 because incidence rates suggested a higher risk of ONJ. The recalculated estimate was 1.20 (95% CI, 0.59 to 2.56).[463]

Denosumab. One SR noted incidence of ONJ with denosumab use; the review cited the FREEDOM RCT and extension study which is discussed in the next section.[469] This review also cited a postmarketing study that reported 47 adjudicated cases of ONJ in 1,960,405 patient-years of denosumab exposure; all patients had risk factors for ONJ development.[469]

Appendix F.3 Table 3. Recent Reviews and Seminal Reports Addressing Osteonecrosis of the Jaw and Bisphosphonate or Denosumab Use

Author, Year (Search Dates)	Focus	Agents	Included Studies (N Cases/N Participants)	ONJ Outcomes
ASBMR Reports				
Adler et al 2016[460] (NR)	Safety of long-term BP use	BPs as a class	NA	Incidence rates range from 1 per 10,000 to 1 per 100,000 person-years Report noted a trend for increased ONJ risk with increased duration of BP use
Reviews				
Lu et al 2020[461] (NR-Dec. 2018)	Overview of SRs including meta-analyses of RCTs or observational studies reporting rare BP harms	Alendronate, ibandronate, etidronate, zoledronic acid, or risedronate	8 SRs addressing alendronate or BPs in general (NR)	Lee 2014 (BP vs. control, N=NR, 9 obs studies) OR=2.57 (95% CI, 1.37 to 4.84); GRADE: Low Lee 2014 (oral BP vs. control, N=NR, 5 obs studies) OR=3.29 (95% CI, 1.39 to 7.77); GRADE: Low
Fink et al 2019[463] (Jan. 1995-Oct. 2018)	SR of RCTs or obs studies of long-term (>3 years) use of OP drug treatment vs. control in men or women age ≥50	Alendronate, zoledronic acid, any BP	2 obs studies of alendronate vs. no treatment (28/220,894)	ONJ without X-ray or pathology review, HR=3.15 (95% CI, 1.44 to 6.87); SOE: Low
			1 obs study of alendronate vs. raloxifene (40/8,354)	ONJ with X-ray/pathology review, HR=7.42 (95% CI, 1.02 to 54.09); SOE: Low
			1 obs study of alendronate vs. raloxifene or calcitonin (46/43,645)	ONJ without X-ray, pathology review, HR=0.86 (95% CI, 0.44 to 1.69); authors recalculated as 1.20 (95% CI, 0.59 to 2.56); SOE: Insufficient
			1 RCT of zoledronate vs. placebo (0/2,000)	SOE: Not rated
Anastasilakis et al 2022[469] (NR)	ECTS “detailed review” of ONJ incidence and characteristics	BPs, denosumab	NR	Variable definitions of medication-related ONJ complicate incidence estimates; higher incidence with IV BPs, potentially because IV agents more often used in cancer patients, and greater incidence ONJ in patients with cancer vs. OP Incidence in persons taking BPs ranged from 0.01% to 0.06%, with higher incidence in persons from Asian countries Data from FREEDOM RCT[467] and extension study suggested an incidence rate of 5.2 per 10,000 person-years (based on 13 cases of ONJ observed). 47 cases of ONJ in 1,960,405 patient-years of denosumab exposure in postmarketing surveillance study; all patients had risk factors for ONJ such as invasive dental procedures, cancer

Abbreviations: ASBMR=American Society for Bone and Mineral Research; BP=bisphosphonates; CI=confidence interval; ECTS=European Calcified Tissue Society; GRADE=Grading of Recommendations, Assessment, Development, and Evaluation; HR=hazard ratio; IV=intravenous; N=number; NA=not applicable; NR=not reported; obs=observational; ONJ=osteonecrosis of the jaw; OP=osteoporosis; OR=odds ratio; RCT=randomized, controlled trial; SOE=strength of the evidence; SR=systematic review; vs.=versus.

Evidence from Observational and Long-Term Extension Studies

Bisphosphonates or Denosumab. As noted in **Appendix F.3 Table 4**, 13 cases of ONJ occurred in the FREEDOM RCT and open-label extension: seven in participants who received denosumab in the RCT and extension and six in participants who received placebo in the RCT and crossed over to denosumab (5.2 cases per 10,000 participant-years).[467] Other recent observational studies also addressed the association between ONJ and use of BPs or denosumab or denosumab alone. One study conducted in Switzerland reported a rate of ONJ cases of 4.5 per 10,000 BP users and 28.3 per 10,000 denosumab users, all of whom had been previously treated with BPs.[470] A 3-year Japanese postmarketing study reported 15 ONJ cases in 3,534 patients; six of these met AAOMS criteria for an incidence rate of 76.2 per 100,000 person-years.[471]

Appendix F.3 Table 4. Recent Observational Studies Addressing Osteonecrosis of the Jaw and Bisphosphonate or Denosumab Use

Author, Year Country Study Design	Population ONJ Cases	Assessment Criteria	Key Findings
Bone et al 2017[467] Multinational RCT + open-label extension (FREEDOM)	2,343 women receiving denosumab in RCT and extension study and 2,207 women receiving placebo in RCT and denosumab in extension, mean age overall at enrollment=72.3 (SD 5.2) 13 cases of ONJ	Adjudication using AAOMS definition	7 cases of ONJ occurred in participants who received denosumab in the RCT and extension; 6 in placebo crossover participants Exposure-adjusted incidence rate=5.2 per 10,000 person-years 11 cases resolved (2 lost to followup), 4 while on denosumab treatment
Everts-Graber 2022[470] Switzerland Case series/Registry	3,068 patients with ≥1 DXA scan and receiving BPs or denosumab between 2015–2019 seen at 1 outpatient center and included in Swiss Society of Rheumatology OP registry; median age=69 years (range=63 to 76 years) 17 cases identified: 12 in denosumab users (9 pretreated with BPs, mean 6.7 years' exposure) and 5 in oral or IV BP users (0 had prior denosumab)	Blinded assessment using AAOMS definition of ONJ	Incidence=28.3 per 10,000 person-years in denosumab users and 4.5 per 10,000 in BP users 9/17 patients with ONJ had risk factors including smoking, cancer, and aromatase inhibitor or steroid use; 9 of 12 patients with ONJ receiving denosumab had received prior BPs (mean BP treatment duration=6.7 years) HR for ONJ with denosumab vs. BP use: 3.49, 95% CI, 1.16 to 10.5, p=0.026 Time to ONJ healing ranged from 2 months to 3.5 years
Tanaka et al 2021[471] Japan Postmarketing analysis	3,534 patients receiving denosumab over 3 years; 1,643 discontinued over the followup period; mean age=75.7 years (SD 9.3) 15 cases of ONJ, 6 met AAOMS criteria	Diagnosis of osteomyelitis and/or ONJ using AAOMS definition	Based on adjudicated cases, IR=76.19 per 100,000 person-years (95% CI, 62.28 to 93.20) ONJ recovered or improved in 12 of 15 patients (3 others lost to followup) 13% of patients has secondary or drug-induced OP ONJ developed earlier in those receiving BPs prior to denosumab than those not receiving prior BPs (timing not reported)

Abbreviations: AAOMS=American Association of Oral and Maxillofacial Surgeons; BP=bisphosphonates; CI=confidence interval; DXA=dual-energy X-ray absorptiometry; HR=hazard ratio; IR=incidence rate; IV=intravenous; ONJ=osteonecrosis of the jaw; OP=osteoporosis; RCT=randomized, controlled trial; SD=standard deviation; vs.=versus.

F.4. Contextual Question 7

What is the evidence for rebound fractures after discontinuation of denosumab?

Summary

We identified recent (within the last 5 years) reviews and observational studies to address this CQ. We also included data from the seminal FREEDOM RCT extension analysis. Overall, studies included relatively few participants, and some included mixed populations of persons receiving denosumab for osteoporosis or for cancer-related bone problems. No consensus definition for “rebound fracture” currently exists. Study followup periods varied and typically did not exceed 24 months post-treatment cessation, while what authors classified as rebound fractures occurred from roughly 2 months’ to 16 months’ post-cessation. Analyses primarily from FREEDOM suggest that the risk of multiple vertebral fractures is increased relatively soon after treatment discontinuation and may be higher in persons with prior fractures.

In studies in which participants had a delay in denosumab dosing, higher fracture risk was similarly estimated to occur with a delay of as little as 4 months. A limitation across studies reporting on rebound fractures is that they were not designed to evaluate causality or estimate potential net benefits to denosumab over the long run, despite the occurrence of rebound fractures after treatment discontinuation.

Detailed Findings

Definition. Bone loss may rebound to levels experienced pretreatment when patients discontinue denosumab.[472] Rebound fractures, typically vertebral fractures, have been described as fractures that occur shortly after cessation of denosumab therapy; however, the timing of fracture occurrence is variable and no consensus definition exists. In the FREEDOM trial of denosumab, fracture assessment occurred in patients who received two to five doses of denosumab or placebo and continued study participation for at least 7 months after the study ended for a maximum of 24 months followup (mean 0.8 years/patient).[473, 474] Followup periods post-treatment cessation in other studies have generally not exceeded 20 months.

Association Between Rebound Fractures and Denosumab Discontinuation

The FREEDOM RCT and extension study[473, 474] assessed the incidence and risk factors for rebound fractures after denosumab use (**Appendix F.4 Table 1**). In a post hoc analysis including participants in both the RCT and extension study and analyzing vertebral fractures specifically, more denosumab discontinuers had multiple vertebral fractures vs. placebo discontinuers (60.7% vs. 38.7%) with rates of 4.2 per 100 in denosumab discontinuers and 3.2 per 100 in placebo discontinuers.[474]

Across two observational studies reporting time to fracture, months to fracture ranged from 1.8 to 16 after the last denosumab dose;[440, 441] in a third cross-sectional study, fractures occurred a median 12 months (mean 13 months) after the last injection[442] (**Appendix F.4 Table 1**). One dose-ranging study reported 17 fractures in eight participants: four women had multiple vertebral fractures, three had single vertebral fractures, and one had a radius fracture.[475] Additionally, in two studies, both including persons with cancer and persons with osteoporosis, more than 50 percent of patients had multiple vertebral fractures post-discontinuation.[476, 477]

In the FREEDOM RCT and extension study, risk factors for rebound fracture included prevalent vertebral fractures, greater gains and losses in hip BMD on therapy and after therapy, and longer duration off therapy. In this study, prior fracture was the strongest predictor of post-treatment fracture (OR 3.9 [95% CI, 2.1 to 7.2]).[474] In addition, the association between duration of denosumab therapy and rebound fracture was not clear. In one observational study, the number of injections was not significantly associated with rebound facture,[477] while in another, women taking denosumab for 2 or more years had more fractures than those taking denosumab for less than 2 years.[476] Both of these studies, however, included participants with cancer and osteoporosis.

Association Between Rebound Fractures and Delays in Denosumab Dosing

Recent studies have also evaluated the association between rebound fractures and delay in denosumab treatment (**Appendix F.4 Table 2**).[478, 479] A typical dosing schedule is every 6 months. Several studies evaluated delays ranging from 1 month to 4 months.[478] In one study, higher vertebral (but not other fractures) fracture rates were estimated with a delay in denosumab therapy of more than 16 weeks vs. treatment within 4 weeks of the last denosumab dose,[478] and fracture incidence rates were significantly increased in patients with a delay of at least 3 months vs. persistent users in a second study.[479]

Appendix F.4 Table 1. Recent Observational and Other Studies Addressing Denosumab and Rebound Fractures

Author, Year Country Study Design	Population (Mean Age)	Denosumab Use and Discontinuation	Key Findings
Trials/Trial Extensions			
Brown et al (FREEDOM) 2013[473] Multinational RCT Cummings et al (FREEDOM+ Extension) 2018[474] Multinational Post hoc, long-term extension study	FREEDOM RCT: 797 trial participants discontinuing denosumab (n=327) or placebo (n=470) after 2 to 5 doses; mean age 73 (SD 5) FREEDOM + 7-year extension: 1,471 participants discontinuing placebo (n=470) or denosumab (n=1,001) after ≥2 doses (age NR)	FREEDOM: Mean followup from last dose of denosumab or placebo + 7 mo)=0.8 years (median=0.5 year) FREEDOM + extension: median 0.2 years (IQR: 0.1 to 0.7)	FREEDOM RCT: 51 vertebral fractures post-treatment in placebo arm and 26 in denosumab In 470 FREEDOM placebo-treated subjects discontinuing placebo and followed for a total of 378 subject-years and 327 denosumab-treated subjects discontinuing treatment and followed for 267 subject-years, overall fracture rate per 100 subject-years=13.5 for placebo and 9.7 for denosumab (HR 0.82 [95% CI, 0.49 to 1.38]) No difference in time to fracture between placebo and denosumab groups in FREEDOM trial FREEDOM + 7-year extension: 31 vertebral fractures (12 multiple vertebral fractures) in placebo discontinuers and 56 vertebral fractures (34 multiple vertebral fractures) in denosumab discontinuers; 14 placebo discontinuers and 23 denosumab discontinuers had ≥1 nonvertebral fracture In post hoc analysis of 1,471 FREEDOM + extension patients, more denosumab users had multiple vertebral fractures post-discontinuation vs. placebo: 60.7% vs. 38.7% Off-treatment exposure-adjusted rate of any new vertebral fractures per 100 subject-years in placebo arm was 8.5 vs. 7.1 in denosumab; rate for multiple vertebral fractures was 3.2 per 100 subject-years in placebo discontinuers vs. 4.2 per 100 in denosumab discontinuers Rate of nonvertebral fractures was 3.8 per 100 (95% CI, 1.8 to 5.8) in placebo discontinuers and in denosumab discontinuers=2.8 per 100 subject-years (95% CI, 1.7 to 4.0) Prior fracture was strongest predictor of post-treatment fractures (OR=3.9); odds of multiple vertebral fracture were 1.6 times higher with each additional year of off-treatment followup
Observational Studies			
McClung et al 2017[475] U.S. (13 centers) Case series McClung et al 2017[475] U.S. (13 centers) Case series	82 women who had received denosumab in a 4-year phase 2 denosumab dose-ranging trial or 4-year extension study and ceased denosumab treatment (68.9)	52 women had received denosumab for 8 months prior to discontinuation (i.e., were in the denosumab arm in the dose-ranging parent trial and extension study); 12 had discontinued denosumab after 2–4 months and restarted for	8 participants had at least 1 post-treatment fracture, 17 total fractures. 4 women had multiple vertebral fractures, 3 had single vertebral fractures, 1 had radius fracture Among the 8 of 82 participants with an osteoporotic fracture in the 12-month post-denosumab followup period, 2/8 had history of prior fracture; all had fracture risk factors Time to fracture after stopping denosumab: Rebound fractures occurred from 1.8 to 13.1 months after last dose of denosumab

Author, Year Country Study Design	Population (Mean Age)	Denosumab Use and Discontinuation	Key Findings
(continued)		the extension study; 8 had discontinued alendronate and started denosumab in the extension study; 10 had taken placebo and started denosumab in the extension 12 months' followup post-discontinuation	Age at fracture ranged from 62 to 79. No participants were receiving OP treatment after stopping denosumab before fracture occurred Study also reported data from 2 women who did not participate in the 12-month followup but had participated in the dose-ranging trial: 1 had multiple vertebral fractures, and 1 had single fracture. Time to fracture after stopping denosumab: Fractures occurred 1 or 3.5 months post-discontinuation in participants 61 and 74 years old Spine radiographs not obtained in the dose-ranging trial; thus, it was not clear if post-treatment fractures were acute or chronic
Burckhardt et al 2021[477] Switzerland Cross sectional (Survey of 39 clinicians from hospitals across country conducted in 2019)	797 women with OP or nonmetastatic breast cancer receiving denosumab (65.3 years); 134 women had breast cancer (fracture incidence not reported separately for OP and cancer patients)	Mean injections=5.9 (range 2 to 20) Mean treatment duration=35 months (range 5 to 120) Mean followup post-discontinuation=27.5 months (SD 15.5)	215 post-treatment vertebral fractures in 82 women (mean 2.6 fractures; 69.5% with multiple fractures) and 16 post-treatment nonvertebral fractures (N women NR) Time to fracture after stopping denosumab: First fracture occurred mean 13 months/median 12 months after last injection; 75% occurred between 6 and 15 months after last injection Number of denosumab injections not significantly associated with rebound fracture occurrence BP use pre-denosumab or post-denosumab associated with lower risk of vertebral fracture and multiple vertebral fractures (HR for BP use pre-denosumab=0.24, for use after denosumab=0.042, for use before and after denosumab=0.048); greater protective effect with post-discontinuation use of BPs BP use post-denosumab discontinuation associated with lower incidence of nonvertebral fractures (0.08)
Anastasilakis et al 2017[476] Greece Review and case series	Total N considered in review NR; N of women considered eligible for case series NR	Time on denosumab ranged from 1 to 5 years, mean 2.9 years Duration followup NR	13 women with post-discontinuation fractures identified in literature search 11 women experiencing fractures in authors' centers Total of 112 fractures in 24 women after stopping denosumab (median 5.0 fractures, range 1 to 9) Time to fracture after stopping denosumab: All fractures occurred 8 to 16 months after last denosumab injection; 92% of patients had multiple vertebral fractures Women with ≥2 years denosumab duration had more fractures vs. those with ≤2 years (mean (SEM) fractures=5.2 (1.4) vs. 3.2 (0.7), p=0.055) 5 of 24 patients were receiving aromatase inhibitors for cancer, 1 was receiving glucocorticoids

Abbreviations: BP=bisphosphonates; CI=confidence interval; FREEDOM=Fracture REduction Evaluation of Denosumab in Osteoporosis Every 6 Months; HR=hazard ratio; IQR=interquartile ratio; n/N=number; NR=not reported; OP=osteoporosis; OR=odds ratio; RCT=randomized, controlled trial; SD=standard deviation; SEM=standard error of the mean; vs.=versus.

Appendix F.4 Table 2. Observational Studies Assessing Fracture After Delays in Denosumab Treatment

Author, Year Country Study Design	Population (Mean Age)	Denosumab Use and Discontinuation	Key Findings
Lyu et al, 2020[478] U.K. Cohort study	2,594 patients initiating denosumab for OP; mean age 75.8 (SD 9.5)	6,144 injections, Treatment delay defined as within 4 weeks of prior injection, 4–16 weeks delay, >16 weeks delay	Fracture risk within 4 weeks of prior denosumab injection: composite fracture 27.3 per 1,000 persons; MOF 14.7 per 1,000 persons; vertebral fracture 2.2 per 1,000 persons Fracture risk with delay of 4 to 16 weeks to next denosumab injection: composite fracture 32.2 per 1,000 persons; MOF 18.1 per 1,000 persons; vertebral 3.6 per 1,000 persons Fracture risk with >16 weeks delay in denosumab injection: composite fracture 42.4 per 1,000 persons; MOF 27.2 per 1,000 persons; vertebral 10.1 per 1,000 persons aHR for fracture between dosing within 4 weeks and delay of 4–16 weeks or >16 weeks were elevated but not significant except for vertebral fractures; aHR for vertebral fracture with delay of >16 weeks vs. within 4 weeks 3.91 (95% CI, 1.62 to 9.45)
Tripto-Shkolnik et al, 2020[479] Israel Cohort study	1,500 patients discontinuing denosumab treatment; mean age 72.4 (SD 9.6) at first denosumab purchase	Patients were included if they had at least 2 denosumab purchases; treatment discontinuation defined as refill gap of ≥3 months Post-discontinuation fractures defined as occurring within 1 year of discontinuation	54 of 1,500 patients had any MOF post-denosumab discontinuation (21 with any vertebral, 12 with multiple vertebral, 13 with hip, 22 with non-hip, nonvertebral fractures); incidence rate for any fracture 5.1 per 100 person-years (95% CI, 3.94 to 6.62) Fracture incidence rate per 100 person-years in discontinuers with prior vertebral fracture 6.81 (95% CI, 4.0 to 11.3) Higher rates of MOF, vertebral fractures, multiple vertebral fractures, hip fractures in discontinuers vs. persistent users: IRRs ranging from 2.23 to 14.63, all p≤0.005 Unadjusted HR for fracture within 1 year in discontinuers vs. persistent users was 2.5 (95% CI, 1.3 to 4.7, p=0.003) in patients with 2 denosumab purchases and 3.18 (95% CI, 1.6 to 6.5, p<0.001) in patient with 3 purchases; HRs not significant in those with 4 or 5 purchases

Abbreviations: aHR=adjusted hazard ratio; CI=confidence interval; HR=hazard ratio; IRR=incidence rate ratio; MOF=major osteoporotic fracture; OP=osteoporosis; SD=standard deviation; U.K.=United Kingdom; vs.=versus.

Appendix G. Risk of Bias in Development Cohorts of Fracture Risk Assessment Instruments for Key Question 2a

The tables included in this section offer risk-of-bias assessments for the development studies and cohorts for five of the six instruments included in addressing the KQ on the predictive accuracy of risk assessment instruments (KQ 2a).

- Fracture Risk Calculator (FRC)
- Fracture Risk Assessment Tool (FRAX)
- Fracture Risk Evaluation Model (FREM)
- Garvan Fracture Risk Calculator
- QFracture
- Women's Health Initiative Fracture Risk Model

Risk of bias was assessed using a modified version of the Prediction model study Risk Of Bias Assessment Tool (PROBAST[111, 112]). This instrument was modified to include additional health equity signaling items. These items as denoted with an "a" after the signaling item in the tables that follow.

The seventh risk assessment instrument included for KQ 2a (Osteoporosis Self-Assessment Tool, OST) was not developed as a fracture risk prediction instrument; therefore, we cannot assess the risk of bias of the development study or cohort for predicting fractures.

Appendix G Table 1. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Step 1 and 2 Specification of Review Questions

Item	Response
Intended use of model:	Predict risk of developing an osteoporotic fracture
Participants including selection criteria and setting:	Patients age 40 years or older
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Low body weight, current smoking, hip fracture in mother or sister, personal fracture history (nonspinal after age 50 years or older), with or without BMD
Outcome to be predicted:	Risk of osteoporotic fracture (with a minimum of 3 years' observation for studies with no specified prediction interval or a median or mean of 80% of the time in studies with a specified prediction interval)
Type of prediction study	Development and validation
Citations	Ettinger B, Hillier TA, Pressman A, Che M, Hanley DA. Simple computer model for calculating and reporting 5-year osteoporotic fracture risk in postmenopausal women. J Womens Health (Larchmt). 2005 Mar;14(2):159-71. doi: 10.1089/jwh.2005.14.159. PMID: 15775734.

Abbreviations: BMD=bone mineral density; FRC=Fracture Risk Calculator.

Appendix G Table 2. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response
Describe the sources of data and criteria for participant selection:	The authors used two different sources for data for development. The risk factors and three major osteoporotic nonspinal fractures (hip, wrist, humerus) came from 1996–2000 fracture incidence derived from inpatient and outpatient databases of the Kaiser Permanente Medical Care Program, Northern California Region. Because the membership rate for each age category remained constant during the study, the authors assumed the cohort of members at risk also remained constant per year. It is not stated that these were first fractures. They used these data to calculate the percentage of female members ages 45–75 years treated for each of the fractures for each year. But because of ICD-9 coding deficiencies, vertebral fractures were underdiagnosed and inaccurate. So for vertebral fractures, they relied on data from the Geelong Osteoporosis Study report (https://pubmed.ncbi.nlm.nih.gov/10525717/, which described a cohort study of persons living in Geelong, Australia, for whom fracture outcome was monitored during the 2-year study period (1994–1996). "The Geelong Osteoporosis Study is a population-based study designed to determine complete fracture rate within a defined region sufficiently large and representative of national demographics to establish reliable rates of fracture. The advantage of this population is that all radiologic facilities are provided through two centralized services." The data were limited to first fractures. The authors of FRC noted that they calculated 5-year vertebral fracture rates for White women but did not explain assumptions made in calculating 5-year rates from 2 years of data from the Geelong data. The model was then validated in 2 "large prospective observational studies," Study of Osteoporotic Fractures (SOF) (https://pubmed.ncbi.nlm.nih.gov/7862179/ and https://pubmed.ncbi.nlm.nih.gov/1952469/) and Canadian Multicentre Osteoporosis Study (CaMOS). SOF included 9,516 White women age 65 years or older who had no previous hip fracture from four clinical centers in the United States. The CaMOS source cited in the FRC article is not accessible (Kreiger N, Tenenhouse A, Joseph L, Mackenzie T, Poliquin S, Brown JP, et al. The Canadian multicentre osteoporosis study (CaMoS): background, rationale, methods. CJA. 1999;18:376–387.). But another CAMOS source (https://pubmed.ncbi.nlm.nih.gov/25451323/) seems to indicate that the baseline cohort of 6,314 women and 2,789 men from 1995–1997 of whom 94.9% were White.

Risk of Bias	Dev	Val
Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data? Incidence of risk factors came from a routine care database, as did the nonspinal fractures for the development cohort. The spinal fractures for the development cohort came from a prospective cohort designed to address fractures. Both validation cohorts were prospective and designed to address fractures	PN	NA
Were all inclusions and exclusions of participants appropriate? No information on exclusions	NI	NA
Was there sufficient representation of individuals from racial and ethnic groups in model development data? The nonspinal fracture data from Kaiser Permanente Medical Care program in the Northern California region included 70% non-Hispanic White females, 7.5% African American or Black females, 8% Latino or Hispanic females, and 13.5% Asian females. The vertebral fracture data were designed to include only White women.	PY for nonspinal fractures, N for spinal fractures	NA
Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only) The model validation cohorts were predominantly or exclusively White but the model did not include race.	NA	NA

Item	Response
Risk of bias introduced by selection of participants (low/high/unclear)	High The predictors and the nonspinal fracture data came from a routine care database, which is not ideal because data were not collected with standardized research protocol.

Appendix G Table 2. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Applicability	
Describe included participants, setting and dates:	Development cohort for nonspinal fractures from inpatient and outpatient databases of the Kaiser Permanente Medical Care Program so likely representative of wide population of women and include races other than White women. The spinal fracture development cohort and the two validation cohorts were large and prospective and intended to be broadly representative, except that two-thirds did not include men, and all were almost entirely White.
Concern that the included participants and setting do not match the review question (low/high/unclear)	Unclear Development and validation cohorts broadly representative of White women but not other women.

Abbreviations: CaMOS=Canadian Multicentre Osteoporosis Study; Dev=development; FRC=Fracture Risk Calculator; ICD-9=International Statistical Classification of Diseases and Related Health Problems, 9th Edition; ICD=International Statistical Classification of Diseases and Related Health Problems; NA=not available; NI=no information; PN=probably no; PY=probably yes; RCT=randomized, controlled trial; SOF=Study of Osteoporotic Fractures; Val=validation.

Appendix G Table 3. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response	
List and describe predictors included in the final model, e.g., definition and timing of assessment:	Risk factors included Age Height Weight Current smoker Prior nonspinal fracture No. of spinal fractures Hip fractures in sister Hip fractures in mother BMD It is unclear how the data were collected and what approaches were used to address missing data. 5-year followup was the primary time point for prediction.	
Risk of Bias	**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants? Routine care database so consistency of data collection for items such as family history-taking is unknown. Unclear how missing data were handled.	PN	
Were predictor assessments made without knowledge of outcome data?	PY	
Are all predictors available at the time the model is intended to be used?	PY	
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? Race and ethnicity were not used in the model.	Y	
Was differential missingness of predictor data in racial and ethnic groups considered? Not applicable, race was not used	NA	
Risk of bias introduced by predictors or their assessment (low/high/unclear)	High Because this is a routine care database, unclear whether predictors were defined as assessed in the same way for all participants. The handling of missing data was also unclear.	
Applicability		
Concern that the definition, assessment, or timing of predictors in the model do not match the review question (low/high/unclear)	Low These predictors are relevant to what would be collected in primary care and included in electronic medical records.	

Abbreviations: BMD=bone mineral density; Dev=development; NA=not available; PN=probably no; PY=probably yes; Val=validation.

Appendix G Table 4. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:	Two primary outcomes: 5-year nonspinal (hip, humerus, and wrist) fracture risk and 5-year spinal fracture risk. Nonspinal fractures came from the Kaiser database and were used to create a 5-year fracture risk outcome for any of 3 limb fractures; the arithmetic function described the relation between age and 5-year fracture risk of any one of the three limb fractures is RISK = 0.0433 * $e^{0.0703*AGE.}$ For spinal fractures, the authors acknowledged the deficiencies in recording vertebral fractures in their database and stated they used data from a population-based study to calculate the 5-year incidence of clinical spinal fracture in their model. That study collected all radiographic reports of first spinal fractures in Geelong, Australia; during a 2-year period; from these rates, the authors of FRC calculated 5-year rates using data for women ages 45 to 75 years. Based on the information in the appendix, it appears that they used the Geelong data to calculate risk of clinical spinal fracture as a function of age. (RISK= $0.0000000000005*AGE^{6.8436}$)

Risk of Bias	Dev	Val
Was the outcome determined appropriately? Nonspinal fractures from a routine care database. Outcome determination methods not specified by protocol so open to coding errors, particularly for humerus and wrist. Hip fracture data may be more accurate given the devastating nature of this outcome. Spinal fractures from a prospective cohort designed to measure fractures, using ICD-9 codes and imaging data.	PN for nonspinal fractures, PY for spinal fractures	NA
Was a prespecified or standard outcome definition used? From routine care database using ICD-9 codes (wrist [813.4x, 813.5x, 813.8x, 813.9x], hip [820.x], and humerus [812.x]), so these are not adjudicated or standardized so they will be open to coding errors. Again, hip fracture is probably okay given how devastating this outcome is.	PN for nonspinal fractures, PY for spinal fractures	NA
Were predictors excluded from the outcome definition? Yes	Y	NA
Was the outcome defined and determined in a similar way for all participants? Within each outcome, possibly, but there was not enough information to determine it. Across outcomes, no, because as noted earlier, fracture risk for nonspinal fractures was derived from ICD-9 data for which consistency is unclear.	PN	NA
3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered? Followup not reported, either overall or by race. Could possibly have differential censoring in different populations.	N	NA
Was the outcome determined without knowledge of predictor information?	Y	NA
Was the time interval between predictor assessment and outcome determination appropriate? 5-year horizon may be short to address some outcomes.	PN	NA
Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)? Actual fractures were used as the outcome. This study did not create separate risk categories by race.	PY	NA
Risk of bias introduced by the outcome or its determination *(low/high/unclear)*	High Outcomes ascertained from routine care database, so these were not adjudicated or standardized so they are open to coding errors. Predictors included in outcome definition. Unknown followup or percentage of individuals dying/getting censored.	

Applicability	
At what time point was the outcome determined:	The outcome was determined over 5 years, but the rate of risk was assumed to be same (other than by age) year-on-year.
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	Unclear for the composite outcome of nonspinal fracture risk.

Appendix G Table 4. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response
Concern that the outcome, its definition, timing or determination do not match the review question (low/high/unclear)	Low The outcome of a fracture as indicated on a medical record seems broadly applicable.

Abbreviations: Dev=development; FRC=Fracture Risk Calculator; ICD-9=International Statistical Classification of Diseases and Related Health Problems, 9th Edition; N=no; PN=probably no; PY=probably yes; Val=validation; Y=yes.

Appendix G Table 5. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	Development cohort: unclear, the Kaiser cohort included "more than 400,000 female members treated for any one of the three nonspinal fractures during the study period (59,628 with prior fracture); 59,772 new osteo fractures; 20,028 hip fractures; 14,528 fracture events, including 3,412 hip fractures." For the Geelong cohort, total women ages 45 to 75 years=32,566, number of vertebral fractures=116 Predictors: 8 variables, but they combined height and weight into BMI with a cut point of <21 kg/m^2 to indicate thinness. Each of the predictors appears to have been weighted by relative risks obtained from various other sources (the authors cited 5 references for each clinical factor). Specifically, the relative risk for thinness for wrist and hip fracture risk was 1.3, current smoker for hip fracture risk was 1.5, mother's hip fracture for hip fracture risk was 1.3, sister's hip fracture for hip fracture risk was 1.6, prior spinal fracture for hip fracture risk was 1.5, prior fractures for spinal fracture risk was 3.2 for 1 prior fracture and 8.0 for 2 or more prior fractures, for nonspinal fracture risk was 1.6 for 1 prior fracture and 2.0 for 2 or more, each SD for BMD for spinal fracture was 2.0 and for nonspinal fracture was 1.5. The relative risk came from the following sources: Hemenway D, Colditz GA, Willett WC, Stampfer MJ, Speizer FE. Fractures and lifestyle: Effect of cigarette smoking, alcohol intake, and relative weight on the risk of hip and forearm fractures in middle-aged women. Am J Public Health 1988;78:1554. Cooper C, Wickham C. Cigarette smoking and the risk of age-related fractures. In: Wald NJ, Baron JA, eds. Smoking and hormone-related disorders. Oxford, New York: Oxford University Press, 1990;93. Forsén L, Bjørndal A, Bjartveit K, et al. Interaction between current smoking, leanness, and physical inactivity in the prediction of hip fracture. J Bone Miner Res 1994;9:1671. Fox KM, Cummings SR, Powell-Threets K, Stone K. Family history and risk of osteoporotic fracture. Study of Osteoporotic Fractures Research Group. Osteoporos Int 1998;8:557. Wasnich RD, Davis JW, Ross PD. Spine fracture risk is predicted by nonspine fractures. Osteoporos Int 1994;4:1. Note that these were not adjusted for each other so likely there was a lot of overlap in risk between the factors (which probably explains why the model overestimated risk so much in the validation studies). Also note that they did not always have RR for spinal and nonspinal factors separately and applied RRs for hip fractures to the other fractures. Then, because the model compared the risk in women who have the risk factor with the risk in the entire female population of the same age (some with and some without the risk factor), RRs needed to be adjusted for prevalence of the risk factor in the population. Using reported osteoporotic fracture rates in another reference (which is listed in what appears to be an error as the Geelong database, but is likely the next reference in the list that is specific to smoking), they estimated population prevalence of the risk factor (prior fracture) by age. Then, they determined the best-fitting regression equation (defined as R^2=0.99) and used this equation to estimate population prevalence of prior fracture in women at various ages. PThinness = ((0.021 * (AGE2)) - (2.6 * AGE) + 92.9)/100 PSmoking = (-0.001 * (AGE2) - 0.1 * AGE + 20.1)/100 PNonspinal fracture = (0.038 * (AGE2) - 3.4 * AGE + 83.5)/100 P1 Spinal fracture = (0.013 * (AGE2) - 0.8 * AGE + 12.3)/100 P_2 Spinal fractures = (0.015 * (AGE2) _ 1.5 * AGE + 39.9)/100 PSister with hip fracture= (0.004 * (AGE2) _ 0.4 * AGE + 9.5)/100 PMother with hip fracture= (-0.005 * (AGE2) + 0.9 * AGE - 31.8)/100
Describe how the model was developed (for example, in regards to modeling technique [e.g., survival or logistic modeling], predictor selection, and risk group definition):	Relative risks from background documents and prevalence estimates from databases were used to adjust the weights of predictors in the model, which then predicted fracture risk (which was itself a function of age). The exact model (combination of factors) was not described. Predictor selection was not described.

Appendix G Table 5. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross-validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	The model was validated externally using two large prospective cohort studies that were designed to study fractures, as noted above.
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	Calibration plot: Predicted vs. observed risk for any one of three nonspinal fractures for every 2.5 percentile intervals of predicted risk from <2.5 to 10+ in two cohorts at 5 years in two cohorts (CaMOS and SOF) Predicted vs. observed risk interval for spinal fractures for every 2.5 percentile intervals of predicted risk from <2.5 to 10+ at 5 years in one cohort (CaMOS) and mean of 3.7 years in a second cohort (SOF)
Describe any participants who were excluded from the analysis:	NR
Describe missing data on predictors and outcomes as well as methods used for missing data:	NR

Risk of Bias	Dev	Val
Were there a reasonable number of participants with the outcome? Kaiser cohort: 59,772 new osteoporotic fractures; 20,028 hip fractures; 14,528 fracture events, including 3,412 hip fractures Geelong cohort: 116 vertebral fractures	Y for nonspinal fractures, N for vertebral	NA
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) NA, model performance was not reported separately by race.		NA
Were continuous and categorical predictors handled appropriately? BMI was entered as a categorical variable, others were handled as expected (categorical or continuous).	PY	NA
Were all enrolled participants included in the analysis?	NI	NA
Were participants with missing data handled appropriately?	NI	NA
Was selection of predictors based on univariable analysis avoided? Regression coefficients (log of HR) from final models were used as weights (Tables 3 and 4). The relative risks were unadjusted for other variables.	N	NA
Risk of Bias (continued)	**Dev**	**Val**
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Not a competing risk model. Given that those with advanced age were in the relevant population, competing risk of death from any cause could be important. The study did specify a shorter prediction interval than 10-year predictors but did not report how many people died and were censored on that basis.	N	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? No race/ethnicity differences reported, not a competing risk model.	NA	NA
Were relevant model performance measures evaluated appropriately? Only calibration was reported.	PN	NA

Appendix G Table 5. Fracture Risk Calculator (FRC) Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response	
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups? Model performance measures not reported separately in different racial and ethnic groups.	N	NA
Were model overfitting and optimism in model performance accounted for? No	N	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? There is no final model and no multivariate analysis	N	NA
Risk of bias introduced by the analysis (low/high/unclear)	High Relative risks unadjusted for other variables. Small number of vertebral fractures, no information on missingness, only calibration was reported	

Abbreviations: BMD=bone mineral density; BMI=body mass index; CaMOS=Canadian Multicentre Osteoporosis Study; Dev=development; FRC=Fracture Risk Calculator; HR=hazard ratio; N=no; NA=not available; NR=not reported; PN=probably no; PY=probably yes; RR=relative risk; SD=standard deviation; SOF=Study of Osteoporotic Fractures; Val=validation; Y=yes.

Appendix G Table 6. Fracture Risk Assessment Tool (FRAX) Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Step 1 and 2

Item	Response
Intended use of model:	Predict risk of developing an osteoporotic fracture
Participants including selection criteria and setting:	Patients age 40 years or older
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Demographic information and clinical and family history with or without BMD
Outcome to be predicted:	Risk of osteoporotic fracture (with a minimum of 3 years' observation for studies with no specified prediction interval or a median or mean of 80% of the time in studies with a specified prediction interval)
Type of prediction study	Development and validation
Citations	Kanis JA on behalf of the World Health Organization Scientific Group (2007). Assessment of osteoporosis at the primary health care level. Technical Report. World Health Organization Collaborating Centre for Metabolic Bone Diseases, University of Sheffield, U.K. 2007: Printed by the University of Sheffield. Kanis JA, Johansson H, Oden A, Dawson-Hughes B, Melton LJ 3rd, McCloskey EV. The effects of a FRAX revision for the USA. Osteoporos Int. 2010 Jan;21(1):35-40. doi: 10.1007/s00198-009-1033-8. Epub 2009 Aug 25. PMID: 19705047. Melton LJ 3rd, Crowson CS, O'Fallon WM. Fracture incidence in Olmsted County, Minnesota: comparison of urban with rural rates and changes in urban rates over time. Osteoporos Int. 1999;9(1):29-37. doi: 10.1007/s001980050113. PMID: 10367027.

Abbreviations: BMD=bone mineral density; FRAX=Fracture Risk Assessment Tool; U.K.=United Kingdom.

Appendix G Table 7. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response
Describe the sources of data and criteria for participant selection:	Describe the sources of data and criteria for participant selection: The 2007 WHO Technical Report describes that candidate risk factor data from 12 “prospectively studied” cohorts were used for model development. In total, these cohorts involved 59,644 participants, 1,141 hip fractures, and 2,218 osteoporotic fractures (Table 5.3). Complete information was available from all cohorts for continuous variables of BMI and BMD, but not all cohorts had complete information on all dichotomous risk factors (explicated in Table 5.4; for example, only 3/12 provided alcohol or arthritis predictor data). Subsequently, in 2009 there was a publication describing a revision based on updated fracture incidence and mortality rates for the U.S. Those updated sources are described below. However, the methods behind model updating were exceedingly unclear (further discussed described in analysis domain). Hip fracture (FRAX 3.0, U.S. model): National hospital discharge data for White non-Hispanic women and men in 2006 from the Healthcare Cost and Utilization Project, inpatient sample. Incidence in 1-year age intervals was calculated from U.S. Census projections for 2006. Fracture rates were assumed to be a constant ratio of those in White population for other groups (see Discussion in article for this). Other osteoporotic fractures (FRAX 3.0, U.S. model): 1989–1991 Olmsted County, MN, data for fracture-specific incidence rates: 2,901 county residents age 35 years or older experienced 3,665 separate fractures during the 3-year period (2,362 experienced a single fracture). Population-based database study (Rochester Epidemiology Project), Mayo Clinic and common medical record system with 2 large, affiliated hospitals. The diagnoses and surgical procedures recorded in records were indexed. The index included the diagnoses made for outpatients seen in office or clinic consultations, emergency room visits, or nursing home care, as well as the diagnoses recorded for hospital inpatients, at autopsy examination or on death certificates. Medical records of the other providers who served the local population, most notably the Olmsted Medical Group, were also indexed. The complete (inpatient and outpatient) medical records were reviewed for all local residents with any diagnosis attributable to rubrics 800 through 829 in the International Classification of Diseases. Of 9,260 potential cases, record review was completed on all but 74 (0.8%) who had not provided an authorization for review of their medical records in accordance with a Minnesota privacy law that took effect in January 1997. All fractures were radiologically confirmed, but original radiographs were not available for review. Searched for fracture diagnoses made by any provider in any setting. Patients attended for complications of fractures prior to study period were excluded (other than this we do not see any discussion of inclusion and exclusion). Mortality (FRAX 3.0, U.S. model): Age-, sex-, and race-specific death rates for 2004 (CDC Vital Statistics) in 1-year intervals for the White population and 5-year intervals for other groups (see Discussion for interval information). Mortality records were based on information reported on death certificates as completed by funeral directors, attending physicians, medical examiners, and coroners.

Appendix G Table 7. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

<table>
<tr><th>Item</th><th colspan="3">Response</th></tr>
<tr><td colspan="2">Risk of Bias</td><td>Dev</td><td>Val</td></tr>
<tr><td colspan="2">Were appropriate data sources used, e.g., cohort, RCT, or nested case-control study data?
Fracture data were retrospectively collected from routine care databases.
Hip fracture: Nationally representative, all-payer hospital discharge data. Based on the Burge paper cited in the 2009 revision development publication, hip fracture incidence rates were defined using primary ICD-9 codes 820.0x, 820.2x, 820.8x, closed only and excluding trauma-related cases according to E-codes. These outcome data will not be standardized/adjudicated in the same way that a prospective cohort study is doing this (they were also not longitudinally linked so cannot verify first vs. recurrent fracture).
Other osteoporotic fractures: Rochester Epidemiology Project. Population-based database study. The WHO Technical Report described the 12 contributing cohorts as prospective, but in fact these data were retrospectively collected from a routine care database. Unclear which trauma definitions were used for this outcome.
Mortality: National Vital Statistics Report</td><td>PN</td><td>NA</td></tr>
<tr><td colspan="2">Were all inclusions and exclusions of participants appropriate?
Very sparse reporting. We just learn from the Rochester data that patients who attended for complications of fractures prior to study period were excluded. Unlikely that these individuals were pharmacologically treated (e.g., bisphosphonates) given the time period.</td><td>NI</td><td>NA</td></tr>
<tr><td colspan="2">Was there sufficient representation of individuals from racial and ethnic groups in model development data?
Hip fracture incidence data were exclusively from White non-Hispanic populations in HCUP. This is confusing because hip fracture incidence is available by racial and ethnic group in this dataset.
Fracture-specific incidence rates (other than hip) were taken from predominately White Olmstead County, MN, 1989–1991 sample. Another publication from the Rochester study reported that 99.1% of Olmstead County is White.
Mortality was race specific. Assumption is that the death hazard function is the only one that has true variation in race; the other outcomes are based on White participants exclusively (hip fracture) or an assumed ratio of events (other fractures).
In the original development paper, the racial and ethnic distribution of participants in the 12 cohorts was not reported. Cohorts were from Europe, Canada, Australian, Scandinavia, the U.S., and Japan. The one U.S. cohort is from Olmstead County, MN, which is 99.1% White.</td><td>N</td><td>NA</td></tr>
<tr><td colspan="2">Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only)
The online tool provides options for “Caucasian,” Black, Hispanic, and Asian.
For the U.S.-based external validations noted in the 2007 WHO Technical Report, Study of Osteoporotic Fractures had 99.7% White participant population. Women’s Health Initiative used classifications of White (87%), Black, Hispanic, American Indian/Alaska Native, Asian/Pacific Islander.
It is unclear how ethnicity was handled or which calculator that American Indian/Alaska Native populations should use.</td><td>PY/NI</td><td>NA</td></tr>
<tr><td>Risk of bias introduced by selection of participants (low/high/unclear)</td><td colspan="3">High
Exceedingly sparse reporting about inclusion and exclusion of participants (and whether inclusion and exclusion were consistent between hip and other fracture sources, which are different). Routine care databases used instead of prospective longitudinal cohorts in model revision.</td></tr>
<tr><td>Applicability</td><td colspan="3"></td></tr>
<tr><td>Describe included participants, setting and dates:</td><td colspan="3">2006 for hip fracture rates
1989–1991 for other fracture rates—from predominately White Olmstead county, MN
2004 for mortality: age, sex, and race specific</td></tr>
</table>

Appendix G Table 7. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response
Concern that the included participants and setting do not match the review question (low/high/unclear)	High Data are quite dated. Hip fracture data use estimates from White populations only, despite separate reporting by race and that the WHO Technical Report acknowledged that fracture rates are heterogeneous in different populations and settings. The very restricted geographical sample of Olmstead County, MN, is also concerning in its applicability. The 2009 revision paper stated that FRAX 3.0 is calibrated to the U.S., but the other fractures equation is “calibrated” to Olmstead County, MN.

Abbreviations: BMD=bone mineral density; BMI=body mass index; CDC=Centers for Disease Control and Prevention; Dev=development; FRAX=Fracture Risk Assessment Tool; HCUP=Healthcare Cost and Utilization Project; ICD-9=International Statistical Classification of Diseases and Related Health Problems, 9th Edition; MN=Minnesota; N=no; NI=no information; PN=probably no; PY=probably yes; RCT=randomized, controlled trial; U.S.=United States; Val=validation; WHO=World Health Organization.

Appendix G Table 8. Fracture Risk Assessment Tool (FRAX)Fracture Risk Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response		
List and describe predictors included in the final model, e.g., definition and timing of assessment:	4 models: Probability of hip fracture with BMD Probability of hip fracture without BMD Probability of other osteoporotic fracture with BMD Probability of other osteoporotic fracture without BMD Predictors: Age (continuous), BMD (continuous), BMI (continuous), parental history of hip fracture, prior fragility fracture, current smoking, oral glucocorticoids, rheumatoid arthritis, >2 units of alcohol/day. Time horizon for prediction is 10 years, but initial and revision papers did not clearly report the time horizon of the contributing cohorts. Reported as 252,034 patient-years of followup for all 12 cohorts combined. BMD was entered as a densitometer-specific BMD or as a T-score. The transformation to a T-score was derived from the NHANES III database for White females ages 20 to 29 years.		
Risk of Bias		**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants? Pg 92 in the WHO Technical Report: A history of current or past smoking was obtained by self-report. There was inadequate information to assess possible dose-response effects. The assessment of alcohol intake differed between cohorts and was converted into a daily intake expressed as units/day. A unit of alcohol is equivalent to 8 g in the United Kingdom, though varies somewhat in different countries. A family history was collected of any fracture in first-degree relatives. In addition, a family history of hip fracture was noted but was available only in three of the cohorts. Prior fracture history of each individual was documented, though the construct of the question varied, particularly the age from which a fracture had occurred. Use of oral glucocorticoids ever during a person's lifetime (ever use) was used to characterize steroid exposure because all but three cohorts did not distinguish between ever use and current use. Neither the dose nor the duration of use was analyzed. The presence or absence of rheumatoid arthritis was by self-report.		N	NA
Were predictor assessments made without knowledge of outcome data? Fracture ascertainment was by self-report in 6 contributing cohorts and many of the contributing predictors are also by self-report (see item 2.1 above). Those with a prior fracture may be more aware of their risk factor status.		N	NA
Are all predictors available at the time the model is intended to be used? Theoretically, yes this could be done. But not all cohorts had complete information on all dichotomous risk factors (section 5). For example, history of smoking and alcohol use not available from Rochester. Table 5.4		Y	NA
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? The WHO Technical Report noted a dramatic heterogeneity of age-adjusted and sex-adjusted incidence for hip fracture in various regions in the world and noted that in the U.S. a higher hip fracture rate in White individuals compared with Black individuals may be in part based on BMD differences, but that BMD differences do not explain lower rates in Hispanic and Asian populations. Because equations are available with and without BMD as an input, race/ethnicity was included for calibration rather than as a proxy. The report also noted that country-level differences in fracture risk may be largely attributable to differences in life expectancy because fracture risk is exponentially higher at older ages. Life expectancy differences were captured in the age-, sex-, and race-specific mortality data used in competing hazards model. But that mortality data are from 2004. On the other hand, pg 113 also states that "the frequency of falling is less in Black people than among Whites, as is the risk of fracture, which might indicate an important genetic factor related to falls. It was not possible to investigate other important skeletally-related factors such as the size and shape of bone or the microarchitecture of trabecular elements in cancellous bone." There were some mixed messages about why race and ethnicity were included in the model.		PY	NA
Was differential missingness of predictor data in racial and ethnic groups considered? Not addressed. There is a lot of missing predictor data regardless of racial and ethnic group distribution (whole predictors are missing in many cohorts). There is not much missing opportunity for missing data in BIPOC populations because of very little representation.		NI	NA

Appendix G Table 8. Fracture Risk Assessment Tool (FRAX)Fracture Risk Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response
Risk of bias introduced by predictors or their assessment (low/high/unclear)	High Self-reported data for many predictors; because fracture assessment was also self-report in many cohorts, this is not independent. Those with a previous fracture might be more aware of their risk factors. Dramatic amount of missing predictor data among cohorts (Table 5.4 in initial development paper). Unknown followup time.
Applicability	
Concern that the definition, assessment or timing of predictors in the model do not match the review question (low/high/unclear)	High Unknown followup time. BMD was transformed into a T-score using reference data from White women ages 20 to 29 years.

Abbreviations: BIPOC=Black, Indigenous, and People of Color; BMD=bone mineral density; BMI=body mass index; Dev=development; N=no; NHANES=National Health and Nutrition Examination Survey; NI=no information ; PY=probably yes; U.S.=United States; Val=validation; WHO=World Health Organization; Y=yes.

Appendix G Table 9. Fracture Risk Assessment Tool (FRAX) Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:	10-year probability of hip fracture or major osteoporotic fracture (hip, clinical spine, shoulder, or wrist). These sites are considered osteoporotic on the general definition that fractures from low-energy trauma can be osteoporotic and that a low-energy trauma would not give rise to a fracture in a healthy individual. Apparently, the coding of fracture sites as osteoporotic in the U.S. Is based on expert opinion. Note that these equations predict both risk of first fracture and subsequent fracture. Hip fracture (FRAX 3.0, U.S. model): National hospital discharge data for White non-Hispanic women and men in 2006 from the Healthcare Cost and Utilization Project, inpatient sample. Incidence in 1-year age intervals was calculated from U.S. Census projections for 2006. For other racial and ethnic groups, fracture rates were assumed to be a constant ratio of those in White populations (see Discussion in 2009 paper and pg 195 in original development paper). Other osteoporotic fractures (FRAX 3.0, U.S. model): 1989–1991 Olmsted County, MN, data for fracture-specific incidence rates: 2,901 county residents age 35 or older experienced 3,665 separate fractures. These were used for the revised model with the exception of vertebral fracture. In the case of vertebral fracture, the version 2.0 estimates comprised not only symptomatic (i.e., clinical) vertebral fractures but also included those found incidentally during routine medical care. In the absence of robust empirical data for the incidence of clinically significant vertebral fractures, for the revised model, it has been assumed that the ratio of clinical vertebral fractures to hip fractures in the U.S. was the same as that from Malmo, Sweden, a methodology used for the construction of FRAX. The removal of incidental or nonclinical vertebral fractures in the revision will reduce the estimated 10-year probability of a major fracture. Mortality (FRAX 3.0, U.S. model): Age-, sex-, and race-specific death rates for 2004 (CDC Vital Statistics) in 1-year intervals for the White population and 5-year intervals in other groups (see Discission for interval information). Mortality records were based on information reported on death certificates as completed by funeral directors, attending physicians, medical examiners, and coroners.

Risk of Bias	Dev	Val
Was the outcome determined appropriately? Clinical vertebral fractures not actually measured; they were just assumed to be a ratio of hip fractures and that ratio was taken from Sweden. Ascertainment of vertebral fracture is problematic because not all vertebral fractures come to clinical attention. Individuals with more access to imaging are going to have a higher ascertainment rate of incidental findings. For racial and ethnic groups other than White populations, hip fracture rates were not actually measured but were assumed to be a constant ratio of total fractures. Fractures obtained by self-report in half the cohorts.	N	NA
Was a prespecified or standard outcome definition used? Hip fracture: Based on the Burge paper cited in the 2009 revision development publication, hip fracture incidence rates were defined using primary ICD-9 codes 820.0x, 820.2x, 820.8x, closed only, and excluding trauma-related cases according to E-codes. In the original development paper, a mix of self-reported and verified fractures was used (see 3.4 for more detail). Other osteoporotic fracture: ICD 800-829 In no dataset were outcomes adjudicated or standardized, so they are open to coding errors.	PN	NA
Were predictors excluded from the outcome definition? However, equations predict either first or subsequent fracture, so event index bias is likely present. This is a problem because the coefficients for recurrent events are likely different than first events. After a first event, individuals may modify risk factors (use of steroids, alcohol, etc.) so the assumption that the weighting of the coefficients will be the same in first event/recurrent event individuals is flawed. This also ignores the natural history of fractures (small bones break first before hip or vertebral involvement).	Y	NA

Appendix G Table 9. Fracture Risk Assessment Tool (FRAX) Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Risk of Bias (continued)	Dev	Val
Was the outcome defined and determined in a similar way for all participants? In the main development paper, there is a mix of self-report and verification of outcomes from databases: Fracture ascertainment was undertaken by self-report in 6 cohorts (Sheffield, EVOS/EPOS, Hiroshima, Kuopio, EPIDOS, OFELY) or verified from hospital or central databases in 8 (Gothenburg, CaMos, DOES, Kuopio, Sheffield, EVOS/EPOS, Rochester, Rotterdam). However, the updated "calibration" is using Rochester and HCUP.	N	NA
3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered? Followup of cohorts not reported, either overall or by group.	NI	NA
Was the outcome determined without knowledge of predictor information? In some studies, fracture ascertainment was self-reported and predictor information was also self-reported.	N	NA
Was the time interval between predictor assessment and outcome determination appropriate? 10-year horizon is reasonable; however, we do not know how long followup was in these cohorts from the primary paper.	N	NA
Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)? Hip fracture is hospitalized hip fracture, which seems reasonable given that most hip fractures are hospitalized. For other fractures, diagnosis codes are from a wide range of settings (inpatient, outpatient, office or clinic consultations, ER, nursing home care, autopsy exam or death certificates). Fractures identified incidentally are likely to be differentially ascertained in populations with greater access to imaging. The model revision (3.0) focused on symptomatic vertebral fractures, which is a change from 2.0, which also included incidental findings.	PY	NA

Risk of bias introduced by the outcome or its determination (low/high/unclear)	High Many outcomes were not actually measured or verified. Predictor and outcome ascertainment not always independent. Followup time and loss to followup unknown. Event index bias was likely present because the model included both first and recurrent fractures; the weighting of coefficients in each situation was likely different.

Applicability	
At what time point was the outcome determined:	Unknown. 10-year prediction model but we do not know mean followup in years from these cohorts.
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	Separate models for hip fracture and other fractures. Frequency of the Olmstead County other fractures reported in the Melton paper, but it is unclear how that is being used in the model revision. There are various distributions for each type of fracture reported in initial paper.
Concern that the outcome, its definition, timing or determination do not match the review question (low/high/unclear)	High Followup/timing issues. Prediction of first and recurrent fracture in the same model is problematic.

Abbreviations: CaMOS=Canadian Multicentre Osteoporosis Study; CDC=Centers for Disease Control and Prevention; Dev=development; DOES=Dubbo Osteoporosis Epidemiology Study; ER=emergency room; FRAX=Fracture Risk Assessment Tool; HCUP=Healthcare Cost and Utilization Project; ICD=International Statistical Classification of Diseases and Related Health Problems; ICD-9=International Statistical Classification of Diseases and Related Health Problems, 9th Edition; MN=Minnesota; N=no; NI=no information ; OFELY=Os des Femmes de Lyon; PN=probably no; PY=probably yes; U.S.=United States; Val=validation; Y=yes.

Appendix G Table 10. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	4 models: prediction of hip fracture with and without BMD; prediction of other osteoporotic fractures (excluding hip fracture), with and without BMD (model without BMD) Age (continuous), BMI (continuous), parental history of hip fracture, prior fragility fracture, current smoking, oral glucocorticoids, rheumatoid arthritis, >2 units of alcohol/day. +interactions with sex, age, quadratic of each variable (8*3); 24+8=32 (model with BMD) Age (continuous), BMD, BMI (continuous), parental history of hip fracture, prior fragility fracture, current smoking, oral glucocorticoids, rheumatoid arthritis, >2 units of alcohol/day. +interactions with sex, age, BMD, quadratic of each variable (9*4); 36+9=45 These represent a conservative count of candidate predictors. WHO Technical Report pg 10 also states that contraceptive pills, age at menopause, age at menarche, hysterectomy, diabetes, and consumption of milk (6 additional predictors) were also considered, but it is unclear how far these were tested with interactions, so we omitted from the counts above. The events per variable were already pretty high so it probably makes little difference. There is also an additional situation with other secondary osteoporosis, where if the field for RA was entered as no but yes for secondary osteoporosis, the same function as used for RA was applied to the situation where BMD is not entered. If BMD was entered, no additional risk was assumed in the presence of secondary osteoporosis since independence of BMD was uncertain. In addition to using a calculator for the full score, simplified paper risk stratification tables are available. Fundamentally, it is unclear how race was included in the model. Assuming stratified because it is not listed as a candidate predictor variable. Also, we assumed it is just the death hazard function using actual race-specific data—hip fractures in BIPOC populations were never measured; instead a correction factor was applied to the rate in White populations.

Appendix G Table 10. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe how the model was developed (for example, in regards to modeling technique [e.g., survival or logistic modeling], predictor selection, and risk group definition):	Competing hazards. For each model, fracture and death as continuous hazard functions were computed using a Poisson regression. The effect of the candidate risk factor, age, and sex on the risk of any fracture, any osteoporotic fracture, and hip fracture alone was examined using Poisson regression models in each cohort separately. A Poisson model was chosen since it has greater power than logistic regression and can accommodate all information with variable durations of followup. In addition, time can be accommodated as an interaction term, and for some risk factors, relative risk may decrease with longer durations of observation. For each risk factor studied, covariates included current age and time since followup with and without BMD. Where appropriate, interaction terms were included. Outcome variables comprised any fracture, any osteoporotic fracture, and hip fracture alone. The results of different cohorts (men separate from women) were then merged using weighted coefficients and fixed effects model used. WHO Technical Report pg 10: Risk factors recommended for use were selected on the basis of their international validity and evidence that the identified risk was likely to be modified by subsequent intervention (modifiable risk). Modifiable risk was validated from clinical trials (BMD, prior fracture, glucocorticoid use, secondary osteoporosis) or partially validated by excluding interactions of risk factors on therapeutic efficacy in large randomized intervention studies (e.g., smoking, family history, BMI). A further step was then to merge these meta-analyses of each risk factor so that account could be taken of the interdependence of the risk factors chosen and, therefore, the risk provided by any combination of risk factors with and without the additional use of BMD. For each risk factor, all significant interactions (p<0.05 based on Table 7.13) that were identified by meta-analysis were entered in the model (with age, time, sex, and the risk factor) with and without BMD. Interactions that were significant for hip fracture risk were also entered in the model for other osteoporotic fractures and also included in the model for death. Where interactions noted in the “mega-analyses” were no longer significant for hip fracture and other osteoporotic fractures, these were omitted in a stepwise manner by dropping the interaction with the largest p-value. For the death hazard, all significant interactions for fracture risk were included and thereafter omitted if appropriate in a stepwise manner, as described for the fracture hazard.
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross-validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	No internal validation noted. Pg 205: evaluated in 11 independent cohorts. Study of Osteoporotic Fracture and WHI from U.S.
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	No calibration data; this is a major issue. Discrimination reported with AUC. No assessment of optimism.
Describe any participants who were excluded from the analysis:	Unclear.
Describe missing data on predictors and outcomes as well as methods used for missing data:	In original development dataset, many cohorts did not have measurement/data for all risk factors (Table 5.4). Apparently, this was handled by setting the coefficient to 0 and calculating score on basis of available risk factors.

Appendix G Table 10. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item / Response	Dev	Val
Risk of Bias	**Dev**	**Val**
Were there a reasonable number of participants with the outcome? 59,644 participants; 1,141 hip fractures; 2,218 osteoporotic fractures Models without BMD had 32 candidate predictors (conservatively): Hip fracture events per variable: 1,141/32= 35.6 Other fractures events per variable: 2,218/32= 69.3 Models with BMD had 45 candidate predictors (conservatively): Hip fracture events per variable: 1,141/45= 23.4 Other fractures events per variable: 2,218/45= 49.3 Depending on how interactions were tested in other candidate predictors, the model with BMD may have had too few events per variable, but probably okay.	PY	
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) Outcomes per group not reported for validation (2007 WHO Report). Representation from racial and ethnic groups other than White is likely very limited. The 2007 WHO Report provided AUCs for external validation cohorts, and 2 are from the U.S. that presumably are relevant to the U.S. equations. The Study of Osteoporotic Fractures has 99.7% White participant population. Women's Health Initiative uses classifications of White (87%), Black (7%), Hispanic (3%), American Indian/Alaska Native (<1%), Asian/Pacific Islander (2%). Model performance was not reported separately by group.	NA	NA
Were continuous and categorical predictors handled appropriately? BMI was entered as a categorical variable; others were handled as expected (categorical or continuous).	Y	NA
Were all enrolled participants included in the analysis? Unclear	NI	NA
Were participants with missing data handled appropriately? When predictor data were unavailable for a cohort, the coefficient was set to 0. Further detail in Table 5.4 of the WHO Technical Report, but for example only 3/12 cohorts provided alcohol or arthritis predictor data.	N	NA
Was selection of predictors based on univariable analysis avoided? Unclear. pg 93 of WHO Technical Report notes that at the individual cohort level, for any risk factor, covariates included current age and time since followup, with and without BMD, but this is not all of the risk factors, so there could be additional confounding.	PN	NA
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Competing hazards. We do not know about loss to followup, so censoring in the fracture hazard functions is unknown.	PY	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? Competing hazards with mortality where mortality was age, sex, and race specific.	PY	NA
Were relevant model performance measures evaluated appropriately? No calibration data; this is a major issue. AUC reported.	N	NA
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups? No calibration data. Discrimination (AUC) reported by cohort but not further by group.	N	NA
Were model overfitting and optimism in model performance accounted for? No. No mention of any resampling.	N/NI	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? Underlying equations/coefficients were not reported.	NI	NA

Appendix G Table 10. Fracture Risk Assessment Tool (FRAX) Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Risk of bias introduced by the analysis (low/high/unclear)	High Unclear if all participants were used in analysis. When predictor data for a cohort were not available (not measured), the predictor was set to 0. High degree of missing predictor data in various cohorts. Selection of predictors not entirely multivariable. Calibration not assessed. Model optimism and optimism not addressed. Underlying equations not reported. Unclear exactly how race was being used—assuming stratified and that race is mainly affecting death hazard function. The 2 things that the analysis did right were keeping continuous measures continuous and using a competing risk model.

Abbreviations: AUC=area under the curve; BIPOC=Black, Indigenous, and People of Color; BMD=bone mineral density; BMI=body mass index; Dev=development; N=no; NA=not available; NI=no information; PN=probably no; PY=probably yes; RA=rheumatoid arthritis; U.S.=United States; Val=validation; WHI=Women's Health Initiative; WHO=World Health Organization; Y=yes.

Appendix G Table 11. Fracture Risk Evaluation Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Step 1 and 2

Item	Response
Intended use of model:	Predict risk of developing a major osteoporotic fracture (primary outcome) or a hip fracture (secondary outcome)
Participants including selection criteria and setting:	Total population of Denmark age 45 years and older from 1998 to 2013
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Major osteoporotic fracture in women: 38 predictive baseline diagnoses from ICD-10 codes Major osteoporotic fractures in men: 43 predictive baseline diagnoses from ICD-10 codes Hip fractures in men or women: 28 predictive baseline diagnoses from ICD-10 codes
Outcome to be predicted:	1 year risk of MOF or hip fracture
Type of prediction study	Development and validation
Citations	https://pubmed.ncbi.nlm.nih.gov/29924428/

Abbreviations: ICD-10=International Statistical Classification of Diseases and Related Health Problems, 10th Edition; MOF=major osteoporotic fracture.

Appendix G Table 12. Fracture Risk Evaluation Model Development Cohort Assessment from Prediction From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response		
Describe the sources of data and criteria for participant selection:	The study used the Danish Civil Registration System to identify persons living in Denmark age 45 years or older on January 1, 2013, and extracted ICD-10 codes from the National Patient Register for those persons between 1998 and 2013		
Risk of Bias		**Dev**	**Val**
Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data?		Y	Y
Were all inclusions and exclusions of participants appropriate? No exclusions reported		Y	Y
Was there sufficient representation of individuals from racial and ethnic groups in model development data? Race not reported		NA	NA
Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only) The model does not include race		NA	NA
Risk of bias introduced by selection of participants (low/high/unclear)	Low		
Applicability			
Describe included participants, setting, and dates:	Development cohorts were likely predominantly White (race NR); applicability unclear to other races		
Concern that the included participants and setting do not match the review question (low/high/unclear)	Unclear		

Abbreviations: ICD-10=International Statistical Classification of Diseases and Related Health Problems, 10th Edition; NA=not applicable; Y=yes.

Appendix G Table 13. Fracture Risk Evaluation Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response		
List and describe predictors included in the final model, e.g., definition and timing of assessment:	Predictors included age and health conditions listed in ICD-10 codes at level 2; ICD-10 codes coded for administrative information were excluded. This yielded 1,564 codes among women and 1,467 among men. If those codes with prevalence less than 0.1% were excluded (958 in women and 931 in men). Codes associated with the condition with a p-value ≥0.01 were retained in the model, leading to 38 predictive baseline diagnoses in women and 43 in men for major osteoporotic fracture and 32 codes for both women and men for hip fractures. Codes were categorized using the Charlson comorbidity index. The prediction was for 1-year risk		
Risk of Bias		**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants?		Y	Y
Were predictor assessments made without knowledge of outcome data?		PY	PY
Are all predictors available at the time the model is intended to be used?		Y	Y
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? Race is not included in the model		NA	NA
Was differential missingness of predictor data in racial and ethnic groups considered? Race is not included in the model		NA	NA
Risk of bias introduced by predictors or their assessment (low/high/unclear)	Low Standardized diagnosis codes		
Applicability			
Concern that the definition, assessment, or timing of predictors in the model do not match the review question (low/high/unclear)	Low These predictors are relevant to what would be collected in primary care and included in electronic medical records.		

Abbreviations: ICD-10=International Statistical Classification of Diseases and Related Health Problems, 10th Edition; NA=not applicable.

Appendix G Table 14. Fracture Risk Evaluation Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response		
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:	MOF defined as at least one code for a hip, clinical vertebral, wrist, or humerus fracture (ICD-10 codes: S120, S121, S122, S220, S221, S320, T08, S422, S423, S720, S721, S722, S525, S526) during 2013 Hip fracture defined as having at least one code (S720, S721, S722) during 2013		
Risk of Bias		**Dev**	**Val**
Was the outcome determined appropriately? The timing of diagnoses vs. fractures may not be clear in a small subset of persons who may have experienced fractures in late 2012 and diagnosed in 2013 but likely a very small number compared to the whole population. Some MOF fractures may be open to coding errors, particularly for humerus and wrist. Hip fracture data may be more accurate given the devastating nature of this outcome		PN for MOF, PY for hip	PN for MOF, PY for hip
Was a prespecified or standard outcome definition used? As noted above, MOF fractures may be open to coding errors, particularly for humerus and wrist. Hip fracture data may be more accurate given the devastating nature of this outcome		PN for MOF, PY for hip	PN for MOF, PY for hip
Were predictors excluded from the outcome definition? Yes		Y	Y
Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered? Followup not reported, either overall or by race but because the data come from national registries, censoring from attrition is not a concern		NA	NA
Was the outcome determined without knowledge of predictor information?		PY	PY
Was the time interval between predictor assessment and outcome determination appropriate? 1-year horizon is short.		PN	PN
Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)? The study did not create separate risk categories by race.		NA	NA
Risk of bias introduced by the outcome or its determination (low/high/unclear)	Unclear for MOF, low for hip Outcomes for MOF ascertained from routine care database, so these were not adjudicated or standardized so they are open to coding errors		
Applicability			
At what time point was the outcome determined:	The outcome was determined over 1 year		
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	Unclear for the composite outcome of MOF		
Concern that the outcome, its definition, timing or determination do not match the review question (low/high/unclear)	Low The outcome of a fracture as indicated on a medical record seems broadly applicable.		

Abbreviations: ICD-10=International Statistical Classification of Diseases and Related Health Problems, 10th Edition; MOF=major osteoporotic fracture; N=no; PN=probably no; PY=probably yes; Y=yes.

Appendix G Table 15. Fracture Risk Evaluation Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response	
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	Women Development cohort N=647,103, MOF fractures: 8,890 (1.37%), hip fractures: 3,008 (0.46%) Validation cohort N=647,103, MOF fractures: 8,804 (1.36%), hip fractures: 3,047 (0.47%) Men Development cohort N=600,567, MOF fractures: 3,345 (0.56%), hip fractures: 1,417 (0.24%) Validation cohort N=600,566, MOF fractures: 3,371 (0.56%), hip fractures: 1,405 (0.23%) Predictor variables were ICD-10 codes categorized by the Charlson comorbidity index: 38 for women for MOF, 43 for men for MOF, 32 for men and women for hip All candidate predictors had prevalence of 0.1% of higher. For women and MOF predictors, the highest prevalence was 8% for fracture of the forearm and dislocation/strains/sprains. For men and MOF predictors, the highest prevalence was for open head wound (10%). For women and hip predictors, the highest prevalence was 8% for fracture of the forearm. For men and hip predictors, the highest prevalence was for open head wound (10%). Detailed prevalence for predictor variables is reported in the development cohort study[477]	
Describe how the model was developed (for example in regards to modelling technique (e.g., survival or logistic modelling), predictor selection, and risk group definition):	All ICD-10 codes other than administrative ones were included, codes with prevalence lower than 0.1% were dropped, and codes with p values of 0.01 or more in logistic regression were retained in iterative analyses (all codes were included and then dropped based on p value).	
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	The population was stratified by gender and then randomly split into development and validation cohorts	
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	1. Produced ROC curve comparing the full model with the model only including age only and calculated area under the curve (AUC); calculated AUC separately for each age group 2. Compared in a step diagram predicted with observed fracture risk with 95% CIs in 1% risk intervals (with one large interval for risk >5%) for MOFs and 0.3% risk intervals for hip fractures (with one large interval for risk >1.5%). The authors categorized risks above 5% for MOFs and 1.5% for hip fractures together because they judged that these risks are high enough to clearly indicate further diagnosis, and hence not require more detailed distinction. 3. Evaluated the calibration of the logistic regression by calculating the Brier score. Random split sample gives some information on overfitting	
Describe any participants who were excluded from the analysis:	None	
Describe missing data on predictors and outcomes as well as methods used for missing data:	None	
Risk of Bias	**Dev**	**Val**
Were there a reasonable number of participants with the outcome?	Y	Y

Appendix G Table 15. Fracture Risk Evaluation Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Yes for women, probably yes for men for hip fractures			
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) NA, model performance is not reported separately by race		NA	NA
Were continuous and categorical predictors handled appropriately? All were categorical		Y	Y
Were all enrolled participants included in the analysis? Yes		Y	Y
Were participants with missing data handled appropriately? No missing participants		NA	NA
Was selection of predictors based on univariable analysis avoided? Yes, all predictors entered into logistic regression and eliminated based on backwards selection		Y	NA
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Not a competing risk model. Competing risk of death from any cause could be important. The study did specify a short prediction interval (1 year) but did not report how many people died and were censored on that basis. No other censoring or sampling (full population)		PY	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? NA, race was not reported		NA	NA
Were relevant model performance measures evaluated appropriately? Calibration and discrimination reported		Y	NA
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups?		NA	NA
Were model overfitting and optimism in model performance accounted for? Random split sample validation gives us some information on optimism. Performance was better in the validation cohort		Y	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? Not specified but coefficients are reported in the supplement, other studies note using the coefficients without modification[162]		PY	NA
Risk of bias introduced by the analysis (low/high/unclear)	Unclear; censoring/competing risks based on death not reported		

Abbreviations: AUC=area under the curve; ICD-10=International Statistical Classification of Diseases and Related Health Problems, 10th Edition; MOF=major osteoporotic fracture; N=no; NA=not available; PN=probably no; PY=probably yes; OC=receiver-operating characteristic; U Y=yes.

Appendix G Table 16. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Step 1 and 2 Specification of Review Question

Item	Response
Intended use of model:	Predict risk of developing an osteoporotic fracture
Participants including selection criteria and setting:	Patients age 40 years or older
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Demographic information and clinical and family history with or without BMD
Outcome to be predicted:	Risk of osteoporotic fracture (with a minimum of 3 years' observation for studies with no specified prediction interval or a median or mean of 80% of the time in studies with a specified prediction interval)
Type of prediction study	Development only
Citations	Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of a nomogram for individualizing hip fracture risk in men and women. Osteoporos Int. 2007 Aug;18(8):1109-17. doi: 10.1007/s00198-007-0362-8. Epub 2007 Mar 17. PMID: 17370100. Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. Osteoporos Int. 2008 Oct;19(10):1431-44. doi: 10.1007/s00198-008-0588-0. Epub 2008 Mar 7. PMID: 18324342. Both articles are writeups of the development of the nomogram. The earlier publication (2007) presented results for a nomogram that included age, BMD, prior fractures, and prior falls for hip fractures. The 2008 publication, although published later, appears to address an earlier stage in the development of the nomogram, where the authors were trying to compare nomogram results with BMD vs. body weight and predict any osteoporotic fracture (barring morphometric and some others).

Abbreviations: BMD=bone mineral density.

Appendix G Table 17. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response		
Describe the sources of data and criteria for participant selection:	The authors used risk and fracture data from the Dubbo Osteoporosis Epidemiology Study (DOES), a cohort study designed specifically to study osteoporosis. All men and women age 60 years or older living in Dubbo, a city (400 km north west of Sydney, Australia), were invited to participate in an epidemiological study. Risk factors and BMD came from these individuals and T-scores were calculated for the BMD, based on "young normal" BMD was obtained from a sample of 52 Australian men and women ages 20 to 32 years. These values were identical to those in the LUNAR Caucasian database (the study collected BMD using LUNAR machines).		
Risk of Bias		**Dev**	**Val**
Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data? Data source explicitly designed for osteoporosis epidemiology.		PY	NA
Were all inclusions and exclusions of participants appropriate? No exclusions; all residents invited to participate.		NA	NA
Was there sufficient representation of individuals from racial and ethnic groups in model development data? 98.6% were Caucasian and 1.4% indigenous Aboriginal.		N	NA
Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only) The model validation cohorts were predominantly or exclusively White but the model does not include race.		NA	NA
Risk of bias introduced by selection of participants (low/high/unclear)	Low Appropriate data sources, no exclusions		
Applicability			
Describe included participants, setting and dates:	Development cohort was almost entirely White, applicability unclear to other races.		
Concern that the included participants and setting do not match the review question (low/high/unclear)	Unclear Development and validation cohorts broadly representative of White women but not other women.		

Abbreviations: BMD=bone mineral density; Dev=development; DOES=Dubbo Osteoporosis Epidemiology Study; N=no; NA=not available; PY=probably yes; RCT=randomized, controlled trial; Val=validation.

Appendix G Table 18. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response		
List and describe predictors included in the final model, e.g., definition and timing of assessment:	In the 2007 (hip fracture) article, the risk factors included Age BMD Prior fracture Prior fall Quadricep strength was included in an initial model but found to only add 1.5% to the predictive power and was dropped Information on age, anthropomorphic data, and lifestyle factors were collected at baseline by interview by a nurse using a structured questionnaire (https://asbmr.onlinelibrary.wiley.com/doi/full/10.1359/JBMR.050520). BMD (g/cm^2) was measured at the lumbar spine and femoral neck by DXA using a LUNAR DPX-L densitometer. The 2008 article (any fracture (any first osteoporotic fracture)) tested weight instead of BMD and found that having BMD made the nomogram more accurate than having weight (AUC of 0.75 instead of 0.72 for women and 0.74 for men. 5-year or 10-year followup were the time points for prediction.		
Risk of Bias		**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants? Routine care database so consistency of data collection for items such as family history-taking is unknown. Unclear how missing data were handled.		PY	
Were predictor assessments made without knowledge of outcome data?		PY	
Are all predictors available at the time the model is intended to be used?		PY	
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? Race and ethnicity were not used in the model		Y	
Was differential missingness of predictor data in racial and ethnic groups considered? Not applicable, race was not used		NA	
Risk of bias introduced by predictors or their assessment (low/high/unclear)	Low Appears that predictors were collected in the same way.		
Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question (low/high/unclear)	Low These predictors are relevant to what would be collected in primary care.		

Abbreviations: AUC=area under the curve; BMD=bone mineral density; Dev=development; DXA=dual-energy x-ray absorptiometry; NA=not available; PY=probably yes; Val=validation; Y=yes.

Appendix G Table 19. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response		
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:	Fractures occurring during the study period were identified for residents of the Dubbo local government area through radiologists' reports from the two centers providing X-ray services. Fractures due to major trauma, underlying disease, those of digits, skull, or cervical spine, or morphometric vertebral fractures were not included. Hip fractures were the focus on the 2007 article, and first osteoporotic low trauma and nonpathological fractures were considered the primary outcome of the 2008 study. In the 2008 article, 92% of those who had fractures in the DOES agreed to have BMD. Fractures were only included if the report of fracture was definite and, on interview, had occurred with minimal trauma (fall from standing height or less). Fractures more than 3 months before study entry were not considered in the analysis. Fractures were classified as hip, vertebrae (symptomatic), wrist, metacarpal, humerus, scapula, clavicle, distal femur, proximal tibia, patella, pelvis, and sternum.		
Risk of Bias		**Dev**	**Val**
Was the outcome determined appropriately? The timing of eligible fractures is unclear, it appears that some fractures may have taken place up to 3 months before study entry and measurement of BMD, from the description of eligible fractures. Fractures came from radiologists' reports		PN	
Was a prespecified or standard outcome definition used?		PY for clinical fractures, NA for morphometric and other fractures (digits, skull, cervical spine)	
Were predictors excluded from the outcome definition?		Y	
Was the outcome defined and determined in a similar way for all participants?		PY	
3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered? Followup not reported, either overall or by race. Could possibly have differential censoring in different populations, but the vast majority of the population was White.		N	
Was the outcome determined without knowledge of predictor information?		Y	
Was the time interval between predictor assessment and outcome determination appropriate? Median duration of followup was 13 years		Y	
Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)?		Y	
Risk of bias introduced by the outcome or its determination (low/high/unclear)	Low In the 2007 study, 8% of sample with fractures did not have BMD measurements, but the Ns are not clearly described. From a total of 1,581 men and 2,095 women age ≥60 years, data were analyzed from 1,358 women and 858 men who had been followed up between 1989 and 2004, which means only 60% of the overall sample was retained.		
Applicability			
At what time point was the outcome determined:	Median 13-year followup		
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	Not applicable.		
Applicability (continued)			
Concern that the outcome, its definition, timing or determination do not match the review question (low/high/unclear)	Unclear The outcome of a fracture by radiologist report seems broadly applicable, but the exclusion of morphometric and other fractures may reduce the applicability of the instrument to all fractures		

Abbreviations: BMD=bone mineral density; Dev=development; DOES=Dubbo Osteoporosis Epidemiology Study; N=no; NA=not available; PN=probably no; PY=probably yes; Val=validation; Y=yes.

Appendix G Table 20. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	1,581 men and 2,095 women age ≥60 years in Dubbo. Of these, 1,028 women and 740 men (48%) were included in the 2007 analysis and 1,358 women (of these 96 women and 31 men sustained at least one hip fracture) and 858 men (60%) in the 2008 analysis (of these 426 women and 149 men sustained at least one fracture). For hip fracture, N events per candidate predictor is not reported, but HRs for women are: for 5+ years of age, HR: 1.95 (95% CI, 1.70 to 2.22); each prior fracture HR: 2.89 (95% CI, 1.85 to 4.50); falls in the past 12 months, each fall HR: 1.42 (95% CI, 0.93 to 2.16); FNBMD for -0.12 g/cm^2 HR: 2.62 (95% CI, 2.21 to 3.11). For men, for 5+ years of age, HR: 2.31 (95% CI, 1.76 to 3.03); each prior fracture HR: 4.23 (95% CI, 2.96 to 6.04); falls in the past 12 months, each fall HR: 1.40 (95% CI, 1.20 to 1.62); FNBMD for -0.12 g/cm^2 HR: 2.61 (95% CI, 1.95 to 3.50). For any fracture, for women, events per candidate predictor were not reported, but HRs for women were: for 5+ years of age, HR: 1.43 (95% CI, 1.34 to 1.53); each prior fracture HR: 2.06 (95% CI, 1.87 to 2.26); falls in the past 12 months, each fall HR: 1.23 (95% CI, 1.10 to 1.38); FNBMD for -0.12 g/cm^2 HR: 1.68 (95% CI, 1.55 to 1.82). For men, for 5+ years of age, HR: 1.67 (95% CI, 1.48 to 1.88); each prior fracture HR: 2.92 (95% CI, 2.43 to 3.52); falls in the past 12 months, each fall HR: 1.38 (95% CI, 1.13 to 1.69); FNBMD for -0.12 g/cm^2 HR: 1.62 (95% CI, 1.43 to 1.83).
Describe how the model was developed (for example, in regards to modeling technique [e.g., survival or logistic modeling], predictor selection, and risk group definition):	For both hip and any fractures, Bayesian model average helped identify the most parsimonious models. 1,000 subsamples, each with 150 subjects, of the entire sample were repeatedly resampled (with replacement) and analyzed. For hip fracture, the most parsimonious one included age, femoral neck BMD, prior fracture, previous fall, and quadriceps strength, but the last one was dropped because it only added 1.5% to the predictive power.
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross-validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	The model was not validated externally. The nomograms were internally validated by the bootstrap method.
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	For hip fracture, the study report maximum calibration error (2% for women, 7% for men), and c-index of 0.85 (unclear if overall or for men or women). No visual depiction of calibration. For any fracture, max calibration error (less than 1%) and a graph of predicted to observed probabilities were shown. Although the c-index is mentioned in methods, it is not reported in the results.
Describe any participants who were excluded from the analysis:	NR
Describe missing data on predictors and outcomes as well as methods used for missing data:	NR, see above, 48% to 60% retained and missing data not described clearly.

Risk of Bias	Dev	Val
Were there a reasonable number of participants with the outcome? PN for hip fractures, unclear for any fractures	PN	NA
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) NA, model performance is not reported separately by race	NA	NA
Were continuous and categorical predictors handled appropriately? BMI was entered as a categorical variable, others were handled as expected (categorical or continuous)	PY	NA
Risk of Bias (continued)	**Dev**	**Val**
Were all enrolled participants included in the analysis?	N	
Were participants with missing data handled appropriately?	NI	NA
Was selection of predictors based on univariable analysis avoided?	PY	NA

Appendix G Table 20. Garvan Fracture Risk Calculator Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response	
Regression coefficients (log of HR) from final models were used as weights (Tables 3 and 4). The relative risks were unadjusted for other variables.		
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Not a competing risk model. Unclear how other complexities were addressed.	NI	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? No race/ethnicity differences reported, not a competing risk model	NA	NA
Were relevant model performance measures evaluated appropriately? Both calibration and validation measures not reported for both models	PN	NA
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups? Model performance measures not reported separately in different racial and ethnic groups, but overwhelming majority was White.	N	NA
Were model overfitting and optimism in model performance accounted for? Yes	Y	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? Difficult to map weights from final model to multivariate analysis for the hip fracture model, not possible in the any fracture model because individual coefficients not reported	PY	NA
Risk of bias introduced by the analysis (low/high/unclear)	High Large missingness, no explanation of effects, some but not all performance measures	

Abbreviations: BMD=bone mineral density; BMI=body mass index; CI=confidence interval; Dev=development; FNBMD=femoral neck bone mineral density; HR=hazard ratio; N=no; NA= not available; NI= no information; NR=not reported; PN=probably no; PY=probably yes; Val=validation; Y=yes.

Appendix G Table 21. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Steps 1 and 2 Specification of Review Question

Item	Response
Intended use of model:	Predict risk of developing an osteoporotic fracture
Participants including selection criteria and setting:	Patients age 40 years or older
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Demographic information and clinical and family history with or without BMD
Outcome to be predicted:	Risk of osteoporotic fracture (with a minimum of 3 years observation for studies with no specified prediction interval or a median or mean of 80% of the time in studies with a specified prediction interval)
Type of prediction study	Development and validation
Citations	Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ. 2012 May 22;344:e3427. doi: 10.1136/bmj.e3427. PMID: 22619194.

Abbreviations: BMD=bone mineral density.

Appendix G Table 22. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response		
Describe the sources of data and criteria for participant selection:	QResearch database. U.K. nationally representative primary care electronic database. All QResearch practices that have been using EMIS computer system for 1 year were included. Random split sample validation of two thirds of practices in derivation set and one third in validation set. Patients ages 30 to 100 years at study entry date and registered with eligible practices at some time between 1 January 1993 and 1 October 2011. Patients needed to have 1 year of complete data in the medical record. Patients with a previous recorded fracture were eligible for inclusion in the cohort.		
Risk of Bias		**Dev**	**Val**
Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data? Routine care database. Authors described this as a "prospective open cohort study," but these data were not being collected with a defined research protocol.		N	NA
Were all inclusions and exclusions of participants appropriate? No exclusions based on missing values. Imputation for missing values for alcohol, smoking, and BMI. Includes those with prior fracture (approximately 2% of population).		Y	NA
Was there sufficient representation of individuals from racial and ethnic groups in model development data? Based on Table 1, it appears that self-assigned ethnic origin was only recorded in ~45% of derivation and validation cohorts. and those with ethnic origin not reported are combined with White participants in later analysis. Missing information about race or ethnicity (or anything else) was associated with fewer contacts with the healthcare system. This study required only a minimum of 1 year of data in the medical record. If White participants have higher fracture risk (as suggested in Table 3), this may result in overstated risk in those with nonrecorded ethnicity, unless they are also White. However, the reference category includes White and nonrecorded ethnicity, so the true estimate of increased fracture risk in Whites in this dataset is unknown. 2011 Census data show that 87% of the United Kingdom is White, yet ~95% are categorized as White or not recorded in Table 1. The percentage of participants from other ethnic groups is small, but given the extremely large size of the database, there are probably sufficient absolute numbers.		PY	NA
Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only) White or not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Caribbean, Black African, Chinese, Other. Two thirds of QResearch practices assigned to derivation dataset and one third to validation dataset so underlying population classifications are the same.		NA	Y
Risk of bias introduced by selection of participants (low/high/unclear)	High Routine care database is not ideal as data are not collected with standardized research protocol. Further, self-assigned race and ethnicity data are not available from ~55% of population.		
Applicability			
Describe included participants, setting and dates:	Primary care practices in the United Kingdom. The website refers to this version of the tool as QFracture-2016, so another update occurred after this main publication and March 2019 is listed as the last update date on the website. It is likely they are not adding new variables, but coefficients are probably getting updated—somewhat unclear.		
Concern that the included participants and setting do not match the review question (low/high/unclear)	Low Recent and nationally representative population of primary care patients in the United Kingdom. This Q product does not include post-code specific deprivation score, making it more transportable to the United States.		

Abbreviations: BMI=body mass index; Dev=development; N=no; PY=probably yes; RCT=randomized, controlled trial; U.K.=United Kingdom; Val=validation; Y=yes.

Appendix G Table 23. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response
List and describe predictors included in the final model, e.g., definition and timing of assessment:	This is the list of candidate risk factors; followed by notes of whether it was retained or combined in final model. 2009 risk factors: (24) • Age at study entry (in single years) • Body mass index (continuous) • Smoking status (non-smoker, ex-smoker, light smoker (<10 cigarettes/day), moderate smoker (10-19 cigarettes/day), heavy smoker (≥20 cigarettes/day) • Parental history of osteoporosis or hip fracture in a first-degree relative (binary variable; yes/no) • Cardiovascular disease (binary variable; yes/no) • Alcohol intake (none, trivial (<1 unit/day), light (1-2 units/day), medium (3-6 units/day), heavy (7-9 units/day), very heavy (>9 units/day) • Rheumatoid arthritis (binary variable; yes/no)—combined with SLE in final model • Type 2 diabetes (binary variable; yes/no) • Asthma (binary variable; yes/no)—combined with COPD in final model • History of falls (binary variable; yes/no) • Chronic liver disease (binary variable; yes/no) • Gastrointestinal conditions likely to result in malabsorption (that is, Crohn's disease, ulcerative colitis, celiac disease, steatorrhoea, blind loop syndrome at baseline) (binary variable; yes/no) • Other endocrine conditions (thyrotoxicosis, primary or secondary hyperparathyroidism, Cushing's syndrome) at baseline (binary variable; yes/no) • At least two prescriptions for systemic corticosteroids in the six months preceding baseline (binary variable; yes/no) • At least two prescriptions for tricyclic antidepressants in the six months preceding baseline (binary variable; yes/no) • At least two prescriptions for hormone replacement therapy (in women) in the six months preceding baseline (binary variable; yes/no)—estrogen-only HRT in final model • Menopausal symptoms in women (binary variable; yes/no)—considered but not included in final model New risk factors examined: (20) • Self-assigned ethnic origin (White or not recorded, Indian, Pakistani, Bangladeshi, other Asian, Black African, Black Caribbean, Chinese, other including multiethnic) • Previous fracture (hip, vertebral, proximal humerus, or distal radius fracture) (binary variable; yes/no) • Use of other antidepressants apart from tricyclic antidepressants (at least two prescriptions in previous six months) (binary variable; yes/no)—combined with tricyclic antidepressants in final model • Chronic obstructive pulmonary disease (binary variable; yes/no)—combined with asthma in final model • Epilepsy (binary variable; yes/no)—combined with anticonvulsants in final model • At least two prescriptions of anticonvulsants in the 6 months preceding baseline (binary variable; yes/no) • Dementia (binary variable; yes/no) • Parkinson's disease (binary variable; yes/no) • Any cancer (binary variable; yes/no) • Systemic lupus erythematosus (binary variable; yes/no)—combined with RA in final model • Chronic renal disease (binary variable; yes/no) • Type 1 diabetes (binary variable; yes/no) • Care or nursing home residence (binary variable; yes/no)—included in equations for men only

Appendix G Table 23. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response		
List and describe predictors included in the final model, e.g., definition and timing of assessment (continued)	All values of these variables were restricted to those recorded in the person's electronic healthcare record before baseline, except for body mass index, alcohol intake, and smoking status. Values recorded closest to study entry date and recorded before the diagnosis of osteoporotic fracture were used (or for patients who did not develop a fracture, before censoring). Assumed that if there was no recorded value of a diagnosis, prescription, or family history, then the patient did not have that exposure. 10-year followup was the primary time point for prediction, but risk equations were also derived for each year from 1 to 15 years so users could select time period for evaluation.		
Risk of Bias		**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants? Routine care database, so consistency of data collection for items such as family history-taking is unknown. If there was no recorded value of a diagnosis, prescription, or family history, then the patient was categorized as not having that exposure.		PN	NA
Were predictor assessments made without knowledge of outcome data?		Y	NA
Are all predictors available at the time the model is intended to be used?		Y	NA
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? It appears that ethnic origin was included for calibration. Table 3 reports adjusted HRs for fractures by ethnic origin where all groups had a statistically significant lower incidence compared with the reference category of White or not recorded. No other rationale for inclusion of race or ethnicity is provided. It is interesting that while race and ethnicity were probably included for calibration, calibration was not reported separately by group.		PY	NA
Was differential missingness of predictor data in racial and ethnic groups considered? Missingness of data is not reported by race and ethnicity. However, multiple imputation was used for missing values for alcohol, smoking, and BMI.		PY	NA
Risk of bias introduced by predictors or their assessment (low/high/unclear)	High Because this is a routine care database, we cannot guarantee that predictors were defined as assessed in the same way for all participants. The handling of missing data through multiple imputation is a strength of the approach.		
Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question (low/high/unclear)	Low These predictors are relevant to what would be collected in primary care and included in electronic medical records.		

Abbreviations: BMI=body mass index; COPD=chronic obstructive pulmonary disease; Dev=development; HR=hazard ratio; HRT= hormone replacement therapy; PN=probably no; PY=probably yes; RA=rheumatoid arthritis; SLE=systemic lupus erythematosus; Val=validation; Y=yes.

Appendix G Table 24. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

<table>
<tr><th>Item</th><th colspan="3">Response</th></tr>
<tr><td>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</td><td colspan="3">Two primary outcomes: osteoporotic fracture defined as a diagnosis of a hip, vertebral, proximal humerus, or distal radius fracture during followup and diagnosis of hip fracture, where these fractures were recorded either on the general practice record or the linked death record.</td></tr>
<tr><th colspan="2">Risk of Bias</th><th>Dev</th><th>Val</th></tr>
<tr><td colspan="2">Was the outcome determined appropriately?
From a routine care database. Outcome determination methods not specified by protocol so open to coding errors. However, the recording of hip fracture is probably fine given how devastating this outcome is it is likely to have been recorded. Mortality and hospital data were true record linkages but otherwise, it is incumbent on the GP to record other incidents in the medical record.</td><td>PY for hip; N for osteo</td><td>NA</td></tr>
<tr><td colspan="2">Was a prespecified or standard outcome definition used?
From routine care database, so these are not adjudicated or standardized so they will be open to coding errors. Again, hip fracture is probably okay given how devastating this outcome is.</td><td>PY for hip; N for osteo</td><td>NA</td></tr>
<tr><td colspan="2">Were predictors excluded from the outcome definition?
However, equations are predicting either first or subsequent fracture so event index bias is likely present. This is a problem because the coefficients for recurrent events are likely different than first events. After a first event, individuals may modify risk factors (use of steroids, alcohol, etc.) so the assumption that the weighting of the coefficients will be the same in first event/recurrent event individuals is flawed. This also ignores the natural history of fractures (small bones break first before hip or vertebral involvement).</td><td>Y</td><td>NA</td></tr>
<tr><td colspan="2">Was the outcome defined and determined in a similar way for all participants?
Predictors and outcome taken from same routine care database for everyone; however, we do not know the consistency of imaging or other diagnostics used across the population.</td><td>NI</td><td>NA</td></tr>
<tr><td colspan="2">3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered?
Followup not reported, either overall or by group. Considering the context of national healthcare system we assumed followup overall is good, but we do not know % dying, which could generate differential censoring in different populations.</td><td>NI</td><td>NA</td></tr>
<tr><td colspan="2">Was the outcome determined without knowledge of predictor information?</td><td>Y</td><td>NA</td></tr>
<tr><td colspan="2">Was the time interval between predictor assessment and outcome determination appropriate?
10-year horizon is reasonable. Developers also created different models for different horizons of followup.</td><td>Y</td><td>NA</td></tr>
<tr><td colspan="2">Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)?
It is unclear whether incidental findings were included. Incidental findings were likely higher in populations with more access to imaging. For other fractures, diagnosis codes were from a wide range of settings (inpatient, outpatient, office or clinic consultations, ER, nursing home care, autopsy exam or death certificates). Fractures identified incidentally were likely to be differentially ascertained in populations with greater access to imaging. The model revision (3.0) focused on symptomatic vertebral fractures, which is a change from 2.0, which included incidental findings.</td><td>NI</td><td>NA</td></tr>
<tr><td>Risk of bias introduced by the outcome or its determination (low/high/unclear)</td><td colspan="3">High
Outcomes ascertained from routine care database, so these are not adjudicated or standardized so they will be open to coding errors. Unknown followup or percentage of individuals dying/getting censored.</td></tr>
<tr><th colspan="4">Applicability</th></tr>
<tr><td>At what time point was the outcome determined:</td><td colspan="3">10 years is the primary time horizon, but equations were also developed for other horizons and these are available on the website.</td></tr>
<tr><td>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</td><td colspan="3">NR</td></tr>
<tr><th colspan="4">Applicability (continued)</th></tr>
<tr><td>Concern that the outcome, its definition, timing or determination do not match the review question. (low/high/unclear)</td><td colspan="3">Low
The outcome of a fracture as indicated on a medical record seems broadly applicable.</td></tr>
</table>

Appendix G Table 24. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Abbreviations: Dev=development; ER=emergency room; GP=general practitioner; N=no; NI=no information ; NR=not reported; PY=probably yes Val=validation; Y=yes.

Appendix G Table 25. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	4 equations: men and women separately and osteoporotic and hip fracture separately Derivation: 3,142,673 (59,628 with prior fracture); 59,772 new osteo fractures; 20,028 hip fractures Validation: 1,583,373 (27,907 with prior fracture); 28,685 new osteo fractures; 9,610 hip fractures Candidate predictors: 44 from predictors section plus 3 fractional polynomial terms for each model (below)=47 Above RFs plus fractional polynomial terms for age and BMI • Osteo fracture: (age/10)2; (age/10)3; BMI/10)-1 • Hip fracture: (age/10)2; (age/10)3; (BMI/10)-2 Events per candidate predictor: 59,772 new osteo fractures/47=1,272 20,028 hip fractures/47=426
Describe how the model was developed (for example, in regards to modeling technique [e.g., survival or logistic modeling], predictor selection, and risk group definition):	Cox proportional hazards model: Separate for osteoporotic fracture and hip fracture and separate for men and women. Robust variance estimates used to allow for clustering of patients with general practices. Graphical methods used to check assumption of proportional hazards. Fractional polynomials used to model nonlinear risk associations with continuous variables (age and BMI). Predictors from the previous QFracture were carried forward for evaluation (except for Townsend deprivation score, which is not further explained), and risk factors recommended in 2012 NICE report were evaluated for inclusion. They retained the predictor if it was significant (threshold not reported that we can see). For example, care home residency was retained in the equation for men only because it was only statistically significant in this population. Clinically similar variables were tested to determine if they could be appropriately combined. They ran a model with separate terms for each variable; if two similar variables were both significant (hazard ratio <0.8 or >1.20, and p<0.01), they were compared with a direct significance test. If this comparison was not significant (at P <0.01) and if the hazard ratios were within 0.2 of each other, the variables were combined into a new variable (for example, either rheumatoid arthritis or systemic lupus erythematosus).
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross-validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	Random split sample validation (420 in derivation, 207 in validation). This can tell us about reproducibility of coefficients but does not tell us anything about transportability—for that it would have been preferred to split the sample by geography or time.
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	Calibration plot: predicted vs. observed risk at 10 years for every tenth of predicted risk (Figure 2) Overall performance measure: R2 Discrimination: D statistic, AUC Reclassification in patients reclassified from high to low risk (low to high) compared to 2009 algorithm Sensitivity for the top 10% of risk predicting a new fracture Internal split sample validation gives us some information about optimism.
Describe any participants who were excluded from the analysis:	None
Describe missing data on predictors and outcomes as well as methods used for missing data:	Loss to followup unknown—given that this is part of the national health system, it is of less concern. We do not know, however, how many people died, which would give us more information about percentage censored. ~25% with missing BMI data, ~12% with missing smoking data, ~55% with missing self-assigned race or ethnicity, ~28% with no alcohol data

Appendix G Table 25. QFracture Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response	
Risk of Bias	**Dev**	**Val**
Were there a reasonable number of participants with the outcome? 59,772 new osteo fractures/47=1,272 20,028 hip fractures/47=426	Y	NA
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) At least 100 participants with the outcome is recommended. The numbers do not appear adequate for all categories except White or not recorded or Indian (these include Pakistani, Bangladeshi, Other Asian, Caribbean, Black African, Chinese, and other ethnic group).	PN	NA
Were continuous and categorical predictors handled appropriately? Age and BMI were entered continuously. Smoking status and alcohol were categorized. (Giving partial credit)	PY	NA
Were all enrolled participants included in the analysis?	Y	NA
Were participants with missing data handled appropriately? Multiple imputation	Y	NA
Was selection of predictors based on univariable analysis avoided? Regression coefficients (log of HR) from final models were used as weights (Tables 3 and 4).	Y	NA
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Not a competing risk model. Given that those with advanced age are in the relevant population, competing risk of death from any cause could be important. However, there is the option to specify a shorter prediction interval, which would reduce this issue. We do not know how many people died and were censored on that basis.	N	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? As above	N	NA
Were relevant model performance measures evaluated appropriately?	Y	NA
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups? Model performance measures not reported separately in different racial and ethnic groups.	N	NA
Were model overfitting and optimism in model performance accounted for? Random split sample validation gives us some information on optimism. More ideal would have been a geographic or temporal split. More efficient would have been bootstrapping or other form of resampling.	Y	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Y	NA
Risk of bias introduced by the analysis (low/high/unclear)	High Primary issue is that a competing risk model was not used, which could be more appropriate because of risk of death in older age groups. We do not know the extent of the issue because we do not know the % who died during followup. Model overfitting and optimism were not accounted for; however, this dataset is huge with high events per variable so not downgrading on that. Performance in different racial and ethnic groups cannot be evaluated because performance was not reported by group, and the number of outcome events was limited in many groups. Both calibration and discrimination were reported, which is a strength.	

Abbreviations: AUC=area under the curve; BMI=body mass index; Dev=development; HR=hazard ratio; N=no; NICE=National Institute for Health and Care Excellence; PN=probably no; PY=probably yes; RF=risk factor; Val=validation; Y=yes.

Appendix G Table 26. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Step 1 and 2 Specification of Review Question

Item	Response
Intended use of model:	Predict risk of developing an osteoporotic fracture
Participants including selection criteria and setting:	Patients age 40 years or older
Predictors (used in prediction modeling), including types of predictors (e.g., history, clinical examination, biochemical markers, imaging tests), time of measurement, specific measurement issues (e.g., any requirements/prohibitions for specialized equipment):	Demographic information and clinical and family history with or without BMD
Outcome to be predicted:	Risk of osteoporotic fracture (with a minimum of 3 years observation for studies with no specified prediction interval or a median or mean of 80% of the time in studies with a specified prediction interval)
Type of prediction study	Development and validation
Citations	Robbins J, Aragaki AK, Kooperberg C, Watts N, Wactawski-Wende J, Jackson RD, LeBoff MS, Lewis CE, Chen Z, Stefanick ML, Cauley J. Factors associated with 5-year risk of hip fracture in postmenopausal women. JAMA. 2007 Nov 28;298(20):2389-98. doi: 10.1001/jama.298.20.2389. PMID: 18042916.

Abbreviations: BMD=bone mineral density.

Appendix G Table 27. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 1 Participants

Item	Response		
Describe the sources of data and criteria for participant selection:	The authors developed the algorithm using the population of postmenopausal women ages 50 to 79 years from 40 clinical centers participating in the observational study component of the Women's Health Initiative (93,676), then validated it using the sample of women enrolled in clinical trials (68,132). Participants in the clinical trial tended to be younger (mean, 62.7 years), taller (161.1 cm [63.42 in]), heavier (76.1 kg [169.1 lb]), less likely to be White (81.5% were White), with a lower proportion of the clinical trial reporting fair to poor health (8.3%), history of fracture after age 55 years (13.1%), either parent breaking a hip (11.8%), and corticosteroid use (0.1%) than in the observational studies. They were also more likely to be physically inactive (19.2%), currently smoking (7.9%), and taking treatment for diabetes (4.8%). The women in the clinical trial had volunteered to participate, were taking trial-required medications, and were following diet plans. The authors tested the addition of BMD to the model by testing ROC curves for the algorithm, DXA, or both in a subset of women with BMD measurements (10,750).		
Risk of Bias		**Dev**	**Val**
Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data? Trial data were more restrictive in inclusion and participations were randomized to treatments that may have affected the outcome.		PY	NA
Were all inclusions and exclusions of participants appropriate? No exclusions; all residents invited to participate. Exclusions not described in the manuscript, but per Wikipedia, "included medical conditions that would be predictive of a survival of less than three years, possessing characteristics or conditions that may diminish study adherence (e.g., substance abuse, mental illness, or cognitive impairment), or concurrent enrollment in another randomized controlled clinical trial."		NA	NA
Was there sufficient representation of individuals from racial and ethnic groups in model development data? In the observational study, 94% were White, 2.4% were Black, 1.0% were Hispanic, and the rest were a combination of American Indian, Asian/Pacific Islander, and unknown. Race/ethnicity for the trial and the subset with BMD not provided.		N	NA
Were racial and ethnic groups classified/categorized in a similar way in the development data and population to whom model is applied? (Validation studies only) The model validation cohort's race/ethnicity data were not reported.		Unclear	NA
Risk of bias introduced by selection of participants (low/high/unclear)	Unclear Appropriate data sources, no exclusions, but validation cohort race/ethnicity NR.		
Applicability			
Describe included participants, setting and dates:	Development cohort were predominantly White, applicability unclear to other races.		
Concern that the included participants and setting do not match the review question (low/high/unclear)	Unclear Development and validation cohorts broadly representative of White women but probably not other women.		

Abbreviations: BMD=bone mineral density; Dev=development; DXA=dual-energy x-ray absorptiometry; N=no; NA=not available; NR=not reported; PY=probably yes; RCT=randomized, controlled trial; ROC=receiver operator characteristic; Val=validation.

Appendix G Table 28. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 2 Predictors

Item	Response	
List and describe predictors included in the final model, e.g., definition and timing of assessment:	Age per each year: 1/2 point per year >50 Self-reported health Fair or poor vs. excellent: 3 points Good vs. excellent: 1 point Very good vs. excellent: 0 point Height per each inch:1/2 point per inch >64 Weight per each pound: 1 point per 25 lb <200 Fracture on or after age 55 y Not applicable vs. no: 0 point Yes vs. no: 2 points Race/ethnicity: White= 3 point Physical activity, metabolic equivalent tasks (METs): 1 point Smoking status Current vs. never: 3 points Parent broke hip, yes vs. no: 1 point Corticosteroid use, yes vs. no: 3 points Use of hypoglycemic agent, yes vs. no: 2 points 5-year followup was the time point for prediction.	
Risk of Bias	**Dev**	**Val**
Were predictors defined and assessed in a similar way for all participants? Routine care database so consistency of data collection for items such as family history-taking is unknown. Unclear how missing data were handled.	PY	NA
Were predictor assessments made without knowledge of outcome data?	PY	NA
Are all predictors available at the time the model is intended to be used?	PY	NA
Did the model avoid using race and ethnicity as a proxy for a biological or other risk factor that could be measured with more accuracy or fidelity? Race and ethnicity were not used in the model	N	NA
Was differential missingness of predictor data in racial and ethnic groups considered? Not applicable, race was not used	N	NA
Risk of bias introduced by predictors or their assessment (low/high/unclear)	Low Appears that predictors were collected in the same way.	
Applicability		
Concern that the definition, assessment or timing of predictors in the model do not match the review question (low/high/unclear)	Low These predictors are probably more intensive than routine collection in primary care (e.g., metabolic equivalents) but are feasible.	

Abbreviations: Dev=development; N=no; NA=not applicable; PY=probably yes; Val=validation.

Appendix G Table 29. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 3 Outcome

Item	Response		
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:	Incidence of hip fracture was collected using a standardized medical update questionnaire completed by all participants. These were collected every 6 months for those in the clinical trial and annually for those in the observational study until the study closed between October 2004 and March 2005. Hip fractures were self-reported and then confirmed both locally and centrally by review of medical records including x-ray and surgical reports. It appears that non-self-reported fractures were not being counted but likely with hip fractures; this may not have been a big number because it is such a big event.		
Risk of Bias		**Dev**	**Val**
Was the outcome determined appropriately? Self-report confirmed by medical reports (X-ray and surgery).		PY	NA
Was a prespecified or standard outcome definition used?		PY	NA
Were predictors excluded from the outcome definition?		Y	NA
Was the outcome defined and determined in a similar way for all participants?		PY	NA
3.4a Was differential followup or ascertainment of the outcome in racial and ethnic groups considered? All of the participants, including those who agreed to being followed up after dropping out of the interventions, are used in this analysis. Participants with missing data in their predictor variables, and 5.5% (n=5,161) of the participants who did not have a hip fracture within 5 years or did not have 5 years of followup were excluded from the logistic regression model. Unclear how many had missing data. The participants who were excluded from the logistic regression model tended to be minorities (28% vs. 16%).		N	NA
Was the outcome determined without knowledge of predictor information?		Y	NA
Was the time interval between predictor assessment and outcome determination appropriate? For the observational study, women were followed for a mean (SD) of 7.6 (1.7) years (median, 7.9 years; interquartile range, 6.9–8.9 years) The mean (SD) followup time for women in the clinical trial was 8.0 (1.7) years (median, 8.0 years; interquartile range, 7.4–9.0 years).		PN	NA
Were proxy outcomes avoided as the predicted outcome, where the meaning of the proxy may differ in racial and ethnic groups (label choice bias)? White race used as proxy		N	NA
Risk of bias introduced by the outcome or its determination (low/high/unclear)	High Reasonable duration (although not 10 years), standardized outcome collection, but N excluded for missing predictors unclear. Potential differential attrition by race, and race is part of the model.		
Applicability			
At what time point was the outcome determined:	Mean of 8 years		
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	Not applicable		
Concern that the outcome, its definition, timing, or determination do not match the review question (low/high/unclear)	Low Broadly applicable population		

Abbreviations: Dev=development; N=no; NA=not applicable; PN=probably no; PY=probably yes; SD=standard deviation; Val=validation; Y=yes.

Appendix G Table 30. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response
Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:	Developed using the observational study component of the Women's Health Initiative (93,676 women, 1,132 hip fractures, 0.16%), then validated it using the sample of women enrolled in clinical trials (68,132 women, 791 hip fractures, 0.14%) Candidate predictors (11 listed below): Number of hip fractures Age per each year: 1/2 point per year >50: 50–59: 102; 60–69: 359; 70–79: 671 Self-reported health: NR Height per each inch: NR Weight per each pound: NR Fracture on or after age 55 y: 313 Race/ethnicity: White: 1,064 Physical activity, metabolic equivalent tasks (METs): 0: 181; <5: 241; 5–12: 292; ≥12: 395 Smoking status Current: 565 Parent broke hip, yes: 240 Corticosteroid use, yes: 41 Use of hypoglycemic agent, yes vs. no: NR
Describe how the model was developed (for example, in regards to modeling technique [e.g., survival or logistic modeling], predictor selection, and risk group definition):	Potential risk factors were identified from the literature and fit 1 at a time in a Cox proportional hazards model, adjusting for age and race/ethnicity. Variables that achieved a modest level of statistical significance (p<0.25), based on the score test were included in the pool of variables used to select a final prediction model. Tenfold cross-validation was used to determine the optimal number of predictors; the training data were divided into 10 parts. Nine-tenths of the data were used to select the best model with k predictors by fitting a hazard regression model, which uses stepwise addition and deletion and considers interactions and nonparametric (spline) terms. For each model, they then evaluated the prediction log-likelihood on the remaining one-tenth of the data that were not used to select the model. For each k, they added the predicted log likelihoods to obtain a prediction score. The value of k that minimized the cross-validated prediction score was taken to be the optimal number of predictors. A hazard regression model with K* predictors was then selected from the entire WHI observational study data. The probability of a hip fracture within 5 years was then calculated using a multivariate logistic regression model fit on the WHI observational study dataset, using the K* variables from the earlier exercise
Describe whether and how the model was validated, either internally (e.g., bootstrapping, cross-validation, random split sample) or externally (e.g., temporal validation, geographical validation, different setting, different type of participants):	Receiver operator characteristic (ROC) curves and the corresponding area under the curve (AUC) in the clinical trial data were used to evaluate how the prediction model preformed on the test data. The 95% confidence intervals (CIs) were obtained by bootstrapping
Describe the performance measures of the model, e.g., (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:	The Hosmer-Lemeshow statistic was used to ascertain lack of fit (calibration) of this model.
Describe any participants who were excluded from the analysis:	Participants with missing data in their predictor variables, and 5.5% (n=5,161) of the participants who did not have a hip fracture within 5 years or did not have 5 years of followup were excluded from the logistic regression model.
Describe missing data on predictors and outcomes as well as methods used for missing data:	See above, not described clearly for predictors vs. outcomes.

Appendix G Table 30. Women's Health Initiative Fracture Risk Model Development Cohort Assessment From Prediction Model Study Risk-of-Bias Assessment Tool—Domain 4 Analysis

Item	Response	
Risk of Bias	**Dev**	**Val**
Were there a reasonable number of participants with the outcome? PN for hip fractures	PN	NA
4.1a Were there sufficient outcomes in racial and ethnic groups to assess model performance separately in these groups? (Model validation studies) Race included in the model, but majority of the population was White, and attrition was skewed to minority participants	PN	NA
Were continuous and categorical predictors handled appropriately? BMI was entered as a categorical variable, others were handled as expected (categorical or continuous)	PY	NA
Were all enrolled participants included in the analysis? Some dropout as described above	N	NA
Were participants with missing data handled appropriately? No adjustment	N	NA
Was selection of predictors based on univariable analysis avoided?	PY	NA
Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately? Not a competing risk model. Unclear how other complexities were addressed.	NI	NA
4.6a Was differential life expectancy in racial and ethnic groups accounted for using competing risk methods? No	N	NA
Were relevant model performance measures evaluated appropriately? Only Hosmer-Lemeshow reported over for development cohort. For the validation cohort, a table comparing observed and predicted fractures against threshold values (T-score above and below -2.5, WHI algorithm score above and below 21 points)	N	NA
4.7a Were relevant model performance measures evaluated appropriately in racial and ethnic groups? How does model performance (calibration, discrimination) compare in racial and ethnic groups? Model performance measures not reported separately in different racial and ethnic groups.	N	NA
Were model overfitting and optimism in model performance accounted for? Yes	Y	NA
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis? Unclear	PY	NA

Item	Response
Risk of bias introduced by the analysis (low/high/unclear)	High Differential racial missingness not addressed, performance measures not fully reported other than Hosmer-Lemeshow for calibration

Abbreviations: AUC=area under the curve; BMI =body mass index; CI=confidence interval; Dev=development; MET=metabolic equivalents; n=number; N=no; NA= not available; NI=no information; NR=not reported; PN=probably no; PY=probably yes; ROC=receiver operator characteristic; WHI=Women's Health Initiative; Val=validation; Y=yes.

Appendix H. Ongoing Studies

Relevant KQ	Title Trial Registry #	Intervention Comparator	Primary Outcomes	Estimated Completion Date
KQs 1 and 3	Models of Primary Osteoporosis Screening in Male Veterans (MOPS) NCT04079868	Intervention: Osteoporosis screening, education, and followup handled centrally by the bone health team Control: No practice management support	Screening rates (%), medication discontinuation (days), medication initiation (%), medication implementation (% of days covered with medication), bone mineral density (gram/sq centimeter), harms (%), primary care provider time; outcomes measured at 2 or 5 years	August 2024
KQs 1 and 3	Effects of FRAX+SARC-F Pre-screening on Preventing Fragility Fracture and Fall in Community-Dwelling Older Adults NCT04709393	Intervention: Receiving FRAX+SARC-F questionnaire prescreening results on estimated fracture risk Control: Not receiving FRAX+SARC-F questionnaire prescreening preliminary results on estimated fracture risk	Proportions of participants diagnosed with osteoporosis in the FRAX+SARC-F prescreening and control groups. Time frame: within 1–6 months	December 2028
KQs 4 and 5	Preventing Osteoporosis Using Denosumab NCT02753283	Intervention: Denosumab, then zoledronic acid Control: Placebo then zoledronic acid Both groups also received vitamin D and calcium	Bone density (total hip and spine)	September 2023

Abbreviations: FRAX=Fracture Risk Assessment Tool; KQ=key question; NCT=National Clinical Trial.