



U.S. PREVENTIVE SERVICES TASK FORCE PROCEDURE MANUAL

AHRQ Publication No. 08-05118-EF
July 2008

Table of Contents

Section 1.	Overview of U.S. Preventive Services Task Force Structure and Processes.....	1
Section 2.	Topic Selection, Prioritization, and Updating.....	14
Section 3.	Topic Work Plan Development.....	21
Section 4.	Evidence Report Development.....	32
Section 5.	Methods for Arriving at a Recommendation.....	44
Section 6.	Other Considerations for Recommendations.....	56
Section 7.	Dissemination of USPSTF Recommendations.....	58
Section 8.	References.....	59

Acknowledgements.....	60
Members of the U.S. Preventive Services Task Force.....	61

Appendixes.....	62
I. Congressional Mandate.....	63
II. USPSTF Disclosure Form.....	64
III. Surveillance Literature Summary.....	67
IV. Roles and Responsibilities of USPSTF Topic Work Group Lead.....	69
V. Solicitation for Nominations for Health Topics.....	70
VI. Work Assignment/Plan Template.....	73
VII. Criteria for Internal Validity.....	84
VIII. Criteria for External Validity.....	88
IX. Standardization of Outcomes Tables.....	91
X. Factors for Starting, Stopping Times for Screening.....	93
XI. Factors for Screening Intervals.....	95

Tables

Table 2.	USPSTF Levels of Certainty Regarding Net Benefit.....	46
Table 3.	Letter Grade of Recommendation or Statement of Insufficient Evidence Assessing Certainty and Magnitude of Net Benefit.....	53
Table 4.	What the USPSTF Grades Mean and Suggestions for Practice.....	55

Figures

Figure 1.	Overview of Process of Recommendation Development.....	9
Figure 2.	Procedures for Development of a Recommendation Statement.....	10
Figure 1a.	Steps in Topic Prioritization.....	20
Figure 3.	Template of an Analytic Framework with Key Questions.....	26

U.S. PREVENTIVE SERVICES TASK FORCE PROCEDURE MANUAL July 2008

Section 1: Overview of U.S. Preventive Services Task Force Structure and Processes

1.1 Purpose

The purpose of this procedure manual is to document the methods used by the U.S. Preventive Services Task Force (USPSTF or Task Force), staff of the Agency for Healthcare Research and Quality (AHRQ), and the AHRQ-designated Evidence-based Practice Centers (EPCs) in developing reviews and recommendations for clinical preventive services.

The USPSTF is charged with making evidence-based recommendations on a wide range of preventive services. AHRQ staff is charged with providing methodologic, scientific and administrative support to the USPSTF. The EPCs are charged with assisting the work of the Task Force by developing an extensive series of technical reports, evidence summaries, and other documents that update topics and that assess entirely new topics. The USPSTF, AHRQ, and the EPCs are dedicated to ensuring high-quality, consistent, and unbiased procedures for reviewing the evidence and making recommendations.

The methods outlined in this procedure manual are intended to ensure that the products created in this process are methodologically sound, scientifically defensible, reproducible, and well documented. This manual may be amended by vote of the Methods Work Group. It is expected that any issues that require broader discussion will be brought to the entire Task Force, and that any necessary changes to the manual made as a result of such discussions will be inserted after each USPSTF meeting. In addition, a sub-group of the Methods Work Group will review the entire Procedure Manual at least annually. The Procedure Manual will be posted on the USPSTF website (<http://www.preventiveservices.ahrq.gov>).

Additional information about the Task Force and its processes can be found in a series of published articles (References 1, 2, 3, 4, and 5).

1.2 Intended Audience

The Procedure Manual is intended to serve as a user's manual for everyone on the USPSTF team—including AHRQ and EPC staff as well as Task Force members. It is designed primarily for internal use as a guide to developing USPSTF recommendations, but may also be of interest to researchers or methodologists. It reflects the current procedures of the USPSTF team.

1.3 History of the U.S. Preventive Services Task Force; Scope of Work

The U.S. Preventive Services Task Force (USPSTF), first convened by the U.S. Public Health Service in 1984, is a leading independent panel of nationally recognized non-federal experts in

prevention and evidence-based medicine. Programmatic responsibility for the Task Force was transferred to AHRQ in 1995. (See Appendix I, Congressional Mandate Establishing the USPSTF.)

The first Task Force concluded its work in 1989 with the publication of the Guide to Clinical Preventive Services. A second Task Force, appointed in 1990, concluded its work with the release of the second edition of the Guide to Clinical Preventive Services in December 1995. In 1998, members of the third Task Force were appointed for five-year terms. The third Task Force released its recommendations incrementally. The current Task Force features a rolling panel of members appointed for 4 years, with the possibility of a 1- or 2-year extension.

Since its inception more than 20 years ago, the USPSTF has worked to fulfill its mission of:

- 1) Assessing the benefits and harms of preventive services in people asymptomatic for the target condition, based on age, gender, and risk factors for disease; and
- 2) Making recommendations about which preventive services should be incorporated routinely into primary care practice.

Task Force recommendations are intended to improve clinical practice and promote the public health. The Task Force's scope is specific: its recommendations address primary or secondary preventive services targeting conditions that represent a substantial burden in the United States and that are provided in primary care settings or available through primary care referral.

Primary preventive measures in a clinical setting are those provided to individuals to prevent the onset of a targeted condition (for example, the routine immunization of healthy children), whereas *secondary preventive measures* identify and treat *asymptomatic* persons who have already developed risk factors or preclinical disease but in whom the condition has not become clinically apparent (for example, screening for diabetes or colon cancer). Accordingly, most counseling topics and chemoprevention would today be categorized as primary prevention. Preventive measures that are part of the treatment and management of persons with clinical disease are usually considered *tertiary prevention* and are outside the scope of the USPSTF.

While the main audience for Task Force recommendations is the primary care clinician, the recommendations also have relevance for and are widely used by policymakers, managed care organizations, public and private payers, quality improvement organizations, research institutions, and patients.

In 1997 the Agency for Healthcare Research and Quality (AHRQ) launched its initiative to promote evidence-based practice in everyday care through establishment of 12 Evidence-based Practice Centers (EPCs). The EPCs develop evidence reports and technology assessments on topics relevant to clinical, social and behavioral, economic, and health care organization and delivery issues—specifically, issues related to those health services that are common, expensive, and/or significant for the Medicare and Medicaid populations. Since 1997, AHRQ has contracted primarily with the Oregon EPC to conduct systematic evidence reviews which serve as the foundation for USPSTF recommendations. AHRQ also contracts with other EPCs that have expertise related to individual topics of interest to the USPSTF.

1.4 USPSTF Members

There are currently 16 members on the Task Force. They represent an array of health-related disciplines including internal medicine, family medicine, behavioral medicine, pediatrics, obstetrics/gynecology and nursing.

1.4.1 Selection of USPSTF Members

Each year, a notice is placed in the Federal Register soliciting nominations for new USPSTF members. This notice is circulated to all USPSTF Partner organizations and distributed via AHRQ's prevention listserv. (See section 1.8 for a definition of a USPSTF partner organization.) Individuals nominated but not appointed in previous years, as well as those newly nominated, are considered in the annual selection process.

Nominated individuals are selected for the Task Force on the basis of the qualifications outlined below and the current needs of the Task Force for particular areas of expertise. Strongest consideration is given to individuals who are recognized nationally or internationally for scientific leadership within their fields of expertise. Applicants must have no substantial conflicts of interest that would impair the scientific integrity of the work of the Task Force, including financial, intellectual, or other conflicts. The AHRQ Director appoints new members.

In order to qualify for nomination to the Task Force, an applicant must demonstrate the following:

- 1) Knowledge and experience in the critical evaluation of research published in peer reviewed literature and in the methods of evidence review;
- 2) Understanding and experience in the application of synthesized evidence to clinical decision-making and/or policy;
- 3) Expertise in disease prevention and health promotion;
- 4) Ability to work collaboratively with peers; and
- 5) Clinical expertise in the primary health care of children and/or adults, and/or expertise in counseling and behavioral interventions for primary care patients. Some Task Force members without primary health care clinical experience may be selected based on other expertise such as medical decision-making, clinical epidemiology, behavioral medicine, and health economics.

1.4.2 Terms of Members

In 2001 the USPSTF underwent a transition to a standing Task Force. Currently, members are invited to serve for a 4-year term, with a possible 1-2 year extension. New members are selected each year to replace those who have completed their appointments. As a result, approximately one fourth of the Task Force is replaced each year.

1.5 USPSTF Meetings

The Task Force meets 3 times a year, in March, July, and November. Meetings are by invitation only. Representatives from USPSTF partner agencies and organizations have standing invitations. Special guests are invited to attend meetings for specific purposes.

1.5.1 Quorum and Voting

A quorum for official votes is 10 members, including the chair. Voting procedures include the following:

- 1) Formal votes are taken for major procedural and methodological decisions, for final recommendations, and for statements about clinical practice. Votes may be taken for other decisions at the discretion of the chair.
- 2) Votes are taken by voice or hand, without secret ballots.
- 3) Votes are recorded as yes, no, abstain, or absent. Members recused by reason of potential conflict of interest are recorded as recused and do not vote.
- 4) Reconsideration of a previously voted statement requires approval of two thirds of those present or 8 members, whichever is larger.
- 5) Members who are unable to be present at the time a vote is taken may leave proxy votes with the chair.
- 6) An electronic vote may be taken if a quorum is not available, or at the discretion of the chair. Members will have one week to respond to an electronic poll. The requirements for a quorum, for recording votes, and for determining a majority are the same as stated above.
- 7) In votes that are less than unanimous, there are no minority reports.

1.5.2 Conflict of Interest

The public must have confidence in the integrity of the process by which the Task Force makes its recommendations. The reputations of the Task Force members as highly regarded researchers, clinicians, and academicians contribute to this objective and must be protected if the Task Force recommendations are to be accepted and implemented. It is also essential that Task Force deliberations benefit from members' vigorous exchange of perspectives that are derived from and shaped by the member's research and/or practice experiences.

Prior to each meeting, Task Force members are asked to disclose any information that may interfere with their abilities to discuss and/or vote on a specific topic. Conflicts may arise, for example, if a member has a financial, business/professional, and/or intellectual interest in areas related to a particular topic. All members are expected to provide full disclosure of their interests related to all topics that will be discussed at each meeting. A committee comprised of AHRQ staff and the USPSTF Chair and Vice Chair review each member's disclosures and issue a recommendation on the member's eligibility to participate on a specific topic(s). Each member is notified by AHRQ staff of the recommendation prior to each meeting. Members are free to

recuse themselves voluntarily from participation in the processes for specific topics; however, a voluntary recusal does not free a member from the obligation to disclose a conflict.

Completed Disclosure Forms and recommendations are kept on file at AHRQ. (See Appendix II.)

1.5.2.1 USPSTF Conflict of Interest Procedures for Task Force Members

To achieve the objectives of the Task Force, AHRQ requires the following procedures:

- 1) Completion of Disclosure Form of potentially conflicting interests; and,
- 2) Review of completed Disclosure Forms and determination of appropriate actions

1.5.2.2 Process for Disclosure

The USPSTF Disclosure Form will be completed by Task Force members prior to each meeting to provide information to AHRQ on financial, business/professional, and intellectual potential conflicts of interest related to the topics addressed. Task Force members are expected to provide full disclosure for new topics, and an updated disclosure, reflecting changes in their situation since the form was last completed, for continuing topics. The USPSTF Disclosure Form will also be completed by new Task Force members prior to their participation. Completed Disclosure Forms will be kept on file by AHRQ.

EPC members will file separate disclosure forms which will also be kept on file with AHRQ.

Although AHRQ's intention is to keep personal information confidential, it is possible that information may be shared with the public if requested under the Freedom of Information Act.

1.5.2.3 Process for Determining Appropriate Actions

AHRQ staff will review the Disclosure Forms and consult as appropriate with the Task Force chair and/or vice-chair in recommending one of the following:

- A) *No action.*
No disclosure or recusal necessary.
- B) *Information disclosure to Task Force only.*
Member may participate as topic lead, and may discuss and vote on the topic.
- C) *Recusal from participation as lead of topic workgroup; information disclosure to Task Force.*
Member may discuss and vote on the topic.
- D) *Recusal from all participation; information disclosure to Task Force.*
Member may not participate as topic lead, and may not discuss or vote on the topic.
Member will leave the meeting room for all discussion and voting
Publicly released recommendations will denote the member's recusal from participation and voting on this topic.

Prior to the meeting or to new member participation, AHRQ staff will notify the Task Force member of the recommended action, and AHRQ's recommendation will be kept on file. If a Task Force member feels that a more conservative action is appropriate than that recommended, he or she can withdraw from any part of the process for that topic.

Task Force member contributions are invaluable and essential to the process of the Task Force. To that end, AHRQ staff, in consultation with the Task Force chair/vice-chair, will make recommendations always keeping in mind the overall goal of protecting the transparency, integrity, and acceptability of Task Force recommendations and products.

1.6 Members' Public Activities

Task Force members are encouraged to discuss, disseminate, and defend Task Force recommendations in various forums. To this end, Task Force members may provide expert testimony or case review on prevention-related health issues relevant to topics discussed by the USPSTF, with the following limitations/considerations:

- 1) If a Task Force member has provided expert testimony or expert review in a medical malpractice case on a given Task Force-related topic in the past five years, that Task Force member should disclose this information in the current USPSTF Disclosure Forms and the current disclosure protocol will be followed:
 - a. the USPSTF Disclosure Form will be completed with updated information by Task Force members prior to each meeting;
 - b. AHRQ staff will review the Disclosure Forms and consult as appropriate with the Task Force chair and/or vice-chair and make a recommendation for action;
 - c. prior to each meeting or to new member participation, AHRQ staff will notify the Task Force member of the recommended action; and
 - d. AHRQ's recommendation will be kept on file.

- 2) In order to avoid any real or apparent financial conflict of interest, Task Force members should not derive greater than \$10,000 per year from medical expert testimony or expert review for medical malpractice cases. Annually, Task Force members will be asked to disclose any testimony or expert review in medical malpractice cases and attest to income over \$10,000 derived from such activity over the past year. Task Force members should provide this disclosure in the USPSTF Disclosure Forms and current disclosure protocol will be followed. If the above financial limit has been exceeded, AHRQ staff, Task Force chair and/or vice chair, and the member will discuss the appropriate recommendation for action related to the member's continuation on the Task Force.

1.7 Procedures for Writing Papers /Documents

Each individual designated as an author of a paper to be submitted to a journal, or other public document created on behalf of the Task Force, should have participated sufficiently in the work to take public responsibility for the content. Authorship credit should be based on (1) substantial contributions to the conception, design, analysis, or interpretation of data or literature; (2) participation in the drafting of the document or its revision for important intellectual content; and

(3) giving of final approval of the version to be published. All three conditions must be met. Members who participate in the work but do not meet these criteria should be listed, with their permission, in the acknowledgments.

The Task Force does not recognize "courtesy" authorship given to Task Force or EPC members based on nominal role or position within a working group. General supervision of the working group, and participation in conference calls or group discussions, are not sufficient for authorship.

The USPSTF works under severe time constraints in producing its products. Accordingly, Task Force and EPC members who wish to be authors should expect to provide component drafts, supporting materials, comment, and feedback on a timely basis to the lead author. (A 1-week turnaround is a typical benchmark.) Individuals who are not able to meet reasonable timelines for a given product should not expect to be listed as authors.

The order of authorship should be a joint decision of the coauthors. Because the order is assigned in different ways, its meaning cannot be inferred accurately unless it is stated by the authors. Authors may wish to explain the order of authorship in a footnote.

This policy is derived from the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*, from the International Committee of Medical Journal Editors. This document is available at <http://www.icmje.org/>

1.8 Partner Organizations

Partner organizations provide ongoing liaison to the USPSTF. They include the major primary care societies and Federal agencies that are stakeholders in the process and products of the Task Force. Partner organization representatives contribute their expertise to the evaluation process and help disseminate the work of the USPSTF to their members and constituents. They are invited to attend and observe the USPSTF meetings and are permitted to comment on the proceedings during the meetings. Partners are sent drafts of the evidence report and recommendation statement and may arrange for these documents to be reviewed in detail by content experts within their organizations. This opportunity for comment by partners is in addition to the peer review that is obtained from experts who are not involved in the Task Force process, and the peer review provided by journals, as described in the next section.

Federal partners currently include the Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid Services (CMS), Department of Defense (DOD) Military Health System, Department of Veterans Affairs (VA) Center for Health Promotion and Disease Prevention, Health Resources and Services Administration (HRSA), Indian Health Service (IHS), National Institutes of Health (NIH), and U.S. Food and Drug Administration (FDA).

Primary care partners currently include the American Academy of Family Physicians (AAFP), American Academy of Nurse Practitioners (AANP), American Academy of Pediatrics (AAP), American Academy of Physician Assistants (AAPA), American College of Obstetricians and Gynecologists (ACOG), American College of Physicians (ACP), American College of Preventive Medicine (ACPM), American Medical Association (AMA), American Osteopathic Association (AOA), America's Health Insurance Plans (AHIP), National Committee for Quality Assurance (NCQA), and National Organization of Nurse Practitioner Faculties (NONPF).

1.9 Review and Comment for USPSTF Reports and Recommendations

As is summarized in the step-by-step description of the process in the next section, review of documents by individuals outside the Task Force occurs at several different points during the development of evidence reports and recommendation statements. This section specifies the different steps that involve external review.

Evidence Reports: The USPSTF incorporates expert review and partner comment of background documents in order to confirm that all relevant outcomes are being considered, that relevant literature has been considered, and that the evidence presented for USPSTF consideration is accurate. For all topics that are new to the USPSTF, and for some updates, the evidence-based practice center selects content area experts to provide peer review at the work plan stage of an evidence review. These comments are shared with the Task Force, and appropriate adjustments made to the work plan.

Once an evidence review is in draft form, and prior to its presentation to the Task Force, it is sent to content area experts and to the USPSTF's federal partner organizations. Their comments are shared with the Task Force at the time of the topic's presentation at a Task Force meeting.

Recommendation Statements: Once the evidence for a topic has been presented at a Task Force meeting, the Task Force will typically vote on a recommendation at that same meeting. The language for the draft recommendation statement then takes shape over the weeks that follow the meeting. The draft recommendation, developed by AHRQ staff and Task Force topic leads, is sent to all federal and primary care partners of the USPSTF, as well as to appropriate clinical specialty societies. The USPSTF requests comments on the clarity, clinical usefulness, and scientific accuracy of the recommendation statement. The Task Force views its role as a decision maker engaged in a deliberative process. Throughout this process, the Task Force maintains its independence by making these decisions without outside influence by professional societies or governmental entities.

Ultimately, the end-users of the recommendations have opportunities to respond to the recommendation statements and their accompanying evidence reviews through editorials in peer-reviewed publications and through formal letters to the editor of peer-reviewed journals. These letters are an important source of feedback to the Task Force.

1.10 Overview of the Process from Topic Selection to Recommendation Dissemination

As the diagram below illustrates (Figure 1), there are three organizations involved in the process that results in formulating Task Force recommendations: the Task Force itself, the Agency for Healthcare Research and Quality (AHRQ), and Evidence-based Practice Centers (EPCs). Each plays a unique role in the process:

Task Force: The Task Force selects and prioritizes topics for review; defines the questions and outcomes of interest in an analytic framework; interacts with AHRQ and the EPC about evidence issues; synthesizes, judges and grades the level of the available evidence; determines the balance of benefits and harms; and makes the recommendation.

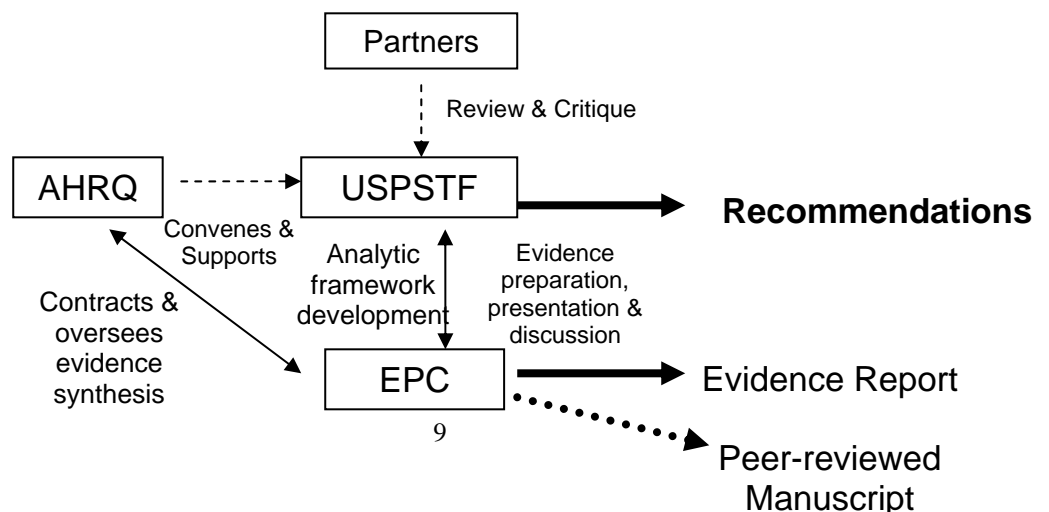
AHRQ: In addition to convening the Task Force, AHRQ provides ongoing administrative, research, and technical support for its operations, including coordination and support of the dissemination of the recommendations. Each topic is under the supervision of an AHRQ Medical Officer, who is responsible for all necessary technical input and coordination for that topic. In addition, AHRQ staff prepares evidence reviews in-house for some updated topics, including all reaffirmation topics.

EPC: Under contract to AHRQ, the EPC conducts systematic reviews of specified questions concerning the evidence on prioritized topics in clinical prevention. These EPC evidence reports serve as the scientific basis for USPSTF recommendations. The EPC's review process includes operationalizing the questions and outcomes of interest specified by the USPSTF for systematic review, locating and retrieving the relevant evidence, evaluating the quality of individual studies, qualitatively or quantitatively summarizing review findings for each question for use by the USPSTF in its evaluation of the evidence, and producing the reports.

USPSTF federal partners (CDC, CMS, FDA, IHS, NIH, and the VA) are invited to review and comment on draft work plans, evidence reports, and recommendation statements. In addition, USPSTF partners representing primary care professional organizations, including AAFP, AANP, AAP, AAPA, ACOG, ACP, ACPM, AMA, AOA, AHIP, NCQA, and NONPF are invited to comment on all draft recommendation statements prior to their release. (See section 1.8.)

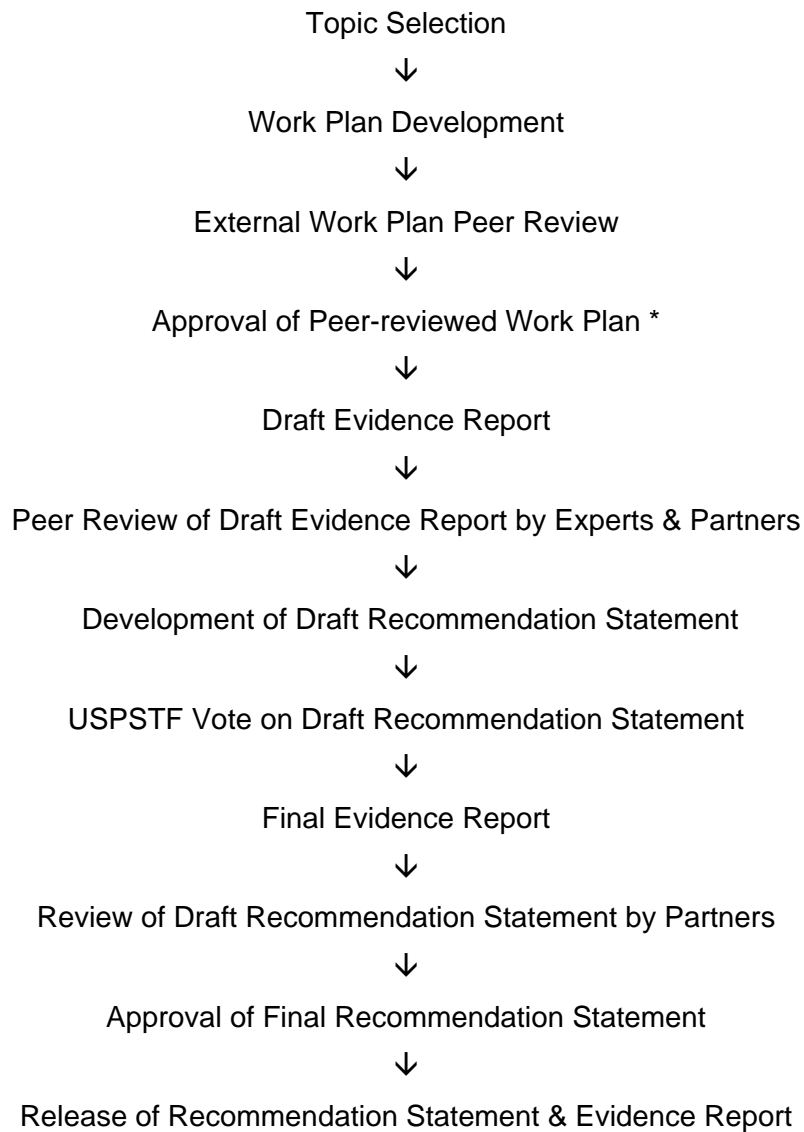
Figure 1. Overview of Process of Recommendation Development

Text Description: This chart consists of 4 rectangles, illustrating the relationships of the 4 entities that contribute to recommendation development: AHRQ, EPC, Partners, and USPSTF. The 3 first entities support the USPSTF, which eventually produces the recommendation.



The procedures for developing a recommendation statement are presented in Figure 2 below. A brief summary follows. However, each step will be described in more detail in later sections of the Procedure Manual.

Figure 2. Procedures for Developing a Recommendation Statement



*This step usually occurs at a Task Force meeting—although work plan review and Task Force approval are not usual in the case of topic updates.

1.10.1 Topic Selection

Topic selection begins with the identification of topics to be considered. Topics can be nominated by organizations, individuals, evidence-based practice centers (EPC), and Task Force members. Suggestions for new topics will be solicited every 2 years, through a notice in the Federal Register and notices to partner organizations. Periodically throughout the year, the Task Force Topic Prioritization Work Group drafts a prioritized list of topics including new topics and updates to be worked on during that year. This list is made according to criteria for prioritization: public health importance (burden of suffering and potential of preventive service to reduce the burden); potential change to a prior recommendation if there was one (for example, because new evidence has become available); and, potential for Task Force impact (controversy, practice not reflective of evidence, timeliness). The ultimate goal is to balance the annual portfolio by populations, types of services (screening, counseling, and preventive medications), disease types (cancer, endocrine system diseases, etc.) and size of projects (updates versus new topics). The Task Force also aims to update topics every 5 years in order to keep the recommendations in the Task Force library current. As part of their support of the work of the Topic Prioritization Work Group, AHRQ staff prepare a standardized summary of current literature in clinical prevention. This literature surveillance process is described in Appendix III

1.10..2 Work Plan Development

For every prioritized topic, a “topic team” is appointed. This team consists of 3-4 “leads” from the USPSTF, members of the AHRQ staff (including a Medical Officer), and evidence review team members, including a lead investigator. The responsibilities of the USPSTF topic team members are described in Appendix IV. AHRQ solicits Task Force members to volunteer to be leads, and issues a work assignment for the topics it assigns to the EPC. The EPC, working with AHRQ staff and Task Force leads, (or, in the case of an in-house evidence review, the AHRQ medical officer working with the Task Force leads) develops a preliminary work plan from the work assignment that contains the literature search strategy, analytic framework, key questions, and a timeline. AHRQ organizes a conference call of the entire topic team to discuss and refine the project scope and finalize the work plan. For new topics, work plans are sent for expert review. Work plans for topic updates are usually not reviewed by experts. If further communication regarding the work plan is needed, this communication is coordinated through AHRQ staff.

1.10.3 External Work Plan Peer Review

Work plans for new topics are sent to a limited number of outside experts in appropriate areas for their review and comments. Work plans for topic updates are not routinely sent to experts for review.

1.10.4 Approval of Work Plan

Work plans for new topics are presented by the EPC to the entire Task Force for the first time at a regular meeting. The EPC’s presentation is followed by comments from Task Force leads. The Task Force then discusses the plan, focusing on any issue of importance, but especially the Key Questions. The work plan is revised by the EPC as requested by the Task Force, and finalized. Work plans for topic updates are approved by the Task Force leads but are not routinely presented to the entire Task Force for discussion.

1.10.5 Draft Evidence Report

Based on the final work plan, the EPC conducts a systematic evidence review to address the questions posed by the Task Force and presents the resulting information to AHRQ in a draft evidence report, with evidence tables.

1.10.6 Review of Draft Evidence Report by Task Force Leads, Experts and Partners

All draft evidence reports are sent to a limited number of experts in the field and six federal USPSTF partners (CDC, CMS, FDA, IHS, NIH, and the VA) for review. In addition, Task Force leads are asked to comment on the draft evidence report.

1.10.7 Development of Draft Recommendation Statement

While the draft evidence report is under review and revision, AHRQ staff schedules one or more conference calls with the Task Force leads to discuss specific recommendations and the content of the clinical considerations section of the recommendation. The Task Force leads draft the recommendation statement with the AHRQ Medical Officer, and together they edit the draft recommendation statement, which is presented to the entire Task Force at its next meeting.

1.10.8 USPSTF Vote on Draft Recommendation Statement

At the Task Force meeting, a representative from the EPC presents the peer-reviewed evidence report findings, and the Task Force leads discuss the evidence and present the draft recommendation statement. The entire Task Force discusses the evidence and recommendation statement. Any proposed changes to the specific language of the recommendation are discussed. The Task Force votes on various formulations of the recommendation statement until one version gains the support needed. It usually takes from 9 to 15 months from the date the work plan is approved to the date the peer-reviewed evidence synthesis and draft recommendation statement are presented to the Task Force for a vote.

1.10.9 Final Evidence Report

Upon receiving all comments from experts, partners, the USPSTF (in particular, the topic leads), and the AHRQ Medical Officer, the EPC edits the evidence report. Often this finalization precedes the review process for the recommendation statement, and does not reflect any comments on the report from the partner comment process undertaken for the recommendation statement. The EPC sends a summary of all comments received and the revised evidence report, indicating how the comments were addressed to the AHRQ Medical Officer. After the AHRQ Medical Officer has reviewed and approved the revised document, the report is considered final. At this point, the EPC may undertake preparation of a manuscript for submission for publication in a peer-reviewed journal. An effort is made to synchronize the publication in the journal with the publication and/or release of the USPSTF recommendation statement.

1.10.10 Review of Draft Recommendation Statement by Partners

One of the Task Force leads or the AHRQ Medical Officer revises the recommendation statement based on discussion at the meeting. This revised recommendation statement is sent to all Task Force leads for approval. It is then sent to appropriate partners for comment.

1.10.11 Approval of Final Recommendation Statement

Task Force leads and the AHRQ Medical Officer discuss the partners' comments on the recommendation statement; Task Force leads decide on any changes. Depending on the magnitude of the proposed changes, the controversial nature of the topic, and the timing of next Task Force meeting, the revised (and hopefully final) recommendation statement may, in some unusual cases, be distributed electronically to Task Force members between meetings.

1.10.12 Release of Recommendation Statement and Evidence Report

An arrangement is often made with *Annals of Internal Medicine*, *American Journal of Preventive Medicine*, *Pediatrics*, *Annals of Family Medicine*, and other appropriate journals to simultaneously publish the recommendation statement (which is published by the journal without substantive editing) and the manuscript derived from the EPC report. The desired timeline from USPSTF vote to recommendation release is 9 months. All recommendation statements and supporting evidence reports are made available on AHRQ's website (www.preventiveservices.ahrq.gov).

Section 2: Topic Selection, Prioritization, and Updating

The USPSTF has a large library of current topics and frequently receives nominations for new topics. The overall goal for new topic selection and prioritization is to provide accurate and relevant recommendations that are as up to date as possible and to balance the overall portfolio of recommendations by populations, types of services (screening, counseling, preventive medications), disease types (e.g., cancer, endocrine disease) and size of project (e.g., update versus new topic). The Task Force also aims to update topics every 5 years, in order to keep recommendations in the Task Force library current according to criteria established by the National Guideline Clearinghouse, an AHRQ-initiated public resource for evidence-based guidelines (www.guideline.gov). The criteria for new topic selection and for prioritization (discussed in detail below) are combined in an assessment of the topic as a whole, rather than used as part of a scoring system.

2.1 Topic Types and Definitions

There are two types of topics in the Task Force library: active and inactive. Among the active topics, there are four categories for consideration: new, updated, reaffirmed, and referred; each of these categories is defined below. The processes for developing work plans, assessing evidence, and making recommendations for active topics are discussed in Sections 3, 4, and 5.

Active Topic Types

New topics are topics chosen by the USPSTF for review and recommendation that have not been previously reviewed.

Updated topics are topics reviewed in the past by the USPSTF that have since undergone an update of the evidence and recommendation. The update may encompass all key questions on a topic (full update) or only a limited set of the key questions in the analytic framework (targeted update).

Reaffirmed topics are topics kept current by the USPSTF because the topic is within the Task Force's scope and a Task Force priority, and because there is a compelling reason for the Task Force to make a recommendation. Topics that belong in this category are well established, evidence-based standards of practice in current primary care medical practice (e.g., screening for hypertension). While the Task Force would like these recommendations to remain active and current in its library of preventive services, it has determined that only a very high level of evidence would change the recommendation. Only recommendations with a current grade of A or D are considered for reaffirmation. The procedure for reaffirming a topic is discussed in Section 4.

Referred topics are topics in which the USPSTF refers providers to another

organization's recommendation. These topics were originally maintained by the Task Force and are retained as active in the Task Force library; however, the Task Force has determined that there is another organization (e.g., Centers for Disease Control and Prevention's [CDC's] Task Force on Community Preventive Services) with evidence-based methods that is better positioned to make accurate and timely recommendations for this topic. The procedure for referring to other organizations is discussed in Section 2.5.

Inactive Topics

Inactive topics are topics the USPSTF has decided to inactivate (e.g., electronic fetal monitoring, home uterine monitoring, and counseling for dental disease) for one or more reasons (see Section 2.4).

2.2 New Topic Prioritization and Selection

The Task Force solicits new topics for consideration through a periodic notice in the *Federal Register* (see Appendix V) and solicitation of professional liaison organizations, including, but not necessarily limited to, the USPSTF's partner organizations. New topic nominations may also be submitted by USPSTF members, members of the public, or others on an ongoing basis, outside of the *Federal Register* notice.

Topic nominations are first considered by the USPSTF's Topic Prioritization Workgroup, which then recommends a selection and prioritization of new topics to the entire Task Force.

The entire Task Force ultimately votes on the selection and priority of new topic nominations. The Topic Prioritization Workgroup receives support from AHRQ staff in gathering background information to facilitate topic selection and prioritization.

The USPSTF first considers whether newly nominated topics are within the scope of the USPSTF (that is, whether the topic relates to a primary or secondary preventive service that is relevant to primary care, and whether the topic address a disease with a substantial health burden). Topics that meet these first two criteria are then prioritized using the following specific criteria:

1. Public health importance (i.e., burden of suffering and expected effectiveness of the preventive service to reduce that burden).
2. Potential for a Task Force recommendation to affect clinical practice (based on existing controversy or the belief that a gap exists between evidence and practice).

2.3 Prioritization and Selection of Active Topics

The Topic Prioritization Workgroup begins prioritization of an active group of topics 2 years after their previous publication.

Step 1. AHRQ staff creates a 1- to 2-page background paper on the topic that includes the following information: previous recommendation statement, estimate of disease burden, relevance to prevention and primary care, existing controversy, and summary of a brief literature search for new evidence.

Step 2. The Topic Prioritization Workgroup reviews and discusses the background paper and places each topic into either the active or inactive category. Topics that are retained as active are considered for referral to other organizations. See Section 2.5 for the process of referring a topic. See Section 2.4 for the process of inactivating a topic.

Step 3. AHRQ staff prepares a request for feedback on all of the remaining active topics to be sent to USPSTF members, partner organizations, and other appropriate stakeholders. Respondents are asked to categorize each proposed topic as high, moderate, or low priority for review in the next 12–18 months, based on the following criteria:

1. Public health importance (i.e., burden of suffering and expected effectiveness of the preventive service to reduce that burden).
2. Potential for a Task Force recommendation to affect clinical practice (based on existing controversy or the belief that a gap exists between evidence and practice).
3. New evidence (e.g., new studies or new analyses of previous data) that has the potential to change prior recommendations.
4. Need for a balanced portfolio of topics.

Step 4. The feedback from USPSTF members, partner organizations, and stakeholders is considered by the Topic Prioritization Workgroup, along with the background paper, in assigning a draft priority category for active topics. The Topic Prioritization Workgroup also reviews and updates the priority of all other active topics (i.e., not currently under review) that have not been referred. The four criteria listed in Step 3, along with resource requirements for the review, are used to recommend priority (low, moderate, or high).

Step 5. The topic categorization and prioritization becomes final after a vote of the full USPSTF membership.

Steps 1–5 are repeated every 5 years (i.e., 7 years after publication) for topics that have not already been placed in the work queue.

AHRQ staff develops the work queue for the next 12- to 18-month cycle using the priority level determined by the USPSTF. Other factors that may be used by AHRQ staff in determining the work queue include: availability of research team, availability of review or funds from a non-USPSTF source, efficiency of combining reviews or research teams on related topics, impending release of relevant study, and age of relevant non-USPSTF review.

2.4 Inactivating a Topic

Inactive topics are topics the Task Force has decided to inactivate (e.g., electronic fetal monitoring, home uterine monitoring, and counseling for dental disease) for one or more of the following reasons:

1. Topic is no longer relevant to clinical practice due to changes in technology, new understanding of disease etiology/natural history, or evolving natural history of the disease.
2. Topic is not relevant to primary care setting, because the service is not implemented in a primary care setting or not referable by a primary care provider.
3. Topic has a low public health burden.
4. Topic is otherwise outside of the USPSTF's scope.

Previously inactivated or referred topics are also eligible as new topic nominations, if appropriate, along with other new topic suggestions.

If a topic is inactivated or referred to another organization, the status on the USPSTF Web site continues to be listed as "active" for a minimum of 5 years from the date of the original recommendation, unless considerations arise beforehand to change the status. After this period, the status changes to "inactive" or "referred."

2.5 Referring a Topic to Other Organizations

The recommendations for some topics in the USPSTF library may be referred to another organization that the Task Force believes is in a better position to make an accurate and timely evidence-based recommendation. This practice avoids redundancy of resource use by the USPSTF. An example is the Advisory Committee on Immunization Practices (ACIP), a non-Federal panel of immunization experts convened by the CDC. In the past, the Task Force has referred recommendations on immunizations to the ACIP. Another example is the CDC's Task Force on Community Preventive Services, which makes evidence-based recommendations on many health promotion topics. The organization identified for referral should have the resources for timely updates of the evidence and a scientifically acceptable methodology for its evidence reviews (see the list of criteria below). Other than immunizations, the Task Force anticipates that few topics will fall into this category. The process for designating a topic for referral is as follows:

1. The Topic Prioritization Workgroup identifies a potential outside organization that makes evidence-based recommendations and decides to consider the topic for referral.
2. AHRQ staff reviews the previous Task Force recommendation statement and evidence report.
3. AHRQ staff reviews the recommendations and review methods of other Federal agencies and professional organizations.

4. AHRQ staff prepares a brief summary of why the topic has been chosen for referral, a reference to the chosen organization's recommendations on the topic, a statement that the organization's methodology may be different from the USPSTF, a new recommendation date, and a statement that the previous evidence review will not be updated.
5. The Topic Prioritization Workgroup decides whether to proceed with a full Task Force discussion.
6. If the Topic Prioritization Workgroup decides to proceed, the AHRQ summary is presented at a Task Force meeting for general discussion. The Task Force then votes on the decision to refer the topic to the specific organization.
7. AHRQ staff adds a single summary paragraph to the USPSTF Web site that includes a link to the organization's recommendation.

Referred topics may be re-activated through the usual new topic nomination process (see Section 2.2). The criteria for referring to another organization's recommendation are:

1. The organization has been identified by the USPSTF as an appropriate source.
2. The organization has a process for updating recommendations in a timely manner.
3. The organization has a written and available evidence-based methodology, including the use of systematic reviews that assess benefits and harms, that the Task Force judges to be adequate for the topic.

2.6 Consideration of an Early Topic Update

Occasionally a study will be published after a recommendation's release that may potentially affect the USPSTF's consideration of the evidence and its conclusions about the certainty and/or magnitude of the net benefit (and the recommendation itself). These studies are brought to the attention of the USPSTF by a number of sources, including the public, USPSTF members, Evidence-based Practice Centers, professional organizations (including USPSTF partners), and advocacy groups. In addition, the USPSTF actively surveys the literature to ensure that the Task Force is aware of new evidence in a timely manner. The process for this active surveillance is discussed in Appendix III. The Task Force uses the following process to consider new evidence and decide whether a recommendation needs to be updated earlier than the usual 5-year timeframe:

1. The USPSTF Program Director at AHRQ is notified of the new evidence and assigns an AHRQ Medical Officer to review the study. Ideally, the Medical Officer who reviews the evidence is the same Medical Officer who performed the previous review. The Medical Officer fills out a review form with the following information:
 - Citation
 - Nominator, affiliation
 - Assigned Medical Officer

- Brief summary/abstract of study
- Number of criteria met (see below)
- Recommendation of Medical Officer/Program Director
- Summary of Topic Prioritization Workgroup and Task Force discussion (to be completed later in the process)
- Action/disposition (to be completed later in the process)

Criteria for evaluating evidence to trigger an early review (order is not necessarily based on criteria weighting):

- New evidence that conflicts with current recommendation
- Large-scale study that may improve certainty of net benefit
- Potential that new evidence will change recommendation grade
- Evidence on a new intervention/strategy not previously considered
- Change in magnitude of benefit or harm that might alter magnitude of net benefit
- Potential of evidence to fill gap in the chain of indirect evidence
- Existing controversy about topic
- Public health burden of condition
- Quality, relevance, or type of study (e.g., a randomized, controlled trial is published on a topic for which the current recommendation is based on observational evidence)
- Published in peer-reviewed journal
- Evidence that directly links the prevention strategy to the primary outcome of interest (i.e., direct evidence of health effect)
- Source of identification of the new evidence (e.g., professional organization, Task Force member, advocacy group)

2. The form and the Medical Officer's recommendation is sent to the Program Director and the AHRQ lead in the Topic Prioritization Workgroup.

3. If appropriate, a discussion of the evidence is placed on the agenda for the Topic Prioritization Workgroup's monthly conference call. If there is an identified current USPSTF member who is a topic lead or expert in the subject area and who is not a member of the Workgroup, then that Task Force member is invited to participate in the conference call. The evidence and the review form are sent to the call attendees with an agenda.

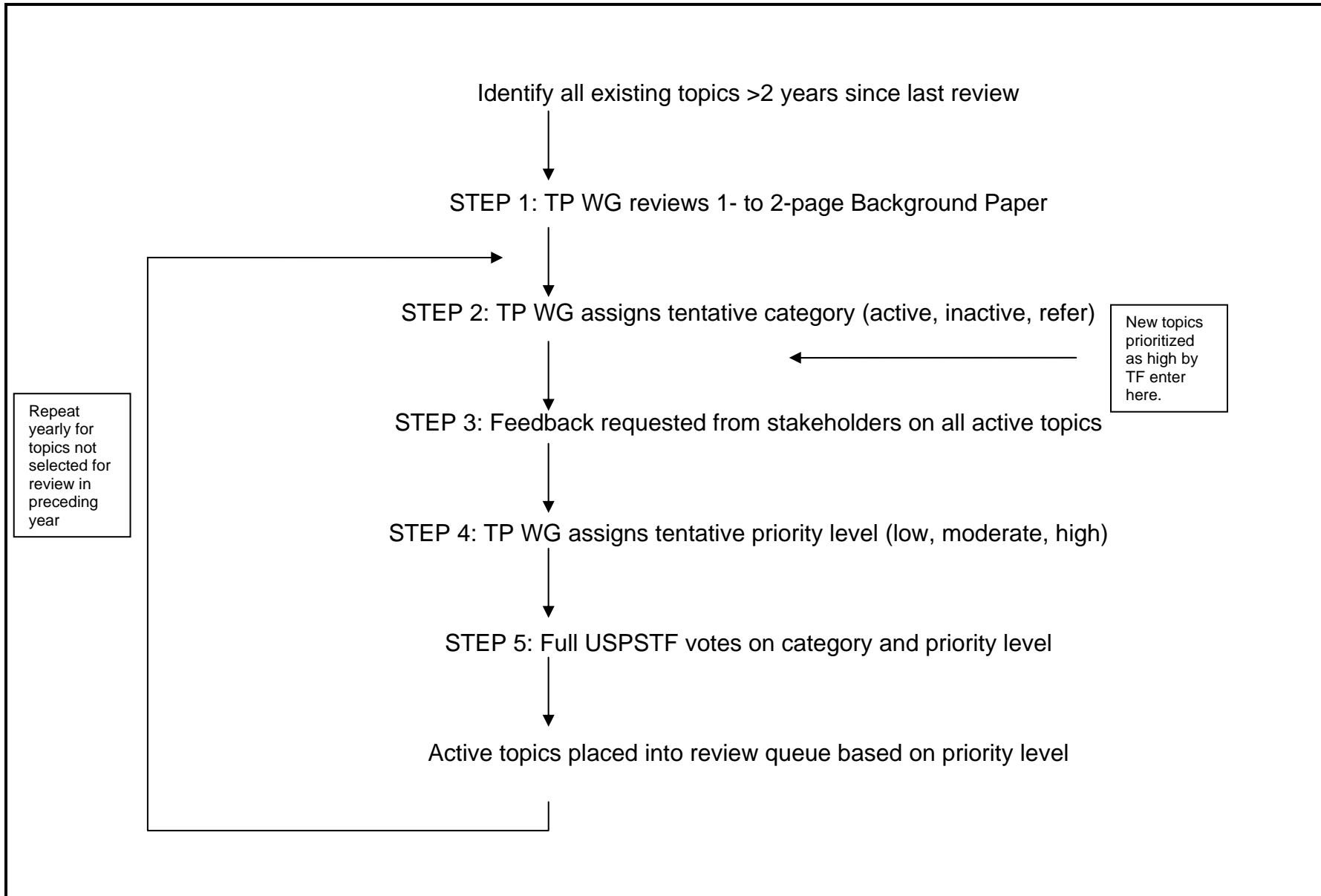
4. The Topic Prioritization Workgroup discusses the evidence and, using the criteria defined above, makes a recommendation to the entire Task Force about whether the evidence should trigger an early update of the review.

5. The USPSTF votes at its next meeting on whether the evidence update should be accelerated due to the new evidence. The Task Force also assigns a priority level (high, moderate, low) based on the usual topic prioritization criteria.

6. If the USPSTF decides to accelerate the update, the USPSTF Program Director at AHRQ places the topic in the review queue.

7. A brief notice from the USPSTF Chair is sent to the nominator about the disposition.

Figure 1. Steps in Topic Prioritization



Note: TP WG = Topic Prioritization Workgroup

Section 3

Topic Work Plan Development

When a topic is prioritized by the Topic Prioritization Work Group for a new or updated recommendation, the scope of the topic and approach to the review must be defined to guide the researchers undertaking the systematic review process. USPSTF leads volunteer for each topic as it begins, based on expertise and interest. An AHRQ medical officer is assigned to oversee the topic and develops a work assignment that outlines the major issues guiding the anticipated review. A lead investigator is assigned from the evidence review team with core review staff also identified. The review team refines the work assignment using the template described below and finalizes it through discussions with the Task Force leads and the AHRQ Medical Officer. For new topics, the work plan is peer-reviewed and presented to the entire USPSTF at one of its regular meetings. The final version of the work plan is approved by the AHRQ Medical Officer. Development of a work plan generally takes from 1 to 4 months, but may take longer.

3.1 Determining Topic Scope and Review Approach

3.1.1 Types of Reviews

The USPSTF has determined that using systematic reviews where possible represents the best method for fully and fairly considering all of the evidence relevant to questions about a clinical preventive service. In order to answer the relevant questions about a clinical preventive service, the EPC usually undertakes a series of related systematic reviews to answer each of the questions in the analytic framework. However, systematic reviews can be labor-intensive and expensive processes. In order to efficiently invest its limited resources in maintaining the most current evidence-based recommendations for the breadth of important clinical preventive services, the USPSTF has developed a series of approaches to defining the scope of the topic that the EPC will systematically review.

These approaches include commissioning:

- Full systematic reviews, ie, systematic reviews of every key question in the analytic framework. This is the default approach for a topic new to the USPSTF.
- Targeted systematic reviews, ie, systematic reviews for selected questions in the analytic framework for which the evidence is not clearly established and/or current. Many updates of USPSTF topics use this approach to review relevant evidence since the last systematic review.
- Staged reviews, ie, systematic reviews for selected questions in the analytic framework, undertaken in a hierarchical manner in order to inform the Task Force as to whether it is possible to establish evidence insufficiency in a parsimonious fashion. In rare cases, the Task Force may see that serious gaps in evidence preclude coming to a recommendation, and a staged review may confirm these gaps and thus lead to an I statement from the Task Force.

3.1.2 Principles for Determining the Review Approach

During work plan development, the topic team considers the scope of the evidence needed for the USPSTF to make its recommendation and the most efficient way to procure the needed information. For reviews undertaken to update existing USPSTF recommendations, this process is based on:

- 1) Examination of the previous USPSTF recommendation(s) including populations and clinical preventive services addressed, to determine their fit with current questions about the clinical preventive service;
- 2) Examination of the previous USPSTF evidence review process for this topic and the review findings in order to identify established evidence, important review limitations, and evidence gaps. For example, in the 1996 Guide to Clinical Preventive Services, behavioral counseling topics had two recommendations: one for the relationship between the behavior and health outcomes (eg, smoking and premature morbidity and mortality) and one for the efficacy of counseling. When updating such recommendations, it is often not necessary to update the behavior-health outcome link, as that is often well-established.
- 3) Determination of current contextual information (eg, changes in understanding of the nature of the disease process, or changes in diagnosis, therapeutics, practice or reimbursement; controversy over any of these elements)

In order to facilitate the consistent development of the review approach across topics, the USPSTF has developed a template to guide the development of the final Work Plan. This template can be considered analogous to a protocol developed for a Cochrane Review. It is also an articulation of the rationale for the scope decisions made in framing the topic.

3.2 Methods Relevant to Work Plan Development

The work plan template (see Appendix VI) is used by AHRQ in the creation of the work assignment, and by the EPC and the Task Force leads in the refinement of the work assignment to produce the final work plan. This template stimulates the thinking and guides the systematic consideration of the factors that experience has shown are important in planning a review to update an existing USPSTF recommendation. Since most reviews conducted for the USPSTF are updates, the work plan template was developed with that purpose in mind. However, the same template can be used to plan and guide the full systematic review for a new topic. Sections addressing the previous USPSTF recommendation, previous review findings, and priority evidence would obviously not be relevant or included.

3.2.1 Analytic Frameworks

The purpose of analytic frameworks is to present clearly in graphical format the specific questions that need to be answered by the literature review in order to convince the USPSTF that the proposed preventive service is effective and safe (as measured by outcomes that the USPSTF considers important). The specific questions are depicted graphically by linkages that relate interventions and outcomes. These linkages serve the dual purpose of identifying

questions to help structure the literature review and of providing an "evidence map" after the review for the purpose of identifying gaps and weaknesses in the evidence. Further details about the design of analytic frameworks are provided in a paper by Woolf et al., published by AHRQ in 1994 (6).

3.2.1.1 Analytic Frameworks Are Not Causal Pathways

Analytic frameworks as used by the USPSTF are not intended to depict comprehensively all factors and variables that cause patients to receive the preventive service or all factors and variables responsible for the intermediate outcomes and health outcomes associated with a preventive service. In short, they do not depict the "mechanism of action" for a preventive service. For example, an analytic framework for cervical cancer screening that is concerned mainly with two questions (whether Pap smears detect early disease and whether early detection lowers mortality) need not specify other co-variables: e.g., the risk factors for cervical cancer, the demographic characteristics of women who are more likely to be screened, the etiologic determinants of cervical cancer, or the pathologic progression of cervical cancer from atypia (atypical cell stage) through invasive disease.

3.2.1.2 Analytic Frameworks Are Not Decision Algorithms or Flowcharts

The use of arrows and boxes gives analytic frameworks the appearance of decision trees and flowcharts but the purpose is quite different. Analytic frameworks are not intended to depict all possible outcomes of a particular event, as is expected of decision analytic models, or to calculate their probabilities. A linkage in an analytic framework for mammography screening might ask whether early detection lowers mortality. The same step in a decision analytic model would be more likely to depict each component in that step (confirmation of suspicious lesion on mammogram, follow-up mammographic views, referral to specialist, biopsy and nodal evaluation, pathologic studies, lumpectomy or mastectomy, adjuvant chemotherapy, treatment for recurrent disease) and to depict decision nodes and probabilities for each step in the process. That is not necessary for analytic frameworks because the evidence question of interest to the USPSTF in this case is whether early detection lowers mortality.

3.2.1.3 Actions Versus Outcomes

Analytic frameworks used by the USPSTF distinguish between actions (eg, the performance of a screening test, or treatment with a drug) and outcomes such as the detection of a disease, reduced morbidity and mortality, a change in patients' behavior, or adverse effects. The performance of a screening test is not itself an outcome. Actions are depicted by arrows, whereas outcomes are depicted by rectangles (See Figure 3 below.)

3.2.1.4 Intermediate Outcomes Versus Health Outcomes

Analytic frameworks used by the USPSTF distinguish between intermediate outcomes and health outcomes. *Health outcomes* are symptoms and conditions that patients can feel or experience, such as visual impairment, pain, dyspnea, impaired functional status or quality of life, and death. *Intermediate outcomes* are pathologic and physiologic measures that may precede or lead to health outcomes. Elevated blood cholesterol level is an intermediate outcome for coronary artery disease; patients cannot "feel" the cholesterol elevation. Myocardial infarction, or death from coronary artery disease, is a health outcome. The USPSTF gives greater weight to evidence of an effect on health outcomes than to evidence of an effect

on risk factors or intermediate outcomes. The fact that a preventive service has a proven effect on an intermediate outcome does not necessarily establish that it can improve outcomes that are perceptible to patients.

When data are available, the Task Force considers data on both all-cause and cause-specific mortality in making its recommendations. (See section 4.5.3 for a fuller discussion of these outcomes.)

3.2.1.5 Conventions for Graphics and Layout

The analytic framework diagram contains three types of items (population, actions, and outcomes). Below the diagram are annotated questions that correspond to specific items in the diagram. The annotated questions are designated in the diagram by superscript symbols. The conventions that follow are illustrated in the accompanying prototype (Figure 3).

The population appears at the left-hand margin of the diagram and specifies the type(s) of patients to whom the evidence about the preventive service pertains. For example, the population description in an analytic framework for cervical cancer screening might be "females who are sexually active or over age 18." Footnotes in this location refer to specific questions about the population that the evidence review must answer in order to evaluate the effectiveness of the preventive service. For example, it may be necessary to know the proportion of the population with a family history of colon cancer.

Actions, such as screening and treatment, appear as arrows linking the population to an outcome or linking one outcome to another. A curved arrow leads to adverse effects of the action. The outcome to which the arrow points should result logically from the action (e.g., early detection of disease for screening, behavior change for counseling, reduced incidence of disease for immunizations or chemoprophylaxis). The name of the action (e.g., "screening with Prostate Specific Antigen [PSA]") appears below the arrow. Each arrow is a linkage in the logical chain of evidence that connects the left side (population) and the right side (health outcomes) of the analytic framework. Overarching linkages directly connect these two sides. Each arrow (linkage) is a "key question" that must be addressed by an evidence review. Outcomes are the intermediate outcomes or the health outcomes resulting from actions or from previous outcomes. A health outcome that follows an intermediate outcome, which typically reflects the natural progression of disease (eg, from "retinopathy" to "visual impairment") is depicted by a dotted line (no arrowhead). Health outcomes are depicted by rectangles with squared corners; intermediate outcomes are depicted by rectangles with rounded corners.

Annotated footnotes are specific key questions that are associated with each arrow and that must be answered by the literature review. The key questions are written in detailed narrative below the analytic framework figure.

3.2.1.6 Revisions

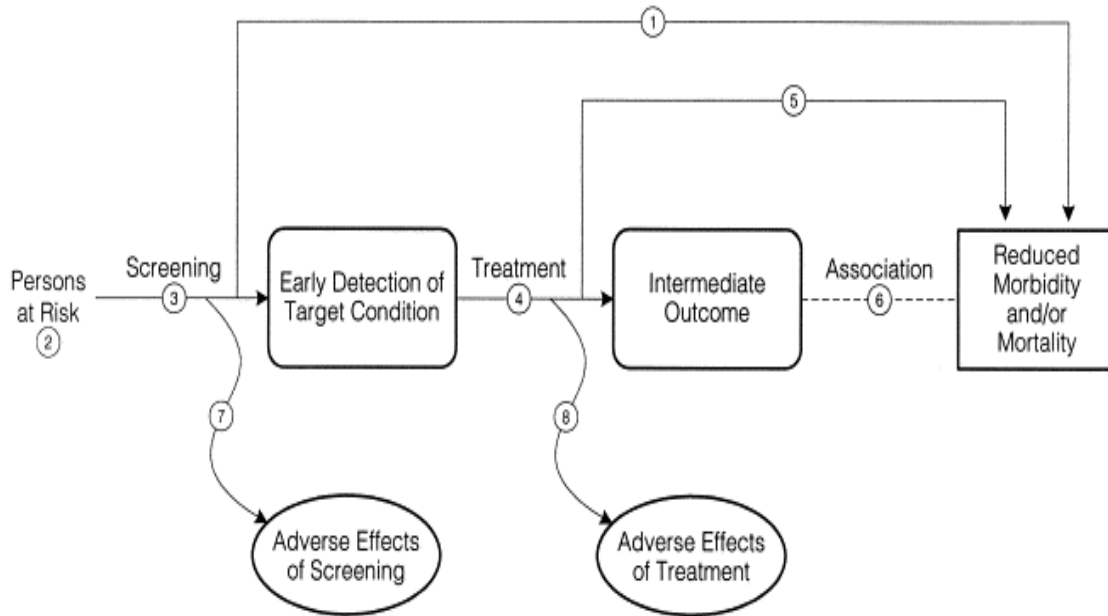
Analytic frameworks are evolving documents that can be influenced over time by the findings of the evidence review and by other considerations. For example, an overarching linkage might later be removed if the USPSTF found no relevant evidence. New linkages may be added when new interventions, outcomes, or logical arguments emerge during the course of the review.

3.2.2 Key Questions

Key questions are an integral part of the approach to conducting systematic reviews the USPSTF uses in its recommendation process. Along with the analytic framework, these questions specify the logic and scope of the topic, and are critical to guiding the literature searching, abstracting, and analysis process.

The purpose of the following definitions is to contrast and compare three types of questions (key questions, updating key questions, and contextual questions) that are used in formulating, conducting, and reporting systematic evidence reviews to update USPSTF recommendations. Updating key questions and contextual questions have been developed to address needs in the updating process, and their use is primarily in updated topic reviews. This function is reflected in their definitions.

Figure 3. Template of an Analytic Framework with Key Questions



Key Questions

1. Does screening for X reduce Morbidity and/or Mortality?
2. Can a group at high risk for X be identified on clinical grounds?
3. Are there accurate (i.e. sensitive and specific) screening tests available?
4. Are treatments available that make a difference in intermediate outcomes when the disease is caught early, or detected by screening?
5. Are treatments available that make a difference in morbidity or mortality when the disease is caught early, or detected by screening?
6. How strong is the association between the intermediate outcomes and patient outcomes ?
7. What are the harms of the screening test?
8. What are the harms of the treatment?

3.2.2.1 Definitions

- 1) **Key Questions**: Key questions, in association with the analytic framework, establish the clinical logic to support implementation of a clinical preventive service in primary care. Key questions articulate the details about relevant patients, interventions, outcomes, and comparisons—details that are essential in order to focus the review on a manageable and clinically relevant topic, and which directly guide the systematic literature review process. Each question is clearly tied to a step in the analytic framework, although certain linkages that are already well established may not have a key question that is actively answered during the review for the USPSTF. In addition, there may be reason to focus on an overarching linkage (and the associated key question) in an analytic framework rather than the intermediate linkages (and their associated key questions). All key questions are reviewed and approved by the topic team in the process of assessing and refining the topic before the detailed literature review is conducted. Key questions addressed in a systematic review are listed in the methods section and used to organize the results section of the report.

Key questions are addressed using up-to-date systematic review methods, under the current guidance and methods of the USPSTF. Each key question is addressed through a distinct literature search and reported separately in the results section of the review report.

- 2) **Updating Key Questions**: Updating key questions are used by the USPSTF to direct the researchers conducting the review to focus on a limited, specific set of key questions from the analytic framework during an updated review process. Updating key questions are used to focus the review for a topic update in two main ways: (1) encouraging the researcher to examine only critical gaps in the evidence that would be required by the USPSTF in order to make a recommendation; and (2) encouraging an evaluation of new research addressing parts of the topic that might confirm or change the evidence basis for a previous recommendation. An updating key question approach is taken in order to maximize efficiency in updating when, in the context of the previous review and recommendation, the USPSTF determines that a systematic review of only a limited part of the analytic framework is necessary in order to conduct a valid update of its evidence-based recommendation.

Updating key questions are addressed in the same manner as key questions and reported in the results section of the review report.

- 3) **Contextual Questions**: These questions are not key questions associated with the analytic framework; however, they represent issues in an updated review for which the USPSTF needs a valid but not necessarily systematic summary of current research in order to provide the context for its vote and recommendation statement. Contextual questions may elicit a range of different types of information, including: (1) updated information for a key question that is not being systematically updated; (2) contextual information on natural history, current practice, prevalence and risk groups, or other aspects of the service for which it is strongly believed there will not be information, but which are part of the Task Force's considerations (eg, screening interval, ages when

screening should be stopped; or newer technologies for screening and/or intervention); or 3) cost-effectiveness.

Contextual questions are not necessarily addressed systematically; however, the approach taken may meet criteria for a systematic review. Comprehensive literature searches are not generally undertaken specifically to answer these questions. Information for contextual questions is gathered in a variety of ways: (1) through targeted literature searches, (2) from authoritative surveys or reviews, (3) from expert input, and (4) opportunistically, while reviewing comprehensive literature searches for key questions. Contextual questions are not listed as separate questions in the methods section of the report and are not reported in the results section. The information resulting from non-systematic review should be included as part of the introduction or in the discussion section, and related as appropriate to the results of the systematic review.

3.3 Scan of Evidence/New Evidence

During the work assignment development for a topic update slated to be undertaken by the EPC, AHRQ conducts a preparatory literature scan to estimate the volume and content of intervening research relevant to this USPSTF topic. The scan focuses particularly on locating new research for critical key questions relating to benefits and harms of the preventive service. Literature searches are conducted in the Cochrane Library databases and in PubMed. Searches incorporate the MeSH terms from the previous evidence review (if available) and cover the time period since the last review (3 months prior to the end-date of the previous search). Searches are limited to the English language, to humans, and, in the case of PubMed searches, to the journals in the abridged Index Medicus (120 "core clinical journals"). For benefits, the search is limited to meta-analyses, systematic reviews, and CCT/RCTs. For harms, the search includes meta-analyses, systematic reviews, RCTs, cohort, and case-control studies. The results of the searches are incorporated into the work assignment as a brief summary of the findings.

At the work plan development stage (after receiving the work assignment), the EPC librarian works with the topic team to develop a strategy for searching the literature to identify existing systematic reviews and other high quality synthesized literature (such as meta-analyses). This search is the first systematic search that will be incorporated into the overall searching done by the EPC for this topic. The purpose is to locate existing synthesized evidence that should be incorporated or built upon in the current systematic review, and the current methods emphasize finding all relevant synthesized evidence.

The following databases and websites are searched: Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database (UK), National Institute for Health and Clinical Excellence - NICE (UK), Institute of Medicine, MEDLINE via PubMed (using the systematic reviews search hedge developed by NLM), and when appropriate, subject-specific databases (e.g. PsycINFO). Timeframes are not limited, except in the case of databases such as MEDLINE and subject-specific databases, which are searched beginning a year prior to the most recent, relevant citation in the previous review.

In order to identify ongoing studies that could affect review scope and/or planning, the EPC librarian and/or topic team search ClinicalTrials.gov, and, if necessary, the National Institutes of

Health CRISP (Computer Retrieval of Information on Scientific Projects) clinical trials data base to identify current trials and/or studies relevant to this review. (<http://crisp.cit.nih.gov>.)

The topic team also checks to determine whether there is a finished, in-process, or planned Community Task Force review (from the CDC's Guide to Community Preventive Services) for the clinical preventive service being reviewed. The timing of this search (work plan stage or later) is left to the discretion of the topic work team.

3.4 Previous USPSTF Review and Recommendations (for an update)

To ensure that the current work plan builds coherently upon the previous USPSTF work on the topic, this section of the work plan succinctly summarizes the conceptual clinical framework and evidence foundation built by the previous reviews and recommendation statement on the topic. The current USPSTF graded recommendations are listed here verbatim, along with the analytic framework (AF), key questions (KQ), summary of evidence table, main findings, and conclusions from the previous review. Methodological or scope limitations and evidence gaps identified in the previous review are listed in bullet form.

3.5 Current USPSTF Review Approach

3.5.1 Inclusion/Exclusion Criteria (Admissible Evidence)

The topic team clearly documents the criteria by which it chooses to admit evidence on a given key question. Such criteria might include, for example, study design (randomized clinical trials [RCTs], cohort studies, and cost-effectiveness analyses [CEAs]), setting, sample size, population studied, language(s) of publication, and year(s) of publication.

No generic criteria for admissible evidence have been established. Rather, the criteria are determined on a topic-by-topic and key-question-by-key-question basis, depending on the questions and the amount and quality of evidence available. All inclusion/exclusion criteria are reviewed and approved by the entire topic team.

One variable in exclusion and inclusion criteria relates to the timeframe of the literature search. For an update of a topic previously covered by the Task Force, the topic team establishes whether the key questions they are posing had been similarly addressed in the previous review. If they were addressed, the team may evaluate key studies previously reviewed, but would not systematically re-review the same literature. An exception to this would be if the Task Force decided to evaluate the validity of this evidence by a method different from that used in the previous review. If a key question has changed, or if the threshold for adequate evidence has changed, the team searches back in time for evidence available before the previous review was published.

In addition, the topic team searches for other systematic reviews on the topic. If another systematic review is found that is rigorous and that addresses the same key question as the topic team, the team may choose to use the data generated in that review rather than redoing the work.

An important aspect of the definition of admissible evidence is the definition of "burden of suffering" of the condition in question. This burden is the ultimate target of implementing the preventive service. Admissible evidence is evidence relevant to the burden of suffering,

including the prevalence of the condition in various populations and the impact of the condition on the health of these populations. The severity of the condition as measured by such metrics as prevalence and severity (eg, number of life-years and quality-adjusted life years lost in a population) is also an important aspect of the burden of suffering. The burden of suffering of a condition defines the maximum possible benefit from prevention of that condition.

3.5.2 Literature Search Strategy

At this point the topic team proposes its full-scale search strategy. At a minimum, all topics include a review of the English-language literature in MEDLINE and the Cochrane Database. Other databases (eg, nursing and allied health or psychology databases) are searched as the topic team deems necessary. To identify sources of synthesized evidence and ongoing studies that could affect review planning, the EPC librarian works with the EPC review team to develop a strategy for searching the literature to identify existing systematic reviews and other high quality synthesized literature (such as meta-analyses). The following databases and websites are searched: Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database (UK), National Institute for Health and Clinical Excellence - NICE (UK), Institute of Medicine, MEDLINE via PubMed (using the systematic reviews search hedge developed by NLM), and when appropriate, subject-specific databases (e.g., PsycINFO). Searches are not limited in terms of timeframes, except in the case of databases such as MEDLINE and subject-specific databases, which are searched beginning a year prior to the most recent, relevant citation in the previous review.

The EPC librarian and/or topic team search ClinicalTrials.gov (and CRISP, when necessary) to identify current trials and/or studies relevant to this review. The review team also checks to determine whether there is a finished, in process, or planned Community Task Force review for the clinical preventive service being reviewed. The timing of this search (work plan stage or later) is left to the discretion of the topic work team.

3.5.3 Use of Topic Experts

By design, EPC review teams consist of generalist clinicians, methodologists, and staff with various levels of content expertise. Where appropriate, the EPC review team engages content experts as consultants to advise about work plan formulation and about operational decisions to be made during the conduct of the review. To allow continuity with the previous USPSTF review, the EPC review team includes previous review team members, where possible, as members of or as consultants to the current review team.

3.6 Peer-review of Work Plan

For full systematic reviews undertaken for new USPSTF topics, the work plan is peer-reviewed (“hardened”) before it is finalized. From 4 to 6 peer reviewers are chosen to provide content expertise, topical research experience, and relevant methodological or policy expertise as appropriate to the topic. Peer reviewer lists are drafted by the EPC and amended and approved by the AHRQ medical officer. The EPC coordinates the peer-review process (by telephone interview or through written communication) and incorporates peer reviewers’ suggestions into the draft work plan. Peer reviewers’ comments are not formally summarized. Instead, peer-reviewed work plans, the list of peer reviewers, and a synopsis of their comments and the resulting revisions are presented at a regular USPSTF meeting for final input and approval by the USPSTF as a whole.

3.7 USPSTF Approval of Final Work Plan

The USPSTF provides final approval of the work plan through its assigned topic work group (for updates) or through a vote at its meetings (for new topics). Once the work plan is approved, the EPC finalizes it and submits it to the AHRQ medical officer as a deliverable.

Section 4

Evidence Report Development

4.1 Literature Retrieval and Data Abstraction for Topic Reviews

After literature searches are conducted, the team of evidence reviewers uses a set of a priori inclusion/exclusion criteria as appropriate to each key question to define whether identified literature is relevant to the review. These criteria are applied twice first at the title, or title and abstract review stage, and a second time at the article review stage. This two-stage process is designed to be efficient, to minimize errors, and to be transparent and reproducible.

Titles and abstracts are reviewed by broadly applying the inclusion criteria for the review. When in doubt at the title/abstract review phase as to whether an article might meet the inclusion criteria, reviewers should err on the side of inclusion so that article is retrieved and can be reviewed at the article stage. All citations are coded with at least an excluded or included code, which is managed in a database and used to guide the further literature review steps. This database is the source of the final tables documenting the review process.

Full-text articles are retrieved for all citations included at the title/abstract stage, and are reviewed by a member of the review team, using inclusion/exclusion criteria for relevance and for quality. Included articles receive codes to indicate the key question(s) for which they meet criteria and excluded articles are coded for a reason for exclusion. The reasons for exclusions could be either the primary reason or the first reason encountered in reviewing the article; and thus the distribution of reasons for exclusion do not necessarily represent the state of the excluded literature. Similarly, all the reasons for exclusion of an individual article may not be listed in the final exclusion table. Before they are abstracted, articles are reviewed to ensure that they meet minimal design-specific USPSTF quality criteria.

The abstract and article review process generally involves a team of reviewers and is conducted using established research methods in order to minimize reviewer drift as well as inter-rater review and coding differences.

4.1.1 Procedures for Abstract and Article Review

Abstracts undergo “dual-review” in that either all abstracts are reviewed separately and reconciled, or at least abstracts excluded by the primary reviewer are re-reviewed by another reviewer to ensure that all appropriate studies are included. Any studies excluded by the first reviewer but included by the second reviewer are included in the next phase.

When the volume of abstracts is very high due to the non-specific nature of searches possible within a specific literature (eg, alcohol misuse), reviewers may use a sampling scheme for quality assurance as follows. For each key question, all of the searches (ML, CCRCT, PsycINFO) will be considered as one search. Reviewers will dual review a set number (1000) of the most recent abstracts that proportionally represents the key databases searched for that key question, and will then review a random subset of the remainder. All abstracts resulting from the CCRCT are dual-reviewed. The other database searches are proportionally reviewed to get up to a total of 1000 abstracts that are dual reviewed, then a random subset of the remainder, to

equal about 20-25% of the total number of articles, will be dual-reviewed. In the case of a sampling approach to dual review, inter-rater reliability is calculated using the kappa statistic.

4.1.2 Database of Abstracts

For each systematic review, the review team establishes a database of all articles located through searches and from other sources (ie, both those included and those eventually excluded from the final set of articles reviewed). Information captured in the database includes the source of the citation (eg, search source, outside source), whether the abstract was included or excluded, the key question(s) associated with each included abstract, whether the article was excluded (with reasons for exclusion) or included in the review, and other coding approaches developed to support the specific review. For example, a hierarchical approach to answering a question may be proposed at the work plan stage, specifying that reviewers will consider a type of study design or a clinical setting only if research data are too sparse for the preferred type of study. While reviewing abstracts and articles, these can be coded to allow easy retrieval during the conduct of the review, if warranted.

4.1.3 Documenting Search Results

Search terms used for each key question, along with the yield associated with each term, are documented in a table or set of tables; these appear in the summary of the literature search (early in any topic review project) and in the final evidence synthesis. Follow-up searches to capture newly published data are conducted periodically as the project progresses; the frequency of these searches depends on the individual topic. A final search is conducted close to the time of completion of the draft evidence report, with the exact timing determined by the evidence review team. The final documentation of the search should indicate the most recent time point searched.

To the extent that it suits the review rationale and is feasible, search dates for different key questions should conform to one another.

4.1.4 Data Abstraction Approaches

- 1) Use of forms: Data may be abstracted in forms developed or adapted for the review, or directly into evidence tables.
- 2) Minimal elements to abstract: Although the Task Force has no standard or generic abstraction form, the following broad categories are always abstracted from included articles: key question, study design, study participant description, details of the intervention or screening test being studied, study results with emphasis on health outcomes where appropriate, and individual study quality information, including specific threats to validity. Information relevant to generalizability is consistently abstracted. Each team uses these general categories, and other categories if indicated, to develop an abstraction form specific to the topic at hand. For example, source of funding may be an important variable to abstract for some topics.
- 3) Abstraction of included articles: The evidence review teams abstract only those articles that, after review of the entire article, both meet criteria for quality and focus on the key question at hand. Abstractions are conducted by trained team members, and evidence review teams may, but do not routinely, double abstract all included articles. Key articles

are always read and checked by more than one team member. All reviewers are trained in the topic, the analytic framework and key questions, and the use of the abstraction instrument. Initial reliability checks are done for quality control.

- 4) Other quality assurance methods: It is desirable to have more than one evidence review team member check data accuracy for key data elements, including data included in a summary table, a meta-analysis, or in calculations supporting a balance sheet/outcomes table.

4.2 Critical Appraisal

By means of its explicit analytic framework and key questions, the Task Force indicates what issues it must examine to make its recommendation. By setting inclusion and exclusion criteria for the searches for each key question, the Task Force indicates what evidence it will consider admissible. The critical aspect used to determine whether an individual study is admissible is its internal and external validity with respect to the key question posed. This initial examination of the “quality” (ie, internal and external validity) of individual studies is conducted with established criteria (see Appendix VII and VIII) by the evidence review team or USPSTF topic work group. If questions arise in the course of this process, Task Force members are asked to review the articles in question. Studies with fatal flaws (ie, with “poor” internal or external validity) are not admissible for further consideration. Likewise, studies of interventions that require training or equipment not feasible in even high quality primary care would be judged to have poor external validity for the key questions posed by the Task Force, and would not be admissible evidence.

Once the admissible evidence has been found, and the internal and external validity of individual studies has been assessed, the Task Force must consider the level of evidence that the studies provide to answer the KQs. The Task Force’s process for determining the level of evidence over a key question involves answering the following 6 critical appraisal questions about the admissible evidence. The Task Force uses these same 6 critical appraisal questions to determine the overall evidence of certainty of net benefit for the entire preventive service, including all key questions in the analytic framework. (See Section 5 for a description of the Task Force’s methods for judging the cumulative evidence and arriving at a recommendation.)

4.2.1 Critical Appraisal Questions

- 1) Do the studies have the appropriate research design to answer the key question(s)?
- 2) To what extent are the existing studies of high quality? (ie, what is the internal validity?)
- 3) To what extent are the results of the studies generalizable to the general US primary care population and situation? (ie, what is the external validity?)
- 4) How many studies have been conducted that address the key question(s)? How large are the studies? (ie, what is the precision of the evidence?)
- 5) How consistent are the results of the studies?
- 6) Are there additional factors that assist us in drawing conclusions (eg, presence

or absence of dose-response effects; fit within a biologic model)?

4.2.2 Levels of Critical Appraisal

The evidence review process involves assessing the validity and reliability of admissible evidence at 3 levels:

- 1) The individual study;
- 2) The key question (ie, linkage in the analytic framework); and
- 3) The entire preventive service.

For individual studies, questions 1-3 and 6 are assessed. That is, a single study will be categorized as to study design and whether internal and external validity are “good,” “fair,” or “poor” to answer the key question. For the key question and entire preventive service levels, all 6 questions must be considered.

For the individual study level, the evidence review team finds admissible evidence and then categorizes the internal validity (ie, quality – Appendix VII) of each study into “good,” “fair”, and “poor” categories. For critical or borderline studies, the Task Force leads (and sometimes the entire Task Force) will also consider the individual studies. The EPC also provides the Task Force with descriptions of factors entering into the determination of external validity (ie, applicability or generalizability – Appendix VIII), as well as descriptions of each study’s research design and the number and description of studies relevant to each key question.

For the key question level, the Task Force, using information about the evidence supplied by the EPC, assesses the level of evidence across each key question using all 6 critical appraisal questions. The body of evidence is often categorized as to the highest level of applicable evidence available. The Task Force categorizes the evidence across each key question into one of 3 categories: “convincing,” “adequate,” or “inadequate.”

For the preventive service, the entire body of evidence in the entire analytic framework is synthesized by the Task Force into categories of “certainty” of the overall evidence: high, moderate, and low. Again, the Task Force uses all 6 critical appraisal questions for this determination. (See Appendix IV regarding topic workgroup procedures for assessing certainty.)

4.3 Assessing Evidence at the Individual Study Level

4.3.1 Critical Appraisal

All individual articles are critically appraised to determine the validity and reliability of the evidence they provide. This assessment is conducted primarily by the topic team (usually led by the EPC or by AHRQ team leaders), with input from Task Force members for critically important or borderline articles. The assessment of internal (ie, “quality”) and external validity (ie, applicability or generalizability) are based on explicit criteria, given in Appendix VII and Appendix VIII.

4.3.2 Internal Validity

The Task Force recognizes that research design is an important component of the validity of the information in a study, for the purpose of answering a key question. Although RCTs cannot answer all key questions, they are ideal for questions of the benefits or harms of various interventions. Thus, for these questions, the current Task Force endorses a slightly revised version of the “hierarchy of research design” used by the second Task Force:

- I: Properly powered and conducted randomized controlled trial (RCT); well-conducted systematic review or meta-analysis of homogeneous RCTs
- II-1: Well-designed controlled trial without randomization
- II-2: Well-designed cohort or case-control analytic study
- II-3: Multiple time series with or without the intervention; dramatic results from uncontrolled experiments
- III: Opinions of respected authorities, based on clinical experience; descriptive studies or case reports; reports of expert committees

In assessing individual studies, all are classified first according to this design code, with additional designations added for other or unconventional designs.

Although research design is an important component of the information provided by an individual study, the Task Force also recognizes that not all studies within a research design have equal internal validity (“quality”). To assess more carefully the internal validity of individual studies within research designs, the Task Force adopted design-specific criteria for assessing the internal validity of individual studies.

These criteria, given in Appendix VII, provide general guidelines for categorizing studies into one of three internal validity categories: “good,” “fair,” and “poor.” These specifications are not meant to be rigid rules; individual exceptions, when explicitly explained and justified, can be made. In general, a “good” study is one that meets all design-specific criteria. A “fair” study is one that does not meet (or does not clearly meet) at least one specified criterion, but has no known “fatal flaw.” “Poor” studies have at least one fatal flaw. A fatal flaw is a deficit in design or implementation of the study that calls into serious question the validity of its results for the key question being addressed.

The Task Force views the level of evidence, whether for an individual study, a key question/linkage, or an entire preventive service, as independent of the magnitude of effect. Thus, a study (or a number of studies) could be classified as “good” even if it (they) found no effect of the preventive service.

4.3.3 External Validity (Generalizability) and Applicability

It is necessary not only to assess the *external validity (generalizability)* of the individual studies that contribute to answering the key questions, but also to assess the body of evidence in order to judge its *applicability* to the population or populations that are the target for the clinical preventive service, to the settings in which the service will be implemented, and to the providers who will deliver the service. In this document, the term "*external validity*" will be used when discussing assessment of individual studies, and the term "*applicability*" will be used when discussing the assessment linkages across key questions and the overall body of evidence, even though the external validity of individual studies is a key element of the applicability judgment. The summative judgment about applicability is more than the sum of the assessment of each of the parts.

For the USPSTF, the study-level assessment of external validity and the assessment of applicability are done separately.

A description of the overall conceptual approach for both components is provided below. Appendix VIII gives detailed information on criteria and process.

4.3.3.1 Assessment of the External Validity of a Study

Judgments about the external validity ("generalizability") of a study pertinent to a preventive interventions address three main questions:

- 1) Considering the subjects in the study, to what degree do the study's results measure the likely clinical results in the asymptomatic people who are the recipients of the preventive service in the United States?
- 2) Considering the setting in which the study was done, to what degree do the study's results measure the likely clinical result in United States primary care situation? and
- 3) Considering the providers who were a part of the study, to what degree do the study's results measure the likely clinical results in providers who would deliver the service in the United States primary care setting?

4.3.3.2 Populations

The subjects that comprise the participants in a study may differ from people receiving primary care in many ways. Such differences may include gender, ethnicity, age, co-morbidities, and other personal characteristics. Some of these differences have a small potential to affect the study's results and/or the outcomes of an intervention. Other differences have the potential to cause large differences between the study's results and what would be reasonably anticipated to occur in asymptomatic individuals or people who are the target of the preventive intervention.

The choice of the study population may affect the magnitude of the benefit observed in the study through exclusion/inclusion criteria that limit the study to people most likely to benefit; other study features may impact the risk level of the subjects recruited to the study. The absolute benefit from a service is often greater for people at increased risk than for people at lower risk.

Because of the presence of certain research design elements, adherence is likely to be greater in research studies than in the usual primary care practice. This may lead to overestimation of the benefit of the intervention when delivered to people who are less selected (ie, who more closely resemble the general population), and who are not subject to the special study procedures.

4.3.3.3 Situations

Factors related to the study situation relative to the situation in U. S. primary care settings must be assessed when assessing the external validity of a study.

The choice of study setting may lead to an over- or under-estimate of the benefits and harms of the intervention as they would be expected to occur in U.S. primary care settings. For example, results of a study in which items essential for the service to have benefit are provided at no cost to patients may not be attainable when the item must be paid for. Results obtained in a trial situation that ensures immediate access to care if a problem or complication occurs may not be obtainable in a usual care situation, where the same safeguards cannot be ensured, and where as a result the risks of the intervention are greater.

4.3.3.4 Providers

When assessing the external validity of a study, factors related to the experience of providers in the study should be considered in comparison with the experience of providers likely to be encountered in primary care in the U.S. Studies may involve providers selected for their experience or their high skill level. Providers involved in studies may undergo special training that affects their performance of the intervention. For these and other reasons, the effect of the intervention may be overestimated or the harms underestimated compared with the likely experience of unselected providers in the primary care setting.

4.3.3.5 Criteria and Process

The criteria used to rate the external validity of individual studies according to the population, the situation, and the setting are described in detail in Appendix VIII. As with internal validity, this assessment of external validity is usually conducted initially by the EPC or AHRQ topic team leader, with input from Task Force members for critically important or borderline studies. This assessment is then used to give each study a rating using the same 3-tiered grading scheme as for internal validity: good, fair, and poor.

The underlying question answered in the grading the external validity of a study as good, fair or poor is:

If the study had been done with the typical U.S. primary care population, situation, and providers, what is the likelihood that the results would be different in a clinically important way?

4.4 Applicability of the Body of Evidence to the Target Population/Situation/Setting

USPSTF members assess the applicability of the body of evidence to populations/situations/settings as one of the components of the overall process of making recommendations.

Judgment about applicability considers the populations, situations, and providers in each study, but it also involves synthesis of the evidence from the individual studies across the key questions, and for the overall body of evidence.

The overall goal of the assessment is to judge whether there are likely to be clinically important differences between the results observed in studies as a whole and the results expected when the intervention is implemented in the U.S. primary care populations/situations/providers.

The following questions are addressed:

1. Can an inference be made from the evidence that the intervention has *any* effectiveness for the U.S. primary care populations/situations/providers?
2. Is the *magnitude of benefit* observed in individual studies that comprise the body of evidence likely to be the same for the U.S. primary care populations/situations/providers?
3. Are the *harms* observed in individual studies that comprise the body of evidence likely to be the same for the U.S. primary care populations/situations/providers?
4. What is the *relationship between benefits and harms* derived from the evidence likely to be for the U.S/ primary care population/situation/providers?
5. Is the time and effort required to provide the interventions that comprise the body of evidence attainable in the U.S. primary care situations/providers?
6. Can people in U.S. primary care populations/situations be expected reasonably to partake of the interventions that comprise the body of evidence considering their time, effort, and cost?
7. Is the extrapolation of data from the body of evidence to large populations of asymptomatic people biologically plausible?

4.4.1 Relative Importance of Efficacy/Effectiveness.

The USPSTF seeks to make recommendations based on projections of what would be expected from widespread implementation of the preventive service within the actual world of U.S. medical practice. For this reason, the Task Force considers carefully the applicability to medical practice of “efficacy” studies, which measure the effects of the preventive care service under ideal circumstances. However, the USPSTF ultimately seeks to base its recommendations on “effectiveness,” which is what results could be expected with widespread implementation under usual practice circumstances.

Questions arise about whether the USPSTF recommendations consider effectiveness in usual practice or in ideal/excellent practice. The “situation” for practices varies widely within the U.S. Some practices have greater support and more resources than others. The TF attempts to make recommendations for all of these practice “situations,” and may specify what resources are required for implementation.

4.4.2 Definition of Primary Care

To further specify the situation that is the object of its concern, the Task Force has adopted the Institute of Medicine’s definition of primary care:

Primary care is the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs,

developing a sustained partnership with patients, and practicing in the context of family and community. This definition acknowledges the importance of the patient clinician relationship as facilitated and augmented by teams and integrated delivery systems. (7)

4.4.3 Primary Care Interventions Addressed by the USPSTF

The USPSTF considers interventions that are delivered in primary care settings or are judged to be feasible for delivery in primary care. To be feasible in primary care, the intervention could target patients seeking care in primary care settings, and the skills to deliver the intervention are or could be present in clinicians and/or related staff in the primary care setting, or the intervention could generally be ordered/initiated by a primary care clinician.

4.5 Other Issues in Assessing Evidence at Individual Study Level

4.5.1 Dealing with Secondary and Aggregate Endpoints

The Task Force adopted a policy of critically appraising all of the endpoints (outcomes) of trials in a similar manner, following the 6 critical appraisal questions listed earlier (section 4.2.1). In its review, the Task Force takes note of the biological plausibility of a study's finding, the supporting evidence, and whether an outcome is a primary or secondary one. Similarly, the Task Force examines composite (aggregate) outcomes carefully. It generally asks 3 questions of these outcomes: (1) are the component outcomes of similar importance to patients? (2) did the more or less important outcomes occur with similar frequency? And (3) are the component outcomes likely to have similar relative risk reduction (RRR)?

4.5.2 Ecologic Evidence

Because biases may be present in ecologic data, the Task Force is careful in its use of this type of evidence. The Task Force rarely accepts ecologic evidence alone as sufficient to recommend a preventive service. Because this evidence is widely accepted by others, the Task Force developed a policy for when it uses ecologic evidence, and how this evidence is critically appraised.

By ecologic evidence we mean data that are not at the individual level; but rather, that relate to the average exposure and average outcome within a population. The "ecologic fallacy" is the erroneous conclusion that there is an association when exposure occurs in some members of a population and an outcome in other members. In addition, ecologic data sets often do not include other potential confounding factors; thus, one cannot directly assess the ability of these potential confounders to explain apparent associations. Finally, some ecologic studies use data collected in ways that are not accurate or reliable.

Ecologic studies usually make comparisons of outcomes in exposed and unexposed populations in one of two ways: (1) between different populations, some exposed and some not, at one point in time (ie, cross-sectional ecologic study); or (2) within a single population with changing exposure status over time (ie, time series ecologic study). In either case the potential for making the ecologic fallacy is a major concern.

As it is not possible to completely avoid the potential for making the ecologic fallacy in these studies, the USPSTF does not usually accept ecologic evidence alone as adequate to establish the causal association of a preventive service and a health outcome. In some unusual situations (eg, cervical cancer screening) ecologic evidence may play the primary role in the Task Force's evidence review, but this is rare.

More frequently, ecologic evidence is considered by the Task Force in the following situations:

- 1) For background, for an understanding of the context in which the preventive service is being considered;
- 2) When well-known ecologic data are being used as evidence by others to justify either recommending or not recommending the service the Task Force is considering;
- 3) Where other evidence is inadequate but the Task Force thinks that good ecologic evidence could add important information;
- 4) When there are reports of dramatic results of ecologic studies.

In the situations above, the Task Force critically appraises ecologic studies. High quality ecologic evidence meets the following criteria:

- 1) The exposures, outcomes, and potential confounders are measured accurately and reliably.
- 2) Other potential explanations and potential confounders are considered and adjusted for.
- 3) The populations and interventions being compared are comparable.
- 4) The populations and interventions are relevant to a primary care population.
- 5) Multiple ecologic studies are present that are consistent/coherent.

4.5.3 Mortality as Outcome: All-cause Versus Disease-specific Mortality

When a condition is a common cause of mortality, all-cause mortality, instead of cause-specific mortality, is a desirable health outcome measure. Few preventive interventions attain the high standard set by use of this outcome. The fact that there is a discrepancy between the effect of the preventive intervention on all-cause and disease-specific mortality is important to recognize and explore. A discrepancy may arise when (1) there is real benefit of the preventive intervention for a targeted condition or (2) because of methodologic issues that are inherent in the study of all-cause mortality:

4.5.3.1 Real Benefit for the Targeted Condition

Three situations can produce this kind of discrepancy. First, when a preventive intervention increases deaths from causes other than the one targeted by the intervention, all-cause mortality may not be decreased even when cause-specific mortality due to the targeted condition is decreased. This indicates a potential harm of the intervention for a condition other than the one targeted.

Second, when the condition targeted by the preventive intervention is rare and/or the effect of the intervention on cause-specific mortality due to the targeted condition is small, the effect on all-cause mortality may be very small or even non-existent.

Third, when the preventive intervention is applied in a population with strong competing causes of mortality, the effect of the preventive intervention considering all-cause mortality may be very small or even non-existent even though the intervention decreases cause-specific mortality due to the targeted condition. For example, preventing death due to hip fracture by implementing an intervention to decrease falls in 85-year women may not decrease all-cause mortality over reasonable time frames for a study because the force of mortality is so large at this age.

4.5.3.2 Methodologic Issues

Methodologic issues can arise because of difficulties in the assignment of cause of death based on records available to or used by a study. In the absence of detail about the circumstances of death, death may be attributed to a chronic condition known to exist at the time of death but without any true contribution to death. Coding conventions for death certificates also result in deaths from some causes being attributed to chronic conditions present at death routinely. For example, it is conventional to assign people with a mention of cancer on the death certificate to cancer as primary cause of death. The result of these methodologic issues is a biased estimate of cause-specific mortality, which may not reflect the true effect an intervention has on death from the targeted condition.

As indicated above, studies that provide data on all-cause and cause-specific mortality may have low statistical power to detect even large or moderate effects of the preventive intervention on all-cause mortality. This is especially true when the disease targeted by the screening test is not common.

When data are available, the Task Force considers data on both all-cause and cause-specific mortality in making its recommendations, taking into account the real and methodologic contributions to potential discrepancies between apparent and true effect.

4.5.4 Subgroup Analyses

The Task Force is interested in targeting its recommendations to those populations or situations in which there would be maximal benefit for the harms and costs involved. Thus, it often takes into consideration subgroup analyses of large studies. It attaches varying levels of credibility to those analyses, however, depending on such factors as: the size of the subgroup; whether randomization occurred within subgroups; whether a statistical test for interaction was done; whether the results of multiple subgroup analyses are consistent within themselves; whether the subgroup analyses were pre-specified; and whether the results are biologically plausible.

4.5.5 Relative Versus Absolute Risk Reduction

The Task Force is interested in reducing risk both for populations and for individuals. For this reason it takes into account both relative (RRR) and absolute risk reduction (ARR) from intervention studies. It generally prioritizes ARR over RRR. That is, it is less impressed with a large RRR in situations of low ARR; it remains interested in an intervention with a low RRR if its ARR is high.

4.6 Incorporating Other Systematic Reviews in USPSTF Reviews

Existing systematic reviews or meta-analyses that meet quality and relevance criteria can be incorporated into topic reviews done for the USPSTF. Quality criteria for reporting meta-analyses are specified by the QUORUM and MOOSE guidelines published, respectively, by The Lancet and the Journal of the American Medical Association (JAMA) (8, 9). The USPSTF has specified its criteria for critically appraising systematic reviews (see Appendix VII and VIII). Relevance is considered at two levels: at the general level of the review or analysis question, and at a more specific level. At the general level, the question would be “Is the review or meta-analysis relevant to one or more of the USPSTF key questions for this review?” The more specific question would be: “Did the review include the desired study designs and relevant population(s), settings, exposure/intervention(s), comparator(s), and outcome(s)?” Recency of the review is also a consideration, and can determine whether a review that meets quality and relevance criteria is recent enough not to require any bridging searches. Finally, existing reviews can be used in several ways in a USPSTF review: (1) to answer one or more key questions wholly or in part; (2) to substitute for conducting a systematic search for a specific time period for a specific key question; or (3) as a source document for cross-checking the results of systematic searches.

4.7 Use of Observational Designs in Questions of the Effectiveness/Efficacy of Interventions

The Task Force prefers large, well-conducted RCTs to determine the benefits and harms of preventive services. In many situations, however, such studies have not been or are not likely to be done. When these studies can be done, and other evidence is insufficient to determine benefits and/or harms, the Task Force advocates that large, well-conducted RCTs be done. It notes that small, poorly-conducted RCTs are often of little use.

In some situations, however, the Task Force does use observational evidence to make recommendations. Multiple, large, well-conducted observational studies with consistent results showing a large effect size that does not change markedly with adjustment for multiple potential confounders may be judged sufficient to determine the magnitude of benefits and harms of a preventive service. Also, large, well-conducted observational studies often provide essential additional evidence even in situations where there are adequate RCTs. Ideally, RCTs provide evidence that an intervention can work and observational studies provide better understanding of the populations where the benefits would be greatest.

Section 5

Methods for Arriving at a Recommendation

The preceding two sections have dealt with the processes of question definition and evidence review, processes which are primarily the work of the EPC. This section begins the description of the specific work of the Task Force in examining and judging the cumulative evidence presented to it, and making recommendations. The steps in this process, as described in this section, include assessing the evidence at the Key Question level and across an entire Analytic Framework, assessing both the certainty of the evidence about, and the magnitude of, the harms and benefits of the service, estimating the magnitude of the net benefit for the service, and the certainty of that estimation, and finally arriving at a recommendation grade for that service in the relevant population.

5.1 Assessing Evidence at the Key Question/Linkage Level

In considering the information provided by the body of evidence across a linkage in the analytic framework (eg, for a key question), all 6 critical appraisal questions (see above) must be considered. The evidence concerning a key question is often categorized by its strongest research design, and then the level of evidence is classified into one of 3 categories: “convincing”, “adequate”, and “inadequate.”

In making this determination, the Task Force considers the evidence described by the EPC in its review. It considers the “aggregate internal validity” of all studies across the key question. This judgment is not a simple summation of the grading of all of the studies in a body of evidence, but often reflects the best research concerning an issue. The general issue is the extent to which at least some studies meet the criteria for internal validity (Appendix VII). Likewise, aggregate external validity refers to the extent to which the best studies are generalizable to primary care populations, situations, and providers (Appendix VIII).

Coherence is used (in addition to consistency) to indicate that a body of evidence “makes sense,” in that it fits together to present an understandable picture of the situation. Coherence in this context includes the concordance between populations, interventions, and outcomes in the studies reviewed. Several studies of an issue may find different results (and thus be inconsistent), but the results may still be understandable (and thus coherent) in terms of the populations they studied or the interventions they used.

5.2 Assessing Certainty of Evidence for the Entire Preventive Service: Evidence Synthesis

As in assessing the evidence across key questions, the Task Force, in discussion with the EPC/topic team leader, also plays the primary role in synthesizing evidence for the entire preventive service.

Assessing evidence at the level of the entire preventive service requires a complex synthesis of all evidence across the entire analytic framework. The question is not simply the level of evidence across each key question/Linkage, but also how these bodies of evidence fit together to provide an accurate estimate of the expected magnitude of benefits, harms, and net benefit

(ie, benefits minus harms) that would be realized from widespread implementation of the preventive service.

The Task Force considers this synthesis of the information provided by the entire body of evidence to be the “certainty” of the overall evidence. The certainty may also be thought of as the width of the “conceptual confidence interval” (CCI) given by the evidence to estimate the magnitude of benefits, harms, and net benefits. This CCI is not a quantitative calculation, but rather a judgment based on the 6 critical appraisal questions given earlier, and how the evidence fits together to complete the linkage from the left side of the analytic framework (population) to the right side (health outcomes). A wide CCI can come from a lack of evidence about one or more key questions; from studies of the wrong study design; from studies of the right design but of poor internal or external validity; from too small and/or too few studies; from inconsistent/incoherent studies; or other aspects of the studies that cloud the interpretation of the magnitude of benefits, harms, and net benefits. When the CCI is wide, then the magnitude cannot be estimated with any confidence, and the entire body of evidence is categorized as having “low certainty.”

When the evidence satisfies most criteria in all of the 6 critical appraisal criteria, and fits together “well enough” to make the connections across the analytic framework, then the CCI is considered to be narrower. In this case, we have a better (although not precise) estimate of the magnitude of benefits, harms, and net benefits. This type of body of evidence is categorized as “moderate certainty.”

When the evidence satisfies criteria for each of the 6 critical appraisal criteria across the analytic framework, and the evidence fits together well, then the CCI is narrow—we have a more precise estimate of benefits, harms, and net benefits. In this case, the body of evidence is categorized as “high certainty.”

The general definitions of the 3 levels of overall evidence are given in Table 2.

The Task Force is careful to separate the concepts of “certainty” of evidence and “magnitude of benefits, harms, and net benefit”. For example, the Task Force may have high certainty of the overall evidence and still determine that there is small (or even zero) magnitude of benefits. Or it may have moderate certainty and determine that there is substantial magnitude of net benefits. The TF first assesses the certainty of the evidence, then the magnitude of benefits, harms, and net benefit. These are used together in the Recommendation Grid to determine the TF recommendation letter.

Table 2. USPSTF Levels of Certainty Regarding Net Benefit

Definition: The U.S. Preventive Services Task Force defines certainty as “likelihood that the USPSTF assessment of the net benefit of a preventive service is correct.” The net benefit is defined as benefit minus harm of the preventive service as implemented in a general, primary care population. The USPSTF assigns a certainty level based on the nature of the overall evidence available to assess the net benefit of a preventive service.

Level of Certainty	Description
High	<p>The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.</p>
Moderate	<p>The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by factors such as:</p> <ul style="list-style-type: none"> - the number, size, or quality of individual studies; - inconsistency of findings across individual studies; - limited generalizability of findings to routine primary care practice; or - lack of coherence in the chain of evidence. <p>As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.</p>
Low	<p>The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of:</p> <ul style="list-style-type: none"> - the limited number or size of studies; - important flaws in study design or methods; - inconsistency of findings across individual studies - gaps in the chain of evidence; - findings not generalizable to routine primary care practice; or - a lack of information on important health outcomes. <p>More information may allow an estimation of effects on health outcomes.</p>

5.3 Dealing with Conflicts among RCTs

When RCTs of the same question appear to contradict each other, the Task Force first clinically appraises the studies, considering whether the studies truly contradict each other. Evidence on the clinical issue from other sources may be useful in this assessment. The Task Force then critically appraises the studies to determine whether the quality of the trials helps to explain any differences. If neither clinical nor epidemiologic reasoning shed light on the differences between the trials, the Task Force at times must admit that it doesn't know why the studies contradict each other. Quantitative synthesis of trials is most useful for aggregating the results of RCTs when the results are generally consistent.

5.4 Assessing Magnitude of Net Benefit

As noted earlier, the Task Force decided that it is important to keep separate the certainty afforded by the evidence from the magnitude of effect (ie, benefits and harms) of the preventive service. To specify the magnitude of the effect of a preventive service, the Task Force separately assesses the magnitude of benefits and harms, and then combines these into a "net benefit" assessment. The Task Force has adopted a four-tiered grading system for the net benefit rating: substantial, moderate, small, and zero/negative. Thus, "substantial" net benefit indicates that benefits substantially outweigh harms, whereas "zero/negative" net benefit indicates that harms equal or even outweigh benefits. This assessment is conducted by the Task Force, in discussion with the EPC and AHRQ team members.

The Task Force defines net benefit as the magnitude of the benefits of the service minus the magnitude of the harms. The Task Force gives equal attention to both benefits and harms since it is well aware that preventive interventions may result in harms as either a direct consequence of the service or for other "downstream" reasons.

Because of lack of evidence, especially evidence using a single, suitable metric, the assessment of "net benefits" is inherently subjective. Thus, the Task Force has not developed specific criteria to judge net benefit.

The Task Force attempts to quantify the magnitude of benefits and harms that would result from implementing the preventive service in the general primary care population. One way of doing so is by using such metrics as "number needed to treat" (NNT, the number of people that would need to be treated for some defined period of time to prevent one adverse health event) or "number needed to screen" (NNS, the number of people that would need to be screened for some defined period of time to prevent one adverse health event). One can also derive a similar "number needed to harm" (NNH, the number of people needed to treat or screen for a defined time to cause one adverse health event). The Task Force does not have a single NNT, NNS, or NNH that it considers to be a threshold for drawing a conclusion about the magnitude of net benefit, due to the often substantial uncertainty in the evidence used to make the estimates.

The Task Force does have a general way of thinking about the concept of net benefit. Net benefit, as used by the Task Force, is substantial in those situations in which either:

1. A large proportion of the total burden of suffering from the target condition (minus the additional burden caused by the preventive service) that would be relieved from society

by implementing the preventive service, even if the target condition is rare, is large (eg, screening for PKU); or

2. A large amount of the burden of suffering would be relieved from society (minus the amount of the additional burden caused by the preventive service) by implementing the preventive service (eg, counseling for smoking cessation).

Note that in both of these situations, a population can be defined that has a substantial burden of suffering from the target condition, even if rare, and there is a prevention strategy that reduces that burden by a substantial amount. Net benefit, however, would only be substantial if harms of the intervention are zero or small (as in the examples cited here). Thus, both the magnitude of harms and the magnitude of benefits are critical factors in determining net benefits.

5.4.1 Assessing Magnitude of Benefits

In situations where the certainty of evidence is high or moderate, the Task Force considers all of the admissible evidence to determine the magnitude of benefit that would be expected from implementing the preventive service in a defined population. Its preferred approach for doing this is the Outcomes Table. In this table, the topic team uses the evidence to estimate the number of people in a hypothetical population who would benefit in specific ways from implementation of the preventive service, over a given time horizon (often 5-10 years). Specific health benefits might include such things as lives extended, cardiovascular events avoided, visual impairment avoided, lung cancers avoided, or alcohol complications avoided. In some situations, the table can be completed easily by simply transferring information from a large, well-conducted RCT of a representative population. Most commonly, however, some cells in the Table are not so easily completed and require calculations based on assumptions—a situation that intrinsically adds uncertainty. Thus the different numbers in an Outcomes Table have different levels of certainty and must be interpreted carefully.

Note that the numbers in an Outcomes Table are meant to shed light on the amount of the burden of suffering from the condition (within a stated population) that can be expected to be prevented by the intervention in question. The magnitude of benefit cannot be greater than the total burden of suffering.

For screening interventions, the benefit may be further limited by such issues as the following:

- 1) the prevalence of the target condition;
- 2) for heterogeneous conditions, the prevalence of that subtype of the condition that would cause important health problems;
- 3) the sensitivity of the screening test (ie, the degree to which the test will detect that subtype of the condition that would potentially cause health problems; rarely 100%);
- 4) the effectiveness of early treatment (compared with later treatment) of the subtype of the condition that would cause health problems. (This quantity is rarely 100%).

The Outcomes Table can show such considerations as these, demonstrating how many people are likely to receive benefit—and in what ways—from implementation of the preventive service.

In situations of limited or absent direct evidence, this type of logic is useful to the Task Force in placing an upper bound on the magnitude of benefit. In other situations, the Task Force may logically be able to judge the lower bounds of the benefit.

5.4.2 Assessing Magnitude of Harms

The Task Force starts with the assumption that harmless interventions are rare. For screening interventions, the Task Force looks for harms of screening and also harms of early treatment. Harms of screening may include such things as psychological harm from labeling and the harms of work-ups to confirm the presence of the condition. The harms of treatment may include the actual physical effects of early treatment as well as the effects of “over-treatment.” These harms of treatment may accrue to patients whose conditions might never have come to clinical attention or for whom the harms of treatment initiated prior to routine clinical detection were different or occurred earlier and/or over a longer period of time. In other words, these are harms of treatment which would not have occurred in the absence of screening. Although harms of counseling are frequently small, harms may include psychological harms from labeling or harms of treatment.

Although there is often less evidence about potential harms than about potential benefits, the Task Force may draw general conclusions from such evidence as the expected yield of screening in terms of false positive test results. If the prevalence of the condition is low and the specificity of the test is less than 100%, then there will be some false positive tests (ie, the positive predictive value may be low). If the work-up is invasive, then the Task Force can infer that there will be at least some harms from many people going through an invasive work-up for no possible benefit (ie, people who had a false positive screening test).

Similarly, if over-treatment is common, and if the treatment has some adverse effects, the Task Force may infer that screening causes at least some harms, even in the absence of a study dedicated to defining harms. This approach does not require an exact estimate of the magnitude of harms but rather a determination that the harms are unlikely to be less than what is known about the number of false positives, the invasiveness of the work-up, and the expected amount of over-treatment. These “lower bounds” of harms can be shown in an Outcomes Table. Care should be taken to call attention to the estimate’s lack of precision.

In another situation, the Task Force may determine that a study gives an upper bound of benefit (or harm), rather than a lower bound. For example, the Task Force might consider the estimate of benefit to be an upper bound if it came from a study of an intervention conducted by highly trained physicians using specialized equipment for people at very high risk.

The Task Force also considers the time and effort required by both patients and the health care system (opportunity costs) to implement the preventive care service. If the time and effort are judged to be clinically important these factors are also considered in the “harms” category. The Task Force usually has general rather than precise estimates of opportunity costs.

Although opportunity costs may be considered in the Task Force’s letter grades, financial costs are not. The Task Force understands, however, that many of its audiences are interested in issues of financial cost. In situations where there is likely to be some degree of health benefit, the Task Force searches for information about costs and cost-effectiveness and provides a summary of this information under “Other Considerations” in its recommendation statement.

5.4.3 Assessing Magnitude of Net Benefit

Once the Task Force has estimated the magnitude of benefits and harms, it faces the further challenge of synthesizing these assessments into an estimate of the magnitude of net benefit. Weighing the balance of benefits and harms can be challenging since they are often measured in different metrics. Benefits are often quantified in terms of lives extended or illness events averted. Harms may be measured in different metrics, such as false positive screening tests or adverse effects of treatment.

As noted above, the Outcomes Table is a critical tool in the Task Force's approach to determining the magnitude of net benefit. The estimates in an outcomes table may not have a great deal of precision but are useful in giving a general idea of the magnitude of the benefits and harms. Both estimates from direct evidence and also estimates based on explicit assumptions should be included, in order to provide likely upper and lower bounds of the magnitude of specific benefits and harms. A Decision Analysis is another approach to provide information about magnitude of benefits and harms based on best estimates from direct evidence and from explicit assumptions. A Decision Analytic model would typically describe benefits to a population over a life time horizon rather than the five or ten years represented in an Outcomes Table.

It is common for direct evidence to be inadequate to complete one or more critical cells in the Outcomes Table. This may be due either to a lack of direct evidence or to gaps in the direct evidence that is available. Common gaps in direct evidence include such factors as lack of evidence about all populations (including risk groups) of interest; lack of availability of the exact interventions (or the experts administering them) used in the large RCTs; or insufficient follow-up to determine long-term effects of interventions. Thus, the Task Force needs to use indirect evidence to calculate upper or lower bounds of benefits and harms. The conceptual confidence interval (CCI), discussed above (section 5.2), places upper and lower limits on the estimated net benefit. This range is bounded by the best-case and worst-case scenario estimates based on available evidence. The interval is not meant to have a statistical interpretation. The Task Force, however, recognizes the danger in this approach, and considers such bounds with appropriate skepticism, using them only with great care.

After data on assembled expected outcomes, whether from an Outcomes Table, or from a Decision Analytic Model, are presented, the Task Force must still weigh benefits and harms (usually very different types of health effects in different metrics) to arrive at net benefits. Clearly, value judgments are involved in this balancing of effects. In making its determination of net benefit, the Task Force strives to consider what it believes are the general values of most people. When the Task Force perceives that preferences among individuals vary greatly, and that these variations are sufficient to change the balance of benefits and harms, it will often suggest shared decision making to incorporate the individual's perspective into the decision.

The Task Force has standardized the Outcomes Table to the extent possible. There will invariably be some variation, depending on the topic. The standard Outcomes Table format is given in Appendix IX.

5.5 Translating Evidence into USPSTF Recommendations

Clarity and comprehensibility are critical for recommendations' widespread use. The Task Force and AHRQ are aware that the recommendations and the letter grades used to define them may be misunderstood, and are therefore taking pains to clarify and refine them. AHRQ has conducted multiple focus groups of clinicians over the period of years from 2004 to 2006) to solicit feedback about the readability and usability of the Task Force recommendations. Themes that emerged included requests for: simplified, succinct recommendations and an easier-to-use format (bold face type, bulleted sections and boxes to highlight key information); recommendations of other professional organizations to easily compare to the USPSTF recommendation; and websites and references for additional information on the topic.

5.6 Principles for Making Recommendations

Task Force recommendations are coded to reflect both the certainty of the evidence and the magnitude of effect (ie, net benefit as discussed in section 5.4.3 above). To be as explicit as possible about its approach to making recommendations, the Task Force developed a set of principles for making recommendations. These principles, listed below, describe in detail the factors that the Task Force does and does not take into consideration in making recommendations, and to whom the recommendations apply.

- 1) Recommendations are evidence-based: there must be scientific evidence that persons who receive the preventive service experience better health outcomes than those who do not, and that the benefits are large enough to outweigh the harms.
 - The supporting evidence can be compiled from data regarding specific linkages in the analytic framework, but in the end the complete causal chain from intervention to outcome must be supported by acceptable evidence.
 - Inferences about supporting evidence can include generalizations from one population to another when there are acceptable grounds to assume the evidence is applicable to both. A screening test can also be considered effective if evidence supports the value of treatment for early stage disease.
 - Recommendations are not based largely on opinion, such as expert opinion or subjective perceptions based on clinical experience. Subjective judgments do enter into the evaluation of evidence and the weighing of benefits and harms.
 - The scientific rationale for the recommendations and the methods used to review and judge the evidence are stated explicitly along with the recommendations.
 - Recommendations describe services that should or should not routinely be offered based on scientific evidence, although it is recognized that in clinical practice and public policy concerns other than scientific evidence (eg, feasibility, public expectations) may take precedence.
- 2) The outcomes that matter most in weighing the evidence and making recommendations are health benefits and harms.

- In assessing health benefits, outcomes that patients can feel or care about (eg, visual acuity, pain, survival) receive more weight than intermediate/surrogate outcomes.
 - In judging the magnitude of benefit, absolute reductions in risk matter more than relative risk reductions.
 - Effectiveness is considered as valuable, if not more valuable, than efficacy. The ability of patients, providers, and the health care system to perform or maintain interventions over time is considered. Interventions may not be recommended at the population level because of concerns about compliance (adherence), but may be advocated for patients and providers who are willing and able to perform the intervention.
 - The direct and indirect harms of preventive services must also be considered, ensuring that they do not outweigh the benefits to the individual and/or population. Because of the ethical imperative to do no harm, especially when caring for asymptomatic persons, in selected circumstances the quality of evidence for harms need not be as strong as that for benefits. Both physical and psychological harms are considered.
 - Judgments about tradeoffs between benefits and harms are generally made at the population level, and involve subjective estimations by the Task Force of the average utilities of the population. For interventions that involve tradeoffs that are highly sensitive to patient utilities, interventions for which the relationship between benefits and harms is influenced heavily by personal preferences, the Task Force may abandon population-based recommendations and advocate shared decision-making at the individual level.
 - Consideration of benefits and harms should not be limited to the perspective of individuals but should also consider population effects (e.g., population attributable risk, decreased exposure to infectious diseases, herd immunity).
- 3) The economic costs (direct and indirect) of preventive services, both to individuals and to society, warrant consideration in making recommendations but are not the first priority.
- Although the USPSTF does not consider economic costs in making recommendations, it realizes that these costs are important in the decision to implement preventive services. Thus, in situations where there is likely to be some effectiveness of the service, the TF searches for evidence of the costs and cost-effectiveness of implementation, presenting this information separately from its recommendation.
- 4) Recommendations are not modified to accommodate concerns about insurance coverage of preventive services, medicolegal liability, or legislation, but users of the recommendations may need to do so.
- 5) Recommendations apply only to asymptomatic persons or to those with unrecognized signs or symptoms of the target condition for which the preventive service is intended. They also apply only to preventive services delivered in, or referable from the clinical setting.

- Persons living in the United States are the target population, although it is understood that the evidence reviews and recommendations may be useful in other countries. Recommendations are not intended for populations with markedly different disease patterns and health care services (eg, developing countries).
- The clinical setting to which the recommendations apply are typically primary care ambulatory practices but can also include offices and clinics of specialists, hospitals, emergency departments, public health departments, urgent care facilities, student health centers, worksites, family planning clinics, nursing homes, and home care.
- The evidence for preventive services delivered outside the traditional clinical context (eg, non-clinic based programs at schools, worksites, shopping centers) is often the same, but the recommendations are not primarily intended for this setting.
- Recommendations apply only to asymptomatic persons or to those with unrecognized signs or symptoms of the target condition for which the preventive service is intended. They also apply only to those preventive services for which at least one component, eg, identification or referral, may be delivered in the primary care clinical setting.

5.7 Grades

The Task Force also adopted a set of grades to apply to the evidence. For graded recommendations, appended rationale statements and statements about clinical considerations allow readers to clearly understand the Task Force’s judgment about the certainty of the evidence, the net benefit of implementation, and the overall recommendation about the use of each preventive service.

The Task Force includes a grade that indicates when evidence is insufficient to make any recommendation, these grade ‘I’ topics are accompanied by the same type of rationale and clinical considerations, but are considered “statements” rather than “recommendations.”

The grades may be best understood in the following grid (Table 3):

Table 3. U.S. Preventive Services Task Force Recommendation Grid: Letter Grade of Recommendation or Statement of Insufficient Evidence Assessing Certainty and Magnitude of Net Benefit

Certainty of Net Benefit	Magnitude of Net Benefit			
	Substantial	Moderate	Small	Zero/negative
High	A	B	C	D
Moderate	B	B	C	D
Low	Insufficient			

Grade A indicates that the certainty of evidence is high that the magnitude of net benefits is substantial.

Grade B indicates that the certainty of evidence is moderate that the magnitude of net benefits is either moderate or substantial, or that the certainty of evidence is high that the magnitude of net benefits is moderate.

Grade C indicates that the certainty of the evidence is either high or moderate that the magnitude of net benefits is small.

Grade D indicates that the certainty of the evidence is high or moderate that the magnitude of net benefits is either zero or negative.

Grade I indicates that the evidence is insufficient to determine the relationship between benefits and harms (ie, net benefit),

The Task Force also adopted a plan for appending to the recommendation grade an explicit rationale statement, giving the Task Force's assessment of the overall certainty of the evidence and the magnitude of net benefit. After the rationale statement, the Task Force adds a statement about Clinical Intervention, providing more specific guidance to clinicians.

5.8 Wording of Recommendation or Conclusion Statements

The Task Force also adopted standardized language for the grades given in Figure 5, as shown below:

A: The USPSTF recommends X service for Y population.

B: The USPSTF recommends X service for Y population.

C: The USPSTF recommends against routinely (providing) X service for Y population. There may be considerations that support (providing) the service in an individual patient.

D: The USPSTF recommends against X service for Y population.

I: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of X service in Y population.

Table 4 provides detailed definitions of the grades, with suggestions for clinical practice.

Table 4. What the USPSTF Grades Mean and Suggestions for Practice

Grade	Grade Definitions	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer/provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer/provide this service.
C	The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is moderate or high certainty that the net benefit is small.	Offer/provide this service only if there are other considerations in support of the offering/providing the service in an individual patient.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality or conflicting, and the balance of benefits and harms cannot be determined.	Read “Clinical Considerations” section of USPSTF Recommendation Statement. If offered, patients should understand the uncertainty about the balance of benefits and harms.

Section 6

Other Considerations for Recommendations

6.1 Consideration of Starting and Stopping Times for Screening

In considering the issues of starting/stopping points, the Task Force believes there are weaknesses in the system of using a chronologic age cutpoint. Clearly, factors beyond chronologic age (including risk of the target condition, risk of harms from the preventive intervention, and efficacy of the intervention) are involved in determining when screening an individual should begin or stop.

The Task Force developed a list of factors to be considered when recommending starting or stopping times. These are given in Appendix X.

6.2 Consideration of Screening Interval

Because clinicians are very interested in the recommended interval for screening tests and yet the evidence for one screening test compared with another is often lacking, the issue of what interval to recommend can be problematic. For some topics a decision analysis may be used by the Task Force to help inform its recommendation (see section 5.4.3). The Task Force may not recommend an interval when the evidence is insufficient. It has, however, adopted a set of factors to consider in determining whether to recommend an interval. These factors are given in Appendix XI.

6.3 Outcomes Tables

The Task Force adopted the concept of having “outcomes tables” prepared by the topic team to assist the Task Force in deliberating about the potential net benefits and evidence gaps (Appendix IX). The topic teams are encouraged to provide quantitative information in such metrics as “number needed to treat” (NNT) or “number needed to screen” (NNS) where this is possible; to provide ranges of values where this is most precise; and to write “not available” where this is necessary. The tables are clear about what is known, at what level of precision, and what is not known.

The Task Force recognizes that the NNT or NNS metric does not itself allow comparison of benefits across topics, as the outcomes and times involved are quite different. It does provide, however, a useful first step in making sure the Task Force is as clear as possible about the actual effects of a proposed intervention.

Although the Task Force understands that complete standardization of Outcomes Tables is impossible, it encourages the use of the standards in Appendix IX where possible.

6.4 Extrapolation Policy

“Extrapolation” is a term that is sometimes used synonymously with “generalizability.” To the Task Force, however, these terms have different meanings. While generalizability is a synonym for external validity, extrapolation refers instead to a separate question in the Task Force’s interpretation of evidence. The issue raised by extrapolation is the extent to which the Task

Force can make inferences across evidence gaps within the Analytic Framework to reach a complete chain of evidence connecting the population with health benefits. Examples include whether the Task Force judges that it can extrapolate evidence about intermediate endpoints to health endpoints. Another example is the question of whether the Task Force can judge long-term health outcomes based on shorter-term outcomes. Still another example is whether the Task Force can judge the effects of screening based on RCTs of treatment. In these cases, the Task Force is not considering the question of whether a study applies to a different population, situation, or provider. Instead, it is judging whether a gap in the evidence within the Analytic Framework can be overcome with logic and biologic plausibility. These 2 factors—logic and biologic plausibility—play the greatest role in the decision about extrapolation. In general, the Task Force extrapolates only when the case for extrapolation is strong.

Section 7 Dissemination of USPSTF Recommendations

The recommendations of the Task Force are widely disseminated to professional audiences: in professional peer-reviewed journals, in an electronic tool (ePSS) available on the AHRQ website and as a PDA application (www.epss.ahrq.gov), in print through the annual *Guide to Clinical Preventive Services* (ordering information is available on the AHRQ website at <http://www.preventiveservices.ahrq.gov>), and as reprints in peer-reviewed journals such as the *American Family Physician* and *American Journal of Nurse Practitioners*.

The AHRQ Prevention Program is committed to improving the health of all Americans by working to make sure that *everyone* in the health care system—clinicians, consumers, providers, payers—knows about and uses evidence-based clinical preventive services.

To achieve this goal, AHRQ's prevention program works with public and private partners to reach specific groups and individuals with information about appropriate clinical preventive services, their benefits, and how to improve access to and use of these services.

For more information, see Put Prevention into Practice, on the AHRQ website (<http://www.ahrq.gov/clinic/prevenix.htm>)

Section 8 References

- 1) Barton MB, Miller T, Wolff T, Petitti D, LeFevre M, Sawaya G, Yawn B, Guirguis-Blake J, Calonge N, Harris R; U.S. Preventive Services Task Force. How to read the new recommendation statement: methods update from the U.S. Preventive Services Task Force. *Ann Intern Med.* 2007 Jul 17;147(2):123-127. Epub 2007 Jun 18.
- 2) Guirguis-Blake J, Calonge N, Miller T, Siu A, Teutsch S, Whitlock E; U.S. Preventive Services Task Force. Current processes of the U.S. Preventive Services Task Force: refining evidence-based recommendation development. *Ann Intern Med.* 2007 Jul 17;147(2):117-22. Epub 2007 Jun 18.
- 3) Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001 Apr;20(3 Suppl):21-35.
- 4) Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D; U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med.* 2007 Dec 18;147(12):871-5.
- 5) Whitlock EP, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence-based approach. *Am J Prev Med.* 2002 May;22(4):267-84. Review.
- 6) Woolf SH. An organized analytic framework for practice guideline development: using the analytic logic as a guide for reviewing evidence, developing recommendations, and explaining the rationale. In: McCormick KA, Moore SR, Siegel RA, eds. *Methodology Perspectives*. AHCPH Pub. No. 95-0009. Rockville, MD: Agency for Health Care Policy and Research, 1994, pp. 105-13.
- 7) Donaldson MS, Yordy KD, Lohr, KN Vanselow NA (eds.). *Primary Care: America's Health in a New Era*. Institute of Medicine, Committee on the Future of Primary Care. Institute of Medicine: National Academy Press, 1996.
- 8) Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUORUM Group. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement. *The Lancet.* November 27, 1999;354:1896-19000.
- 9) Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.

Acknowledgments

The Methods Workgroup of the USPSTF acknowledges the special efforts of Russ Harris and Evelyn Whitlock, along with the Procedure Manual workgroup: Lucy Marion, Tess Miller, Ginny Moyer, Judy Ockene and Barbara Yawn, in the creation of this document.

Current and Recent Members of the Methods Workgroup are:

USPSTF

Ned Calonge
Thomas DeWitt, Chair
Allen Dietrich
David Grossman
Russ Harris *
George Isham
Michael LeFevre
Rosanne Leipzig
Lucy Marion
Judith Ockene
Diana Petitti
George Sawaya
Sandy Schwartz
Steven Teutsch *
Tim Wilt

* Former members of the USPSTF

EPC

Tracy Beil
Roger Chou
Susan Norris
Evelyn Whitlock

AHRQ

Mary Barton, Lead
Janice Genevro
Claire Kendrick
Kenneth Lin
Iris Mabry
David Meyers
Tess Miller
Tricia Trinité
Tracy Wolff

Members of the U.S. Preventive Services Task Force

The current membership (July 2008) of the USPSTF is: Ned Calonge, MD, MPH (Colorado Department of Public Health and Environment); Thomas DeWitt, MD (Cincinnati Children's Hospital Medical Center); Allen Dietrich, MD (Dartmouth Medical School); Kimberly Gregory, MD, MPH (Cedars-Sinai Medical Center); David Grossman, MD, MPH (Group Health Cooperative); George Isham, MD, MS (Health Partners, Inc.); Michael LeFevre, MD, MSPH (University of Missouri School of Medicine); Rosanne Leipzig, MD, PhD (Mount Sinai School of Medicine); Lucy Marion, PhD, RN, FAAN (Medical College of Georgia); Bernadette Melnyk, PhD, RN, CPNP/NPP, FAAN, FNAP (Arizona State University); Virginia Moyer, MD, MPH (Baylor College of Medicine); Judith Ockene, PhD, Med (University of Massachusetts Medical School); Diana Petitti, MD, MPH (University of Southern California); George Sawaya, MD (University of California, San Francisco); J. Sanford (Sandy) Schwartz, MD (University of Pennsylvania School of Medicine and The Wharton School); Timothy Wilt, MD, MPH (Minneapolis Veterans Affairs)

Appendixes

Appendix I
Congressional Mandate
Establishing the U.S. Preventive Services Task Force

Under Title IX of the Public Health Service Act, AHRQ is charged with enhancing the quality, appropriateness, and effectiveness of health care services and access to such services. AHRQ accomplishes these goals through scientific research and promotion of improvements in clinical practice, including prevention of diseases and other health conditions, and improvements in the organization, financing, and delivery of health care services (42 U.S.C. 299-299c-7 as amended by Public Law 106-129 (1999)):

- (1) ESTABLISHMENT AND PURPOSE. – The Director may periodically convene a Preventive Services Task Force to be composed of individuals with appropriate expertise. Such a task force shall review the scientific evidence related to the effectiveness, appropriateness, and cost-effectiveness of clinical preventive services for the purpose of developing recommendations for the health care community, and updating previous clinical preventive recommendations.
- (2) ROLE OF THE AGENCY.-The Agency shall provide ongoing administrative, research, and technical support for the operations of the Preventive Services Task Force, including coordinating and supporting the dissemination of the recommendations of the Task Force.
- (3) OPERATION. - In carrying out its responsibilities under paragraph (1), the Task Force is not subject to the provisions of Appendix 2 of Title 5, United States Code.

Appendix II

U.S. Preventive Services Task Force Disclosure Form

This Disclosure Form relates to current topics as listed on page 3 of this form. Those topics for which you have disclosed interests on a previously submitted a Disclosure Form, please update any new interests since the last disclosure was provided. For any new topics, please provide a full disclosure. Period of disclosure is 12 months prior to the date of form completion and continues until final reports are completed.

Disclosure of Significant Financial Interests

Please list each corporation, company, firm, research organization, educational institution, or other organization or institution (proprietary and not-for-profit, domestic and foreign) in which you, your spouse, and dependent children have significant financial interests that are related to the subject matter discussed at this meeting (see p. 3 for topic list). List anything of monetary value that when aggregated for you, your spouse, and dependent children exceeds \$10,000 in value as determined through reference to public prices or other reasonable measure of fair market value, or represents more than 5% ownership interest in any single entity. Please specify source and type of payment. Such interests include but are not limited to: (a) salary or other payments for services (e.g., consulting fees or honoraria); (b) equity interests (e.g., stocks, stock options or other ownership interests); or (c) intellectual property rights (e.g., patents, copyrights and royalties from such rights).

Disclosure of Business and Professional Interests

Please list below your business and professional affiliations with entities that are related to or have interest in the subject matter discussed at this meeting (see p. 3 for topic list). List the name of any corporation, professional society, association, or panel, company, firm, government agency (federal, state and local), research organization, educational institution, or other organization or institution (proprietary and not-for-profit, domestic and foreign) in which your services have been, or will be or is expected to be provided, with or without compensation, including on a part-time or seasonal basis, as an (a) officer; (b) medical staff; (c) board member; (d) owner; (e) trustee; (f) director; (g) expert advisor; or (h) consultant.

Disclosure of Intellectual Interests

AHRQ recognizes that potential intellectual interests are likely to be numerous because Task Force members are chosen for their national reputations on prevention issues; and their work may be sufficiently well-known that users of Task Force products might doubt the objectivity of the process if such members are known to have taken leadership roles in discussion and vote on recommendations regarding that topic. These may range from trivial to substantial. The Task Force previously identified the following as possible intellectual conflicts: (a) public comments and testimony; (b) expert testimony for malpractice; (c) leadership role on a panel; (d) substantial career efforts/interests in a single topic area.

Please list instances in which you have disclosed publicly, either by official verbal or written statements, your position for or against counseling, screening, or intervention that may affect objectivity pertaining to the topics discussed at this meeting.

As a member of the USPSTF, I affirm the following:

◆ I have listed all personal financial interests from the past 12 months in the Disclosure Form (including equity positions, consulting agreements, or employment arrangements) that are related to the topics, including related products that will be discussed at this meeting. (Please see p. 3 for topic list.) Period of disclosure is 12 months prior to the meeting and continues until the final reports are completed.

◆ I have listed all my business and professional affiliations in the Disclosure Form from the past 12 months that have entities related to or have interest in the subject matter of the topics discussed at this meeting. (Please see p. 3 for topic list.) Period of disclosure is 12 months prior to the meeting and continues until the final reports are completed.

◆ I have declared any other real or apparent intellectual conflict(s) of interest related to the subject matter of the current and future topics. (Please see p. 3 for topic list.)

Print Name _____

Signature _____

Date _____

If for any reason you feel you cannot sign this statement as worded or if you have further questions, please contact **Tess Miller** at (301) 427-1585.

PLEASE FAX COMPLETED FORM TO (301) 427-1597 ATTN Gloria Washington

Attachment: List of all current and future (next quarter) topics

APPENDIX III

AHRQ Procedure for Preparing Surveillance Literature Summary

LitWatch Instructions

LitWatch is to be completed every 4-8 weeks, to ensure timely data gathering. Articles for consideration should address primary or secondary preventive interventions in the general primary care setting. (For example, an article on cancer screening in diabetics who work in a dye factory would not be included.) Topics of interest are screening, preventive services, effectiveness of early treatment of screen-detected disease, cost-effectiveness analysis for prevention services or early treatment, new technologies, and methodologies. Special attention should be given to topics currently or previously reviewed by the USPSTF. Articles on vaccines (except HPV or a new breakthrough vaccine), community interventions, or general review articles (unless systematic) should not be included.. Articles should be considered for inclusion in the LitWatch if they meet the following criteria: possible impact on past USPSTF recommendations, new evidence, and/or importance to current reviews of the USPSTF.

The Table of Contents (TOC) for the following journals will be reviewed for potential articles. A list of potentially important articles will be sent to the USPSTF Scientific Director, the AHRQ lead for the Topic Prioritization Work Group (TP WG), and other medical officers and team members. The USPSTF Scientific Director and the AHRQ lead for the TP WG will review the list for studies that may impact significantly the current or past recommendations and bring these to the attention of the TP WG for discussion.

1. AAP Grand Rounds (American Academy of Pediatricians Grand Rounds). Released monthly, e-TOC available.
2. AFP (American Family Physician). Released on the 1st and 15th of the month, e-TOC not-available. Access the Table of Contents on the journal website.
3. Am Journal of EPI (American Journal of Epidemiology). Released on the 1st and 15th, e-TOC available.
4. AJPM (American Journal of Preventive Medicine). Released 10 times per year, e-TOC available.
5. Annals of Family Medicine. Released bimonthly, e-TOC available.
6. Annals of Internal Medicine. Released 1st and 15th, e-TOC available.
7. BMJ (British Medical Journal). Released weekly (Saturdays), e-TOC available.
8. JAMA (Journal of the American Medical Association). Released weekly (Wednesdays), e-TOC available.
9. Journal of Medical Screening. Released quarterly, e-TOC available.
10. JNCI (Journal of the National Cancer Institute). Released 2nd and 16th, e-TOC available.
11. The Lancet. Released weekly (Saturday), e-TOC available.

12. MMWR (Morbidity and Mortality Weekly Report, Centers for Disease Control and Prevention). Released weekly (Friday), e-TOC available.
13. NEJM (New England Journal of Medicine). Released weekly (Thursday), e-TOC available.
14. Obstet Gynecol (Obstetrics and Gynecology). Released monthly, e-TOC available.
15. Pediatrics. Released monthly, e-TOC available.
16. Preventing Chronic Disease. Released quarterly, e-TOC available.
17. ACPM Journal Watch (American College of Preventive Medicine) Released quarterly, no electronic notification established.

Appendix IV

Role and Responsibilities of the USPSTF Topic Work Group Lead

How designated?

AHRQ staff will solicit volunteers for position from among assigned topic leads at the time work assignment is sent to the EPC (or, for internally reviewed topics, once the medical officer has a complete work assignment and commences work).

What are the roles?

At start of, and throughout a topic:

- Scheduling of EPC/TF lead calls will take this individual's schedule into account first. The Topic Work Group lead should be particularly prompt responding to request to schedule calls.
- The calls may be scheduled if and only if this individual plus at least one other Task Force topic lead can be on the call.
- Other members of the topic workgroup who cannot make a scheduled call can direct comments prior to or after a call to a) the Topic Workgroup lead b) all parties, eg, via email or c) the AHRQ medical officer on the topic.

Near the Task Force meeting at which the topic will be voted:

- At the call with the EPC to go over the draft evidence review, and the subsequent call for the TF leads on the topic prior to the task force meeting, the Topic Workgroup Lead is responsible for leading the process of assessing the grouped evidence on each key question by use of one of the following categories: Convincing, Adequate, or Inadequate.

It is expected that the EPC will provide quality ratings of individual studies, and group the evidence by key question together in a summary table for the draft report.

- In addition, the other "lead" responsibilities that the Topic Workgroup lead and the medical officer will facilitate the group in completing during these calls are:
 - assessment of certainty of net benefit across all the key questions in the analytic framework
 - assessment of magnitude of impact of the intervention
 - proposed grade(s) (arrived at from the intersection of the two preceding elements)

It is an expectation that this individual be responsible for presenting the certainty and grade proposal at the TF meeting.

[APPROVED BY USPSTF AT JULY 2007 MEETING]

Appendix V

Solicitation for Nominations for New Primary and Secondary Health Topics to be Considered for Review by the United States Preventive Services Task Force

AGENCY: Agency for Healthcare Research and Quality (AHRQ), DHHS

ACTION: Solicit for new topic nominations

SUMMARY: The Agency for Healthcare Research and Quality (AHRQ) invites individuals and organizations to nominate primary and secondary prevention topics pertaining to clinical preventive services that they would like the United States Preventive Services Task Force (USPSTF) to consider for review. A list of topics that have been recently reviewed or are currently under review by the USPSTF is attached.

The USPSTF is an independent panel of experts that makes evidence-based recommendations regarding the provision of clinical preventive services. Clinical preventive services include screening, counseling and preventive medications. The USPSTF makes recommendations about preventive services for asymptomatic people – people without recognized signs or symptoms of the specific conditions targeted by the preventive service.

Topics can be nominated by individuals, organizations, evidence-based practice centers (EPC) and USPSTF members. The USPSTF will consider nominations and prioritize topics for review based on the following set of criteria: public health importance (burden of suffering, potential of preventive service to reduce the burden); new evidence that has the potential to change prior recommendations including inactive ones; and, potential for greatest Task Force impact (e.g., clinical controversy, practice does not reflect evidence, inappropriate timing in delivery of services). The USPSTF will prioritize topics for which there is a performance gap and the potential to significantly improve clinical practice. Individuals and organizations may nominate new topics or topics previously reviewed by the USPSTF.

Basic Topic Nomination Requirements: Nominations must be no more than 500 words in length and must include the following information. Nominations may include an appendix that contains references and supporting documents (not included in word count).

1. Name of topic
2. Rationale for consideration by the USPSTF, to include:
 - a. Primary or secondary prevention topic (screening, counseling or preventive medication).
 - b. Primary care relevance (applicable clinical preventive service must be initiated in the primary care setting which can be defined as family practice, internal medicine, pediatrics or obstetrics/gynecology and provided by a primary care provider).
 - c. Description of public health importance (burden of disease/suffering, potential of preventive service to reduce burden, including effective interventions). Citations and supporting documents are recommended.
 - d. Summary of new evidence, if any, that has potential to affect the Task Force's recommendation on a previously reviewed topic. Please refer to www.preventiveservices.ahrq.gov for USPSTF recommendations. Citations and supporting documents are recommended.

- e. Description of potential impact of USPSTF's review of the topic, i.e., change in clinical practice, research focus, etc.

DATES: Topic nominations should be submitted by [30 DAYS OF THIS NOTICE PUBLICATION DATE] in order to be considered for 2006-2008. AHRQ will not reply to submissions in response to the request for nominations, but will consider all topic nominations during the selection process. If a topic is selected for review by the USPSTF, the nominator will be notified by AHRQ.

ADDRESSES: Please submit nominations to:

Gloria Washington
ATTN: USPSTF Topic Nominations
Center for Primary Care, Prevention & Clinical Partnerships
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville MD 20850
Fax: 301.427.1597
E-mail: gloria.washington@ahrq.hhs.gov

FOR FURTHER INFORMATION CONTACT: Therese Miller at therese.miller@ahrq.hhs.gov or Gloria Washington at gwashing@ahrq.hhs.gov.

ARRANGEMENT FOR PUBLIC INSPECTION: All nominations will be available for public inspections by appointment at the Center for Primary Care, Prevention & Clinical Partnerships, 301.427.1500, weekdays between 10:00am and 5:00pm (eastern time).

SUPPLEMENTARY INFORMATION:

Background

Under Title IX of the Public Health Service Act, AHRQ is charged with enhancing the quality, appropriateness and effectiveness of health care services and access to such services. AHRQ accomplishes these goals through scientific research and promotion of improvements in clinical practice, including prevention of diseases and other health conditions and improvements in the organization, financing and delivery of health care services (42 U.S.C. 299-299c-7 as amended by Pub. L.106-129 (1999)).

The United States Preventive Services Task Force (USPSTF) is an independent expert panel, first established in 1984 under the auspices of the U.S. Public Health Service. Currently, under AHRQ's authorizing legislation noted above, the Director of AHRQ is responsible for convening the USPSTF to be composed of individuals with appropriate expertise. The mission of the Task Force is to rigorously evaluate the effectiveness of critical preventive services and to formulate recommendations for primary care clinicians regarding the appropriate provision of preventive services. The USPSTF transitioned to a standing Task Force in 2001. Current Task Force recommendations and associated evidence reviews are available at www.preventiveservices.ahrq.gov.

Topic Nomination Solicitation

The purpose of this solicitation for new topics by AHRQ and the USPSTF is to create a balanced portfolio of relevant topics for the current Task Force library. The library is based on

populations, types of services (screening, counseling, preventive medications), and disease types (cancer; heart and vascular disease; injury and violence-related disorders; infectious diseases; mental disorders and substance abuse; metabolic, nutritional and endocrine diseases; musculoskeletal conditions; obstetric and gynecological conditions; pediatric disorders; and, vision and hearing disorders). Selection of suggested topics will be made on the basis of qualifications of nominations as outlined above (see basic topic nomination requirements) and the current expertise of the USPSTF.

Dated:

Carolyn M. Clancy, MD
Director

Notice is released with a list of current topics and topics in progress

Appendix VI

WORK ASSIGNMENT/PLAN TEMPLATE USPSTF Topic Reviews

For each topic, a preliminary draft of this template document will be circulated prior to the first topic conference call with TF leads, EPC researchers, and the AHRQ medical officer. This document will then be revised based on the initial topic call and finalized by the EPC to represent the ‘final work plan.’

Instructions: This template will be used to create a work assignment and for the final work plan. The text in italics provides questions that should be answered about this clinical preventive service as part of that particular section. The questions themselves can be omitted, but the other template items (headers, boilerplate text) should be incorporated into the final work plan for each topic for consistency. Level 1 and 2 Headings should be in Arial font. Level 3 Headings and all body text should be in Times New Roman font.

Project Title:

AHRQ Medical Officer:

Oregon EPC Project Lead Investigator:

Oregon EPC Project Staff:

USPSTF Leads:

Section I. Purpose and Background

Purpose

This report will be used by the United States Preventive Services Task Force (USPSTF) to... *What is the history of this USPSTF recommendation (new, update from 1996, update from update or initial recommendation made after 2000)?*

Condition Background

Condition definition. What defines the disease/condition of interest?

Prevalence and burden of disease/illness. What is the prevalence of the disease/condition overall and in various sub-populations? (If case distribution varies significantly by sub-population, consider whether background questions about high

prevalence groups need to be addressed here, and also whether [and how] issues related to prevalence in subgroups will be handled in this review as part of the ‘scope’ section.)

Who is primarily affected by this disease/condition? (If this condition affects a significant proportion of the population, consider whether there are primary as opposed to secondary causes of the condition, and how issues related to etiology will need to be specified in the review as part of the ‘scope’ section.)

Etiology and natural history. What causes the disease/condition? (If there are multiple causes, consider whether background questions about etiology need to be addressed here and whether a section describing how they are addressed in this review needs to be included in the ‘scope’ section.)

What are the consequences of this disease/condition if left untreated? Is there heterogeneity in its natural history? (If yes, consider whether background questions about natural history also apply to this condition.)

Risk factors. What are risk factors for the condition? Can the individuals primarily affected be practically distinguished as high risk? (If high-risk identification seems to be a potential approach as part of this clinical preventive service, consider whether additional background questions about prevalence and populations with risk factors need to be addressed here or need to be considered in determining the scope of this review.)

Rationale for screening/screening strategies. What is the rationale for screening or early intervention? How is the disease/condition detected? (If there are multiple ways of detecting this disease/condition, consider whether background questions about detection/screening need to be addressed here and whether a section addressing how this review will structure its inclusion/exclusion criteria in order to address this issue should be included in the ‘scope’ section.)

Interventions/treatment. What preventive intervention(s) are of interest? How do they work? (If there are multiple ways of treating this disease/condition, consider whether background questions about intervention/treatment need to be addressed here and whether a section addressing how the review will handle this issue needs to be included in the ‘scope’ section.)

Current clinical practice. What factors in current practice, or in the context of the culture, are important in understanding this clinical preventive service? Have major changes or controversies about this clinical preventive service emerged since the last recommendation was issued? If it is needed as context for the review, a table showing the Recommendations of Other Groups may be included here. (If current practice issues seem important, consider whether background questions about current clinical practice need to be addressed here.)

See Section V for more potential background questions that may apply to a particular topic.

Section II. Previous Review and USPSTF Recommendations (This section is applicable to update topics only)

Previous USPSTF Recommendation(s)

Say: “In (year of recommendation), the USPSTF concluded (statement of the evidence) to recommend (recommendation statement). (Recommendation grade)”

Also list all of the recommendation language in any 1996 or earlier USPSTF topics, even if not a graded recommendation statement.

Previous USPSTF Conclusions

List all conclusions made by the USPSTF in the prior recommendation and rationale statement.

Previous Review Analytic Framework and Key Questions

Insert analytic framework and key questions from the previous topic review.

Previous Review Findings

Insert summary of evidence table with overall quality assessment for each key question. Insert additional main findings from the evidence review as needed in a bullet format. Conclusions should not be included in this section.

List any ongoing studies identified in the previous review.

Previous Review Conclusions

List all conclusions made by the authors of the prior systematic review in the prior Evidence Synthesis and/or article. Make clear which conclusions appear to be based on a stable evidence base and could be used as foundational evidence in this review.

Identified Limitations from Previous Review Process

Identify limitations cited in the prior systematic review, article, and/or recommendation and rationale statement. Identify and list scope limitations or review method limitations you identify from the previous review.

Evidence Gaps Remaining After the Last Review

Summarize the previous review findings, conclusions, and limitations into a series of evidence gaps remaining at the completion of the last review. Order the evidence gaps

into the logical sequence of the analytic framework with section subtitles of Overarching (Direct) Evidence, Screening, Treatment/Intervention, Harms, Costs, and other specific topics (such as Potential Preventable Burden, Current Practice).

Priority Evidence for Updating *(Optional section)*

What studies (types of designs and comparisons) and settings (timeframes, countries, populations) are most applicable to this update? Consider defining the types of evidence necessary to change the USPSTF recommendation, as in the Table below, if that assists in focusing the review.

Table X: Types of evidence required to alter recommendations

2002 Recommendation	Level of evidence required to change the recommendation
I recommendation: Insufficient evidence to recommend for or against routine screening.	<ol style="list-style-type: none"> 1. Good quality RCT or other controlled trials comparing screened and unscreened patients in primary care settings, showing improved outcomes among screened populations. 2. Good quality RCT or other controlled trials comparing screened and unscreened patients in primary care settings, showing no improvement in outcomes among screened populations. 3. Good evidence that the potential benefits of screening and treatment outweigh any potential harms. 4. Good evidence that the potential harms of screening and treatment outweigh any benefits.

Section III. Scan of New Evidence Since Previous Recommendation (see Section V for search strategy)

Organize, summarize, and cite new evidence from the evidence scan and from the systematic review searches in the same order the evidence gaps were presented. Section subtitles should also be the same. Make sure you indicate where priority evidence may or may not be available, based on these initial literature scans.

Previously Identified Pending Studies

Provide follow-up data on previously identified pending studies. Identify new relevant ongoing studies and their expected completion dates and contacts.

Section IV. Update Review Approach

Outline the proposed overall approach to this topic (if an update), answering the following questions and any others necessary to capture a summary of the approach being proposed for this systematic review. Should this be a staged, focused, or complete systematic review? Which key questions in the analytic framework will be addressed? How should they be addressed in terms of sequence and priority? Which key questions will not be updated as their evidence is viewed as “established”? Are there areas that will be updated non-systematically (contextual questions)?

Analytic Framework and Key Questions

All Key Questions.

Update Key Questions.

Contextual Questions.

Review Scope

Specify who and what will be addressed in terms of populations, interventions, comparisons, outcomes, and settings.

Populations. Define the sex, age, and other sociodemographic or medical characteristics of the study participants addressed in this review.

Diseases. Define the spectrum of the disease/condition that the review will include and exclude, including the rationale. (See the background questions on disease/condition to be answered as needed to support your approach). If this condition affects a significant proportion of the population, address how this issue will affect the review. If there are multiple causes of the disease describe how they will be addressed in this review.

Screening interventions. Define the means of detection/screening that will be reviewed, including the rationale (see background questions on screening/detection to be answered as needed to support your approach). If there are multiple ways of detecting this disease discuss how this will be managed in the review. Define outcomes and gold standards as appropriate.

Treatment interventions. Define the methods of treatment/intervention that will be reviewed, including the rationale (see background questions on intervention/treatment to be answered as needed to support your approach). Define outcomes as appropriate.

Study designs. What study designs (types of designs and comparisons) are minimally acceptable for evaluating each key question?

Settings. What settings (timeframes, countries, populations) are minimally acceptable for evaluating each key question?

Costs. Define how cost/cost effectiveness will be addressed in this review. At a minimum, cost-effectiveness articles should be retrieved when encountered as part of the systematic review for key questions (see background questions for other potential cost considerations).

What is Excluded from This Review?

Be clear about decisions to exclude populations, interventions, comparisons, outcomes, or settings, and their rationale.

Search Criteria

Include databases and time periods for each key question, as well as search terms for existing systematic evidence reviews and meta-analyses.

Budget and Timeline

As negotiated with AHRQ after work plan conference calls with AHRQ and the TF leads. Submitted with final work plan as part of the deliverable. Details budget in two phases (work plan development/topic refinement and review phase). Timeline includes major milestones, including expected date for presentation at a future USPSTF meeting. Timeline should be inserted here, but budget should be submitted as a separate document.

Plan to Use Outside Experts and Their Contact Information

Provide information on peer review of workplan as appropriate. If known at the time the final workplan is submitted, justify use or nonuse of previous review team members to provide continuity.

References Cited in Work Plan

Section V. SUPPORTIVE MATERIALS FOR WORK ASSIGNMENT/ WORK PLAN DEVELOPMENT

Methods to Scan for New Primary Evidence Since the Last Review:

Purpose: To locate the main new evidence for a topic, particularly related to update key questions.

Method: AHRQ staff will perform literature searches in the PubMed and Cochrane databases on the benefits and harms of the preventive service. The literature search will use the MeSH terms from the previous evidence review (if available), will search for studies published since the last review (3 months prior to the end-date of the previous search), will be limited to the English language, will be limited to humans, and will be limited to the journals in the abridged Index Medicus (120 "core clinical journals" in PubMed). For the literature search on benefits, the search will be limited to meta-analyses, systematic reviews, and RCTs; for harms, the search will include meta-

analyses, systematic reviews, RCTs, cohorts, and case-control studies. AHRQ staff will prepare a brief (1-2 paragraph) summary of the results of their literature search as part of the work assignment.

Methods of Searching for New Synthesized Evidence Since the Last Review

Purpose: To identify sources of synthesized evidence and ongoing studies that could affect review planning.

Method: At the work plan development stage (after receiving the work assignment), the Oregon EPC librarian works with the topic team to develop a strategy for searching the literature to identify existing systematic reviews and other high quality synthesized literature (such as meta-analyses). The following databases and websites are searched: Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database (UK), National Institute for Health and Clinical Excellence - NICE (UK), Institute of Medicine, MEDLINE via PubMed (using the systematic reviews search hedge developed by NLM), and when appropriate, subject-specific databases (e.g., PsycINFO). Searches are not limited in terms of timeframes, except in the case of databases such as MEDLINE and subject-specific databases, which are searched beginning a year prior to the most recent, relevant citation in the previous review.

The EPC librarian and/or topic team search ClinicalTrials.gov (and CRISP, when necessary) to identify current trials and/or studies relevant to this review. The topic team also checks to determine whether there is a finished, in process, or planned Community Task Force review for the clinical preventive service being reviewed. The timing of this search (work plan stage or later) is left to the discretion of the topic work team.

Other sources for articles: Expert reviewer suggestions during work plan development and/or work plan peer review.

Definitions of Terms

Primary Care: Primary care is the provision of integrated, accessible health care services by clinicians who are accountable for addressing a *large* majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community. (*Primary Care: America's Health in a New Era*. Institute of Medicine (IOM): National Academy Press, 1996.)

Primary Care Interventions Addressed by the USPSTF: The USPSTF considers interventions that are delivered in primary care settings or are judged to be feasible for delivery in primary care. To be feasible in primary care, an intervention could be applicable for patients seeking care in primary care settings, and the skills to deliver the intervention are typically present in clinicians and/or related staff in the primary care setting, or the intervention can generally be ordered/initiated by a primary care clinician.

Types of Questions Answered by Evidence Reviews for USPSTF

Key Questions

Definition: Key questions, in association with the analytic framework, establish the clinical logic to support implementation of a clinical preventive service in primary care. Key questions articulate the details of relevant patients, interventions, outcomes, and comparisons, which are essential in focusing the review on a manageable and clinically relevant topic and which directly guide the systematic literature review process. Each question is clearly tied to a step in the analytic framework, although certain linkages that are already well established may not have a key question that is actively answered during the review for the USPSTF. In addition, there may be reason to focus on an overarching linkage (and the associated key question) in an analytic framework rather than the intermediate linkages (and their associated key questions). All key questions are reviewed and approved by the topic team in the process of assessing and refining the topic before the detailed literature review. Key questions addressed in a systematic review are listed in the methods section and used to organize the results section of the report.

Methods: Key questions are addressed using up-to-date systematic review methods and under the current guidance and methods of the USPSTF. Each key question is addressed through a distinct literature search and reported separately in the results section of the review report.

Update Key Questions

Definition: Update key questions are associated with an analytic framework for a clinical preventive service and are used by the USPSTF to direct the researchers conducting the review to focus on a limited, specific set of key questions from the analytic framework during the updated review process. Update key questions are used to focus the review for a topic update in two main

ways: 1) on examining only critical gaps in the evidence that would be required by the USPSTF in order to make a recommendation, and 2) on evaluating new research addressing parts of the topic that might confirm or change the evidence basis for a previous recommendation. An update key question approach is taken to maximize efficiency in updating when, in the context of the previous review and recommendation, the USPSTF determines that a systematic review of only a limited part of the analytic framework is necessary to validly update its evidence-based recommendation.

Methods: Update key questions are addressed in the same manner as key questions and reported in the results section of the review report.

Contextual Questions

Definition: These questions are not key questions associated with the analytic framework, but represent issues in an updated review for which the USPSTF needs a valid but not necessarily systematic summary of current research in order to provide the context for its vote and recommendation statement. Contextual questions may represent a range of different types of information, including: 1) updated information for a key question that is not being systematically updated; 2) contextual information on natural history, current practice, prevalence and risk groups, or other aspects of the service for which it is strongly believed there will not be information, but which are part of the TF considerations (eg, screening interval, stopping ages for screening, or newer technologies for screening and/or intervention); 3) cost effectiveness.

Methods: Contextual questions are not necessarily addressed systematically, although the approach taken may meet criteria for a systematic review. Comprehensive literature searches are not generally undertaken specifically to answer these questions. Information for contextual questions is gathered in a variety of ways: 1) through targeted literature searches, 2) from authoritative surveys or reviews, 3) from expert input, and 4) opportunistically while reviewing comprehensive literature searches for key questions. Contextual questions are not listed as separate questions in the methods section of the report and are not reported in the results section. The information resulting from their non-systematic review by the researcher conducting the review should be included as part of the introduction or in the discussion section, and related as appropriate to the results of the systematic review.

Potential Background Questions: (these may or may not apply to a particular topic)

Etiology:

What primarily causes the disease/condition?

Are there other causes of the disease condition?

Are these causes common?

How are these other causes to be treated in this review?

Is disease/condition due to other causes detected in the same way?

Does it have the same prognosis/disease impact?

Does it affect the same population?

If included, how should the disease (due to different causes) be distinguished in this review?

Prevalence in key sub-populations:

Do a disproportionate number of clinical cases come from one or a few sub-populations (e.g., 90% of abdominal aortic aneurysm (AAA) ruptures are in older male ever-smokers)?

Natural history:

Is there “pseudo-disease” present among the apparently diseased population?

Risk factors and populations with risk factors:

What is the prevalence of cases in high-risk individuals?

Is there a significant absolute difference in risk between “high” and “low” risk individuals? (eg, with the breast cancer Gail Model, “high” risk has an absolute 5-year risk of 1.66% - still a very low risk, although it is “high” compared with others who have a risk of <1%). This is the “discriminatory” value of a risk calculation. Even though a set of risk factors may double or triple the risk, the risk factors may still be poor at discriminating those who will “get” the clinical condition from those who will not if the risk they are doubling or tripling is small. If risk factors do not discriminate, they may have little clinical use.

Detection/Screening:

If there are multiple ways of detecting, is one (or more) most valid?

If there are multiple ways of detecting, is one (or more) most relevant to current practice?

Which means of detection are of interest for this review?

How will different means of detection be prioritized, combined, or compared in synthesizing the literature?

Intervention/Treatment:

If there are multiple ways of treating the disease/condition, is one (or more) most valid?

If there are multiple ways of treating the disease/condition, is one (or more) most relevant to current practice?

Which treatments are of interest for this review?

How will different treatments be prioritized, combined, or compared in a literature synthesis?

Outcome:

How do we know that treatment is successful (health outcomes of interest to physicians, patients, or their families; intermediate outcomes often measured with established or potential relationship to health outcomes)?

Over what time frame should treatment success be evaluated considering initial results and maintenance of treatment success?

If there are multiple measures of health outcome, is one (or more) most valid?

If there are multiple measures of health outcome, is one (or more) most relevant to practitioners and patients?

How will outcomes be prioritized, combined, or compared in a literature synthesis?

Current Clinical Practice:

What is the current level of use of the service (e.g., What percentage of eligible patients has had it? What percentage of actual practices is doing this regularly?)

How adequate is the provision of the diagnostic, treatment, or monitoring aspects of the preventive service (e.g., follow up of positive mammograms, treatment of hypertension) that are presumed to be in effect for the treatment to be effective?

Cost: (These do not replace the need to identify cost-effectiveness studies as part of key questions)

How much additional “effort”, extra training or equipment does this require from a practice, particularly to get a new service implemented?

Does it require “tracking” people to be sure they get it done regularly and are appropriately followed up? (This would likely not be quantitative, but would be more descriptive.)

What is the cost of delivering the service (actual service costs and time costs for patient and practice) on a population basis?

What is the cost of follow up, especially with respect to false positives?

Appendix VII

Criteria for Assessing Internal Validity of Individual Studies

The Methods Work Group for the US Preventive Services Task Force (USPSTF) developed a set of criteria by which the internal validity of individual studies could be evaluated. The USPSTF accepted the criteria, and the associated definitions of quality categories, that relate to internal validity at its September 1999 meeting.

This appendix describes the criteria relating to internal validity and the procedures that topic teams follow for all updates and new assessments in making these judgments.

All topic teams use initial “filters” to select studies for review that deal most directly with the question at issue and that are applicable to the population at issue. Thus, studies of any design that use outdated technology or that use technology that is not feasible for primary care practice may be filtered out before the abstraction stage, depending on the topic and the decisions of the topic team. The teams justify such exclusion decisions if there could be reasonable disagreement about this step. The criteria below are meant for those studies that pass this initial filter.

Presented below are a set of minimal criteria for each study design and then a general definition of three categories: “good,” “fair,” and “poor,” based on those criteria. These specifications are not meant to be rigid rules but rather are intended to be general guidelines, and individual exceptions, when explicitly explained and justified, can be made. In general, a “good” study is one that meets all criteria well. A “fair” study is one that does not meet (or it is not clear that it meets) at least one criterion but has no known “fatal flaw.” “Poor” studies have at least one fatal flaw.

a. Systematic Reviews

Criteria:

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

Definition of ratings from above criteria:

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.

Case-Control Studies

Criteria:

- Accurate ascertainment of cases
- Nonbiased selection of cases/controls with exclusion criteria applied equally to both
- Response rate
- Diagnostic testing procedures applied equally to each group
- Measurement of exposure accurate and applied equally to each group
- Appropriate attention to potential confounding variables

Definition of ratings based on criteria above:

Good: Appropriate ascertainment of cases and nonbiased selection of case and control participants; exclusion criteria applied equally to cases and controls; response rate equal to or greater than 80 percent; diagnostic procedures and measurements accurate and applied equally to cases and controls; and appropriate attention to confounding variables.

Fair: Recent, relevant, without major apparent selection or diagnostic work-up bias but with response rate less than 80 percent or attention to some but not all important confounding variables.

Poor: Major selection or diagnostic work-up biases, response rates less than 50 percent, or inattention to confounding variables.

ii. Randomized Controlled Trials and Cohort Studies

Criteria:

- Initial assembly of comparable groups
 - for RCTs: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups
 - for 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, cross-overs, adherence, contamination)
- Important differential loss to follow-up or overall high loss to follow-up
- Measurements: equal, reliable, and valid (includes masking of outcome assessment)
- Clear definition of interventions
- All important outcomes considered
- Analysis: adjustment for potential confounders for cohort studies, or intention to treat analysis for RCTs.

Definition of ratings based on above criteria:

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

Fair: Studies will be graded “fair” if any or all of the following problems occur, without the fatal flaws noted in the “poor” category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; 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 done for RCTs.

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

a. Diagnostic Accuracy Studies

Criteria:

- Screening test relevant, available for primary care, adequately described
- Study uses a credible reference standard, performed regardless of test results
- Reference standard interpreted independently of screening test
- Handles indeterminate results in a reasonable manner
- Spectrum of patients included in study
- Sample size
- Administration of reliable screening test

Definition of ratings based on above criteria:

Good: Evaluates relevant available screening test; uses a credible reference standard; interprets reference standard independently of screening test; reliability of test assessed; has few or handles indeterminate results in a reasonable manner; includes large number (more than 100) broad-spectrum patients with and without disease.

Fair: Evaluates relevant available screening test; uses reasonable although not best standard; interprets reference standard independent of screening test; moderate sample size (50 to 100 subjects) and a “medium” spectrum of patients.

Poor: Has fatal flaw such as: Uses inappropriate reference standard; screening test improperly administered; biased ascertainment of reference standard; very small sample size or very narrow selected spectrum of patients.

Appendix VIII

Criteria for Assessing External Validity (Generalizability) of Individual Studies

Each study that is identified as one that provides evidence to answer a KQ is assessed by according to its external validity (generalizability) using the following criteria.

Criteria:

Study Population:

The degree to which the people who were involved as subjects in the study constitute a special population because they were selected from a larger eligible population or were for other reasons unrepresentative of people who are likely to seek or be candidates for the preventive service. The selection has the potential to affect the following:

- absolute risk: The background rate of outcomes in the study could be greater or less than what might be expected in asymptomatic people because of the inclusion/exclusion criteria, because of non-participation, or for other reasons.
- harms: The harms observed in the study could be greater or less than what might be expected in asymptomatic people.

The following are features of the study population and the study design that may cause experience in the study to be different from what would be observed in the US primary care population:

- demographics (age, gender, ethnicity, education, income): The criteria for inclusion/exclusion or non-participation do not encompass the range of people likely to be candidates for the preventive services in the US primary care population
- co-morbidities: the frequency of co-morbid conditions in the study population does not represent of the frequency likely to be encountered in people who seek the preventive service in the U.S. primary care population.
- special inclusion/exclusion criteria: There are other special inclusion/exclusion criteria that make the study population unrepresentative.
- refusal rate (ratio of included to not-included but eligible participants): The refusal rate among eligible study subjects is high, making the enrollees in the study unrepresentative even of the people eligible for the study.
- adherence (run-in phase, frequent contact to monitor adherence): The design of the study has features that may make the effect of the intervention in the study greater than it would be in a clinically observed population.
- stage in natural history of disease; severity of disease: the selection of subjects for the study includes people with at a stage that is earlier or later than would be found in people who are candidates for the preventive service.
- source, intensity of recruitment: The sources for recruiting subjects for the study and/or the effort and intensity of recruitment may distort the characteristics of the study subjects in ways that could increase the effect of the intervention as it is observed in the study.

Situation

The degree to which the clinical experience in the situation in which the study was conducted is likely to be reproduced in other settings

- healthcare system: The clinical experience in the system in which the study was conducted is not likely to be the same as experience in other systems because, for example, the system provides essential services for free when these services are only available at a high cost in other systems.
- country: The clinical experience in the country in which the study was conducted is not likely to be the same as in the U.S. because, for example, services available in the U.S. are not widely available in the other country of study conduct or vice versa.
- selection of participating centers: The clinical experience in which the study was conducted is not likely to be same as in offices/hospitals/settings in which the service will be delivered to the U.S. primary care population because, for example, the centers have ancillary services not available generally.
- time, effort, and system cost for the intervention: The time, effort, and cost to develop the service in the study is more than would be available outside the study setting.

Providers

The degree to which the providers in the study have the skills and expertise likely to be available in general settings

- training to implement the intervention: The intervention in the study was done after giving providers special training not likely to be available or required in U.S. primary care settings
- expertise, skill to implement intervention: The providers included in the study had expertise and/or skills at a level that is higher than the level likely to be encountered in typical settings.
- ancillary providers: The study intervention relied on ancillary providers who are not likely to be available in typical settings.

Global Rating of External Validity (Generalizability):

External validity is rated “good” if:

- the study differs minimally from the US primary care population/ situation/ providers and only in ways that are unlikely to affect the outcome; it is highly probable (>90%) that the clinical experience with the intervention observed in the study will be attained in the US primary care setting.

External validity is rated “fair” if:

- the study differs from the US primary care population/ situation/ providers in a few ways that have the potential to affect the outcome in a clinically important way; it is only moderately probable (50%-89%) that the clinical experience

with the intervention in the study will be attained in the US primary care setting.

External validity is rated “poor” if:

- the study differs from the US primary care population/ situation/ providers in many way that have a high likelihood of affecting the clinical outcomes; the probability is low (<50%) that the clinical experience with the intervention observed in the study will be attained in the US primary care setting.

Appendix IX

Standardization of Outcomes Tables

Although it is not possible to completely standardize outcomes tables because of the diversity of issues among services, the following proposal gives some preferred entries for authors to use in outcomes tables where possible:

1. Time frame:

For services with extended time frame:

10 years. (5 years has typically been used, though 5 years is a short time for many consequences of screening and prophylactic interventions. Since all these calculations require some assumptions, extrapolating to 10 years seems reasonable.) : Alternatives: 5 years, lifetime

For services with short time frame (pregnancy, for example)

1 year.

2. Population:

Express this number as per 1000 individuals targeted, eg, per 1000 women age 40-49

Rationale: Services with large magnitude of impact should have substantial numbers of outcomes when expressed per 1000 individuals, those with less than 1 outcome averted will clearly be interpreted as having relatively small impact.

3. Interventions:

Interventions should be shown in columns and described.

For repeated services, eg, annual fecal occult blood test (FOBT), number of services should be identified.

4. Parameter estimates:

Important parameters should be provided, such as appropriate:

Screening results: sensitivity, specificity

Prevalence of condition

Adherence (to screening, treatment)

Effectiveness

Costs where appropriate (screening, treatment, costs averted)

Intermediate outcomes

Number identified (with and without the condition)

Number treated

5. Outcome Measures (harms and benefits)

Deaths where relevant

Important health outcome, eg strokes averted or cancers caused

QALY's if possible

Harms: adverse events/states

6. Number needed to screen/treat/counsel

Express in outcome terms (NNS to avert one death)

NNC to achieve change in behavior should only be shown if NNC for a health outcome also shown.

Appendix X

Factors to be Considered when Recommending Starting and Stopping Times for Screening

A) Evidence regarding the following factors should be considered when recommending initiation of screening:

- 1) Whether the attributable risk and potential burden of the targeted condition is limited to or increased significantly in subgroups of people who are easily identified by one or more of the following factors: age, gender, ethnicity, particular behaviors (e.g., sexually active), and/or comorbid or biological risk factors;
- 2) Whether the potential to avert risk and burden is decreased by competing risks such as short life expectancy;
- 3) Whether the accuracy of available screening tests differs or is uncertain in particular subgroups of people;
- 4) Whether the feasibility, efficacy, and/or harms of treatment for the risk factor or target condition differ in particular subgroups of people;
- 5) Whether available research about 1-4 is limited to particular subgroups of people, particularly if there is biological or epidemiological knowledge that suggests the risk of disease, the accuracy of the screening test, and/or the efficacy of the treatment may vary significantly across a particular subgroup.

B) Evidence regarding the following factors should be considered when recommending termination of screening:

- 1) Whether risk and potential burden of the targeted condition is absent or decreased significantly in subgroups of people who are easily identified by one or more of the following factors: age, gender, ethnicity, particular behaviors (e.g., sexually inactive), and/or biologic or physical factors (e.g., surgical removal of target organ);
- 2) Whether the potential to avert risk and burden is decreased by competing risks such as short life expectancy;
- 3) Whether the accuracy of screening tests differs or is uncertain in particular subgroups of people;
- 4) Whether the feasibility, efficacy, and/or harms of treatment for the risk factor or target condition differ in particular subgroups of people;
- 5) Whether available research about 1-4 is limited to particular subgroups of people, particularly if there is biological or epidemiological knowledge that suggests the risk of disease, the accuracy of the screening test, and/or the efficacy of the treatment may vary significantly across a particular subgroup;

- 6) Whether the natural history of the target condition suggests a long development or precursor period and prior screening tests have been negative.

Appendix XI

Factors to be Considered when Recommending Screening Interval

Evidence regarding the following factors should be considered when recommending an interval for screening:

- 1) Whether the natural history of the target condition suggests a short or long development or precursor period;
- 2) Whether the incidence rate of the risk factor or target condition remains stable or varies markedly over time according to parameters such as age, particular behaviors, other risk factors, or other medical conditions;
- 3) Whether prior negative or positive screening tests significantly affect the probability of future negative or positive screening tests;
- 4) Whether direct research evidence demonstrates similar or different outcomes in people assigned to different screening intervals.