U.S. PREVENTIVE SERVICES TASK FORCE

PROCEDURE MANUAL

U.S. Preventive Services
TASK FORCE

May 2021

Foreword

Created in 1984, the U.S. Preventive Services Task Force (Task Force) is an independent, volunteer panel of 16 national experts in prevention and evidence-based medicine. Our mission is to improve the health of people nationwide by making evidence-based recommendations on clinical preventive services and health promotion in primary care settings.

The Task Force is committed to making the recommendation development process as clear and transparent as possible so that health care professionals, partners, and the American public are fully informed every step of the way. We share the USPSTF Procedure Manual with this goal in mind.

This Procedure Manual describes the methods used by the Task Force to ensure that its recommendations are scientifically sound, reproducible, and well documented. It is intended as a guide for anyone who is interested in the Task Force, Task Force members, and those who support the Task Force's work, including staff of the Agency for Healthcare Research and Quality and its designated Evidence-based Practice Centers.

The Manual provides a high-level description of the Task Force's structure, governance, and processes for selecting topics, reviewing evidence, soliciting and responding to public input, and arriving at a recommendation. Researchers seeking a more detailed description of methods used to conduct a systematic evidence review may want to review the methods described on the Web site of AHRQ's Effective Health Care Program (effectivehealthcare.ahrq.gov) or read the evidence reviews that are posted with each final Task Force recommendation.

It is important to the Task Force that our colleagues, partners, and the American public understand our procedures. We hope that you will find the USPSTF Procedure Manual helpful and will share it with others who may find it beneficial. If you have any questions or comments, please contact the USPSTF Coordinator at info@uspstf.net.

Together, we can work to improve the health of people nationwide.

Albert L. Siu, M.D., M.S.P.H.,
Chair, U.S. Preventive Services Task Force

Michael LeFevre, M.D., M.S.P.H.
Immediate Past Chair, U.S. Preventive Services Task Force

For the most up to date methods and procedures, please see
https://uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes#methods.

TABLE OF CONTENTS

Figures

Tables

Appendixes

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

1.1 Purpose

The U.S. Preventive Services Task Force's (USPSTF's) mission is to improve the health of people nationwide by making evidence-based recommendations about clinical preventive services and health promotion.

This Procedure Manual documents the methods used by the Task Force to ensure that its recommendations and the reviews on which they are based are of consistently high quality, methodologically sound, scientifically defensible, reproducible, unbiased, and well documented.

The USPSTF is assisted in fulfilling its mission by the Agency for Healthcare Research and Quality (AHRQ), which provides scientific, administrative, and dissemination support to the USPSTF, and by AHRQ-designated Evidence-based Practice Centers (EPCs), which develop the evidence reviews, evidence summaries, and other documents that inform the USPSTF's deliberations. In addition to documenting the USPSTF's methods, this Manual also provides a summary overview of the methods used by AHRQ and EPC staff to support the USPSTF.

1.2 Intended Audience

The Procedure Manual is a user's manual for everyone on the USPSTF team—including AHRQ and EPC staff in addition to 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, methodologists, and members of the public. It is intended to be a "living" document that is constantly updated as methods and processes evolve.

In developing this Manual, the Task Force drew, in part, from a series of articles published by its members, past members, AHRQ staff, and other researchers. A list of these sources is provided in Section 10. Researchers and methodologists seeking further details on the Task Force's methodology may find these articles useful as a complement to the Manual.

1.3 History of the USPSTF

The 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 support for the Task Force was transferred to AHRQ in 1995. The Affordable Care Act of 2010 reauthorized the USPSTF with a slightly different and expanded mandate. Due to the Nation's greater emphasis on prevention, insurers are now required to cover preventive services that are recommended by the USPSTF with a grade of A or B, along with those recommended by the Centers for Disease Control and Prevention's (CDC's) Advisory Committee on Immunization Practices (ACIP), Bright Futures, and the Health Resources and Services Administration's (HRSA's) guidelines for women's health. The Affordable Care Act requires insurers to cover these services with no deductible and no co-pay (**Appendix I**).

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 1996. In 1998, members of the third Task Force were appointed for 5-year terms. The third Task Force released its recommendations incrementally.

Since 2001, the Task Force has featured a rolling panel of members appointed for 4 years, with a portion of the membership being replaced each year. Additionally, Task Force methods were described in a special issue of the *American Journal of Preventive Medicine* that year, including methods for developing recommendations on behavioral counseling and use of analytic frameworks. (See Section 10 for reference.) Following this publication, the Task Force began systematically using analytic frameworks to structure literature reviews and develop recommendations on every topic.

The Task Force now releases its recommendations both incrementally and in periodic publications similar to the "Guide to Clinical Preventive Services."

1.4 Scope of Work

Since its inception almost 30 years ago, the USPSTF has worked to fulfill its mission of improving the health of all Americans by making evidence-based recommendations about clinical preventive services and health promotion.

The Task Force comprehensively assesses evidence and makes recommendations about the effectiveness of clinical primary and secondary preventive services, including screening tests, counseling about healthful behaviors, and preventive medications for children, adolescents, adults, older adults, and pregnant women.

Its recommendations focus on interventions to prevent disease, so they only apply to persons without signs or symptoms of the disease or condition under consideration. USPSTF recommendations address services offered in the primary care setting or services referred by primary care professionals.

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.

1.5 USPSTF Members

There are currently 16 members on the Task Force. Members are nationally recognized experts in prevention, evidence-based medicine, and primary care who are also skilled in the critical evaluation of research and the implementation of evidence-based recommendations in clinical practice. Members' fields of practice include behavioral health, family medicine, geriatrics, internal medicine, pediatrics, obstetrics and gynecology, and nursing. Currently the Task Force is led by a Chair and two Vice-Chairs. Details on the roles and responsibilities of the Task Force members are provided in **Appendix IV**.

1.5.1 Selection of USPSTF Members

Each year, the Secretary of HHS selects new members to replace those members who are completing their appointments. Anyone can nominate a new Task Force member at any time on the Task Force Web site.

The nomination process and required qualifications are described on the Task Force Web site. As of December 2013, the required minimum qualifications are as follows.

Demonstrated knowledge, expertise, and national leadership in the following areas:

1. The critical evaluation of research published in peer-reviewed literature and in the methods of evidence review
2. Clinical prevention, health promotion, and primary health care
3. Implementation of evidence-based recommendations in clinical practice, including at the clinician-patient level, practice level, and health system level

Some USPSTF members without primary health care clinical experience may be selected based on their expertise in methodological issues, such as meta-analysis, analytic modeling, or clinical epidemiology. For individuals with clinical expertise in primary health care, additional qualifications in methodology would enhance their candidacy.

To obtain a diversity of perspectives, AHRQ particularly encourages nominations of women, members of minority populations, and persons with disabilities.

Applicants must have no substantial conflicts of interest, whether financial, professional, or intellectual, that would impair the scientific integrity of the work of the USPSTF and must be willing to complete regular conflict of interest disclosures.

Applicants must also have the ability to work collaboratively with a team of diverse professionals who support the mission of the USPSTF. Applicants must have adequate time to contribute substantively to the work products of the USPSTF.

1.5.2 Terms of Members

In 2001 the USPSTF transitioned to a standing Task Force. Currently, members are invited to serve for a 4-year term, with a possible 1-year extension. New members are selected each year to replace those who have completed their appointments.

1.6 USPSTF Meetings

The Task Force meets three 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.

Formal votes are taken for major procedural and methodological decisions, and for draft and final recommendations.. Votes may be taken for other decisions at the discretion of the Chair. Detailed voting rules are provided in Section 7.4. Key provisions are as follows:

1. All motions on recommendations (at any stage) requiring a vote are passed when two thirds of the current Task Force membership vote "yes."
2. Motions on procedural, methodological, and other decisions which require a vote are passed when a majority of current Task Force membership votes "yes."
3. Votes are submitted as "yes," "no," "abstain," or "absent." Votes are taken by voice, hand, or email, without secret ballots.
4. Members recused for reason of potential conflict of interest are recorded as recused and do not vote.
5. In votes that are less than unanimous, there are no minority reports.
6. A vote must be held to reconsider the grade of a previously voted draft or final recommendation statement. Two thirds of the current Task force membership must approve the request to reconsider. If the request to reconsider is approved, the topic leads review and present the evidence supporting the motion. The Task Force then votes on the new recommendation either in person or by email.

1.7 Conflict of Interest

1.7.1 Introduction

The public must have confidence in the integrity of the process by which the Task Force makes its recommendations.[1] 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 recommendation statements 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.

The intent of requesting disclosure of any potential conflict of interest is to ensure that the USPSTF provides a balanced, independent, objective, and scientifically rigorous product (including its recommendation statements) by understanding other interests that could potentially influence the work and decision-making of its members. The USPSTF requires each member to disclose all information regarding any possible financial and non-financial conflicts of interest prior to each meeting for all topics under development or that will be discussed at each meeting. Previous disclosures for continuing topics must also be updated to reflect changes in a member's situation since the form was last completed.

It is important to note that disclosures are not considered actual conflicts of interest until the value and nature of the disclosure is reviewed by the Task Force chairs.

1.7.2 Process for Completing Disclosure Forms

The USPSTF Disclosure Form will be completed by Task Force members prior to each meeting to provide information on potential financial and non-financial conflicts of interest related to USPSTF topics under consideration. Task Force members are expected to provide full disclosure for new topics and topics in development, as well as an updated disclosure that reflects changes in their situation for continuing topics.

All members are expected to provide full disclosure of their own interests as well as the interests of immediate family members (which includes their spouse/partner, dependent children, and parents) and those of other close personal relationships.

The period of disclosure is 36 months prior to the date of form completion. The exception is publications related to the topic, for which there is no time limit, and research grants, for which the period of disclosure is 36 months from the end of the grant period. Completed Disclosure Forms will be kept on file. Further information on each type of disclosure required is provided below.

Disclosure of Significant Financial Interests
Financial disclosures refer to relationships with entities that could influence, or give the appearance of influencing, the outcome of a USPSTF decision. Entities could be individuals, organizations and corporations, or other groups with established or future business in the matter of a USPSTF decision. A relevant financial interest is a situation in which a Task Force member, immediate family member, or close personal relation has the potential for direct or indirect

[1] Institute of Medicine, Clinical Practice Guidelines We Can Trust (2011). Available at http://iom.nationalacademies.org/reports/2011/clinical-practice-guidelines-we-can-trust.aspx. Accessed 11/10/15.

financial gain or loss related to a Task Force product. Task Force members should disclose financial relationships for themselves, their immediate family members, and close personal relationships. It is important to note that Task Force members report all relevant financial relationships regardless of the amount. Relevant financial interests include, but are not limited to:

a. Ownership or owning individual stocks (stock shares, options, warrants), and bonds or other debt or other significant proprietary interests or investments in any third party that could be affected by a USPSTF decision on a specific topic. (Diversified non-sector mutual funds in which stocks are chosen by an independent fund manager may not need disclosure)
b. Having an employment, independent contractor or consulting relationship or other contractual arrangements, whether written or unwritten, with an entity that could be financially or reputationally affected by a Task Force decision
c. Receiving a proprietary research grant or receiving patents, royalties or licensing fees from such an entity
d. Participating on an entity's proprietary governing board or advisory council
e. Participating in an entity's speakers bureaus
f. Receiving honoraria or travel from such an entity
g. Receiving payment as an expert witness for a plaintiff or a defendant associated with such an entity
h. Receiving remuneration for services with respect to transactions involving parties with a financial interest in the outcome of a USPSTF decision. This may include clinical specialty practice.

There is no set minimum dollar amount for financial disclosure because any relevant financial relationship could be considered significant.

Financial interests that do not need to be disclosed include:

a. Income from seminars, lectures, teaching engagements, service on advisory committees or review panels for public entities or nonprofit organizations that do not have a vested interest in the specified topics
b. Diversified mutual or retirement funds

Disclosure of Significant Non-Financial Conflicts of Interest

Non-financial conflicts of interest are other relationships, activities, or stated positions that could influence or give the appearance of influencing the work of a member of the USPSTF. In addition, non-financial COIs are considered to be any strongly held beliefs related to a topic area that would make it difficult for a Task Force member to work on any new or related topic. Task Force members should disclose these relationships, activities or stated positions for themselves, their immediate family members, and close personal relationships. These disclosure requests are intended to identify strongly held opinions that may not be open to alternative conclusions even if provided with adequate evidence to the contrary. It also includes interests or institutional relationships that are not direct financial COIs but may influence or bias the individual.

The Task Force recognizes that potential non-financial interests are likely to be numerous because Task Force members are chosen for their national reputations on prevention issues; and their work may be very well-known. As a result, 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. Task Force members are required to disclose substantial non-financial interests including, but not limited to:

a. Public comments and testimony
b. Leadership role on a panel
c. Substantial career efforts/interests in a single topic area
d. Previously published opinions
e. Advocacy or policy positions

In addition, potential non-financial interests requiring disclosure include any relationships with or investments in governmental organizations, healthcare organizations, professional societies, or other organizations that you have reason to believe may benefit or be harmed by Task Force recommendations. This includes services that are provided on a part-time or seasonal basis, service that has occurred in the past or is anticipated in the future, and includes services for which compensation may have been provided as an:

- Officer
- Medical staff
- Board member
- Director

- Expert advisor
- Consultant

Non-Financial Interests that do not need to be disclosed include:

a. Employment from nonprofit organizations such as government agencies and nonprofit entities that do not have a vested interest in the specified topics
b. General membership in a professional society
c. Attendance at presentations or conferences related to the topic of interest

Prospective Task Force Members

Prospective Task Force members will be verbally informed of the USPSTF COI policy by the Task Force chair and/or co-chair during the review of their candidacy. Appointees will be required to submit a Disclosure Form prior to finalizing their appointment. The USPSTF Disclosure Form will also be completed by new Task Force members prior to participation in their first in-person meeting.

1.7.3 Process for Determining Appropriate Actions

After disclosures are submitted and prior to each meeting or to new member appointment, all disclosures will be aggregated and reviewed by the Task Force Chairs. The Task Force Chairs will determine the final action on the member's eligibility to participate on a specific topic(s) which also is kept on file.

Each member is notified of the final action. 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. For example, 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.

For disclosures and assessment of potential COI of Task Force chairs, the two chairs not under review determine a final action. This process is followed for each of the three Task Force chairs.

Prior to each meeting, Task Force members will receive a summary of all disclosures that will be publicly announced during the meeting. At the start of each meeting, the Task Force Chairs will announce these disclosures and provide an opportunity for members to ask questions and engage in discussion.

Below is a list of disclosures representing potential conflicts and the possible actions that can be recommended by the Task Force Chairs for each disclosure.

Table 1. Description of Disclosures and Recommended Actions

Level	Type of Disclosure	Range of Possible Recommended Actions	Description
1	• No financial disclosures of any value • No non-financial disclosures that would impact the judgment of the Task Force member Financial Interests that do not need to be disclosed: • Income from seminars, lectures, teaching engagements, service on advisory committees or review panels for public entities or nonprofit organizations that do not have a vested interest in the specified topics • Diversified mutual or retirement funds Non-Financial Interests that do not need to be disclosed: • Employment from nonprofit organizations such as government agencies and nonprofit entities that do not have a vested interest in the specified topics • General membership in a professional society • Attendance at presentations or conferences related to the topic(s) of interest	No Action	No disclosure or recusal necessary
2	• Providing public comments, expert testimony, or participation in speaking bureaus on a relevant topic (excluding speaking engagements on behalf of a product) • Any relevant financial disclosure valued at $1000 or less • Participation in any governmental organizations, professional societies, or other organizations (as an officer, medical staff, board member, director, expert advisor, or consultant) related to the topic(s) of interest • Serving as editor or deputy editor of an academic journal, book or website	Information disclosure to Task Force only.	Member may participate as primary lead, and may discuss and vote on the topic
3	• Any relevant financial disclosures valued at more than $1000 • Participation in any proprietary companies (as an officer, medical staff, board member, director, expert advisor, or consultant) related to the topic(s) of interest such that the member would stand to gain financially from a specific outcome of a recommendation statement. • Speaking engagements on behalf of a product • If a member has a significant non-financial interests in a specific outcome of a recommendation statement	Possible exclusions from Task Force roles as a result of Level 3 disclosures Include: Member may not serve as primary lead for topic workgroup	Member may not participate as the primary lead of the topic workgroup specific to the conflict, but may serve as a lead on the topic workgroup and discuss and vote on the topic.
		Member may not serve as the primary spokesperson for the topic	Member may not participate as the primary spokesperson for the topic specific to the conflict, but may serve as a lead on the topic workgroup and discuss and vote on the topic.

Level	Type of Disclosure	Range of Possible Recommended Actions	Description
	• The member has one or more publications or research grants that are likely to be part of the evidence review, and that address key questions in the analytic framework, or that express opinions related to the topic. Whether or not action is needed will depend on the specific content of the publications and/or grants, and the source of funding of any grants.	Member may not serve as a lead on the topic workgroup	Member may not participate as a lead in the topic workgroup specific to conflict, but may discuss and vote on the topic.
		Recusal from all participation in topic activities	Member may not participate as a lead on the topic workgroup specific to conflict 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.

The member may choose to disclose to the Task Force chairs either a strongly held opinion that results in the potential for bias, or a personal or family illness that may lead to bias but which should be held confidential. This may result in recusal from a particular topic, at the discretion of the Task Force chairs.

If a relationship could be classified in more than one level (for example, service as a medical editor [Level 2] that is compensated with more than $1000/year [Level 3]) it would be classified at the higher level (Level 3, in this case).

1.7.4 Process for Sharing USPSTF Disclosures and Actions with the Public

The USPSTF posts a summary of Level 3 disclosures for any topic on the COI page of the USPSTF website. Additionally, International Committee of Medical Journal Editors (ICMJE) disclosure forms from USPSTF authors are available for each Recommendation Statement from the journal in which the Recommendation Statement is published.

Policy for Other Affiliated Groups
EPC members file separate disclosure forms consistent with EPC procedures and are kept on file.

1.8 Partner Organizations

Partner organizations provide ongoing liaison to the USPSTF. They include Federal agencies that are stakeholders in the process (Federal Liaisons) and Dissemination and Implementation Partners that represent primary care clinicians, consumers, and other stakeholders involved in the delivery of primary care. Partner organization representatives contribute their expertise, help disseminate the work of the USPSTF to their members and constituents, and help put the recommendations into practice. They are invited to attend and observe the USPSTF meetings and are permitted to comment on the proceedings during the meetings.

Like the public, partners are invited to review draft research plans, evidence reviews, and recommendation statements, 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.

Federal Liaisons currently include the Centers for Disease Control and Prevention (CDC); Centers for Medicare and & Medicaid Services (CMS); Community Preventive Services Task Force; 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 Cancer Institute (NCI); National Institutes of Health (NIH); Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion (ODPHP); Substance Abuse and Mental Health Services Administration (SAMHSA); and the U.S. Food and Drug Administration (FDA).

Dissemination and Implementation Partners currently include AARP, America's Health Insurance Plans (AHIP), American Academy of Family Physicians (AAFP), American Academy of Nurse Practitioners (AANP), American Academy of Pediatrics (AAP), American Academy of Physician Assistants (AAPA), American Congress of Obstetricians and Gynecologists (ACOG), American College of Physicians (ACP), American College of Preventive Medicine (ACPM), American Medical Association (AMA), American Osteopathic Association (AOA), American Psychological Association (APA), Canadian Task Force on Preventive Health Care, Community Preventive Services Task Force, Consumers Union, National Association of Pediatric Nurse Practitioners (NAPNAP), National Business Group on Health, National Committee for Quality Assurance (NCQA), and the Patient-Centered Outcomes Research Institute (PCORI).

1.9 Overview of the Process

As illustrated in **Figure 1**, four groups are involved in the process that results in formulating Task Force recommendations: the Task Force, AHRQ, the EPC, and Task Force partners. Each plays a unique role in the process.

The **Task Force** selects and prioritizes topics for review, approves the analytic framework, determines the questions and outcomes of interest, interacts with the EPC about evidence issues, judges and grades the level of the available evidence, determines the balance of benefits and harms, and makes the recommendation.

AHRQ convenes the Task Force and provides ongoing administrative, research, and technical support for its operations, including coordination of and support for the dissemination of recommendations. An AHRQ Medical Officer joins the topic team to provide technical input and assist with coordination. In addition, AHRQ staff occasionally prepares in-house evidence reviews for some update and reaffirmation topics (see Sections 2 and 4 for more information on reaffirmations).

Under contract to AHRQ, **EPCs** conduct systematic reviews of specified questions concerning the evidence on prioritized topics in clinical prevention. EPC evidence reviews 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; drafting an analytic framework that illustrates the questions, populations, interventions, and outcomes of interest; locating and retrieving the relevant evidence; evaluating the quality of individual studies; qualitatively and/or quantitatively summarizing review findings for each question for use by the USPSTF in its evaluation of the evidence; and producing the reports. Further details about EPCs are available at www.ahrq.gov/research/findings/evidence-based-reports/overview/index.html.

USPSTF partners are invited to review and comment on draft research plans, evidence reviews, and recommendation statements. Partners are encouraged to disseminate Task Force recommendations to their members. Further details about the role of partner organizations are provided in Section 1.8.

Lastly, anyone can nominate new Task Force members and new topics for the Task Force to consider. In addition, the USPSTF seeks feedback from the **public** on its draft research plans, evidence reviews, and recommendation statements.

The procedures for developing a recommendation statement are presented in **Figure 2**. A brief summary follows. Each step is also described in more detail in subsequent sections of the Procedure Manual.

1.9.1 Topic Selection

Topic selection begins with the identification of topics to be considered. Anyone—including individuals, organizations, EPCs, and Task Force members—can nominate a new topic for Task Force consideration or request an update of an existing topic through an online nomination form on the USPSTF Web site. Once a year, the Task Force Topic Prioritization Workgroup drafts a prioritized list of topics, including new topics and updates, to be started during that year. This list is made according to the following criteria for prioritization: public health importance (burden of suffering and potential of preventive service to reduce the burden); potential change to a prior recommendation (e.g., because new evidence has become available); and 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). The ultimate goal is to balance the annual portfolio of topics by population, type of service (screening, counseling, preventive medication), type of disease (e.g., cancer, endocrine disease), and size of project (e.g., update vs. new topic). The Task Force also aims to update topics every 5 years in order to keep its library of recommendations current.

Task Force leads) develops a preliminary work plan as described in Section 3. AHRQ organizes a conference call of the entire topic team to discuss and refine the project scope and finalize the work plan.

1.9.3 Work Plan External 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.

1.9.4 Research Plan Development

Based on the full final work plan, a "research plan" that contains only the analytic framework, key questions, and inclusion/exclusion criteria is created for public comment.

1.9.5 Draft Research Plan Public Comment

All draft research plans are posted on the USPSTF Web site for public comment for a period of 4 weeks. USPSTF partners are encouraged to submit comments via the Web site.

1.9.6 Finalization and Approval of Work Plan

The work plan is revised based on public and partner comments and expert review. Work plans for new topics are usually presented by the EPC to the entire Task Force. The EPC's presentation is followed by comments from Task Force topic 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 topic leads, but are not routinely presented to the entire Task Force for discussion.

1.9.7 Draft Evidence Review

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 in a draft evidence review, with evidence tables. The EPC presents a summary of the draft evidence review to the leads by teleconference before discussion and deliberation by the entire Task Force.

1.9.8 Review of Draft Evidence Review by Task Force Leads and External Experts

All draft evidence reviews are sent to a limited number of experts in the field for review (**Appendix XV**). In addition, Task Force topic leads and AHRQ Medical Officers are asked to comment on the draft evidence review.

Figure 1. Group Roles in the Task Force's Recommendation Development and Dissemination Processes

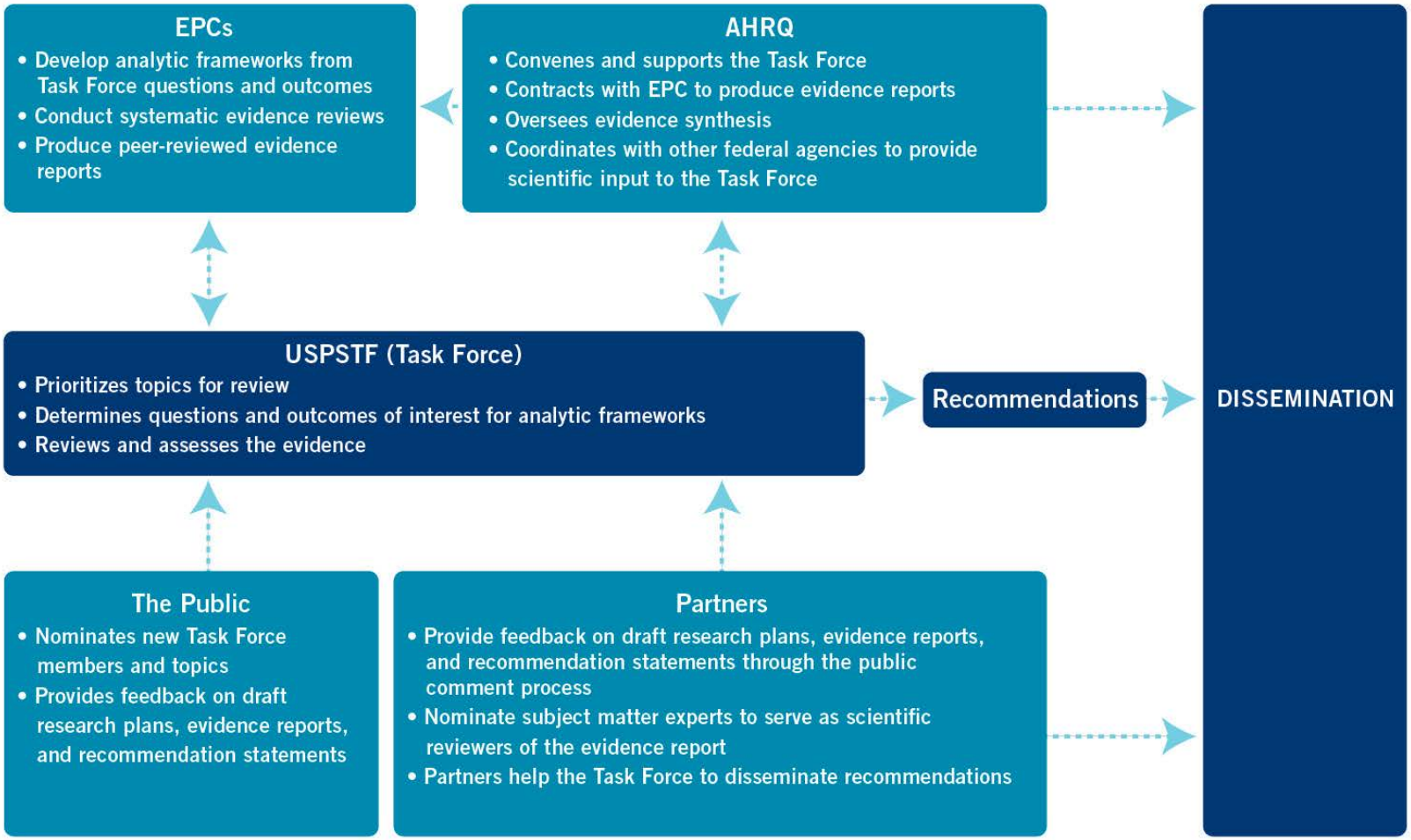
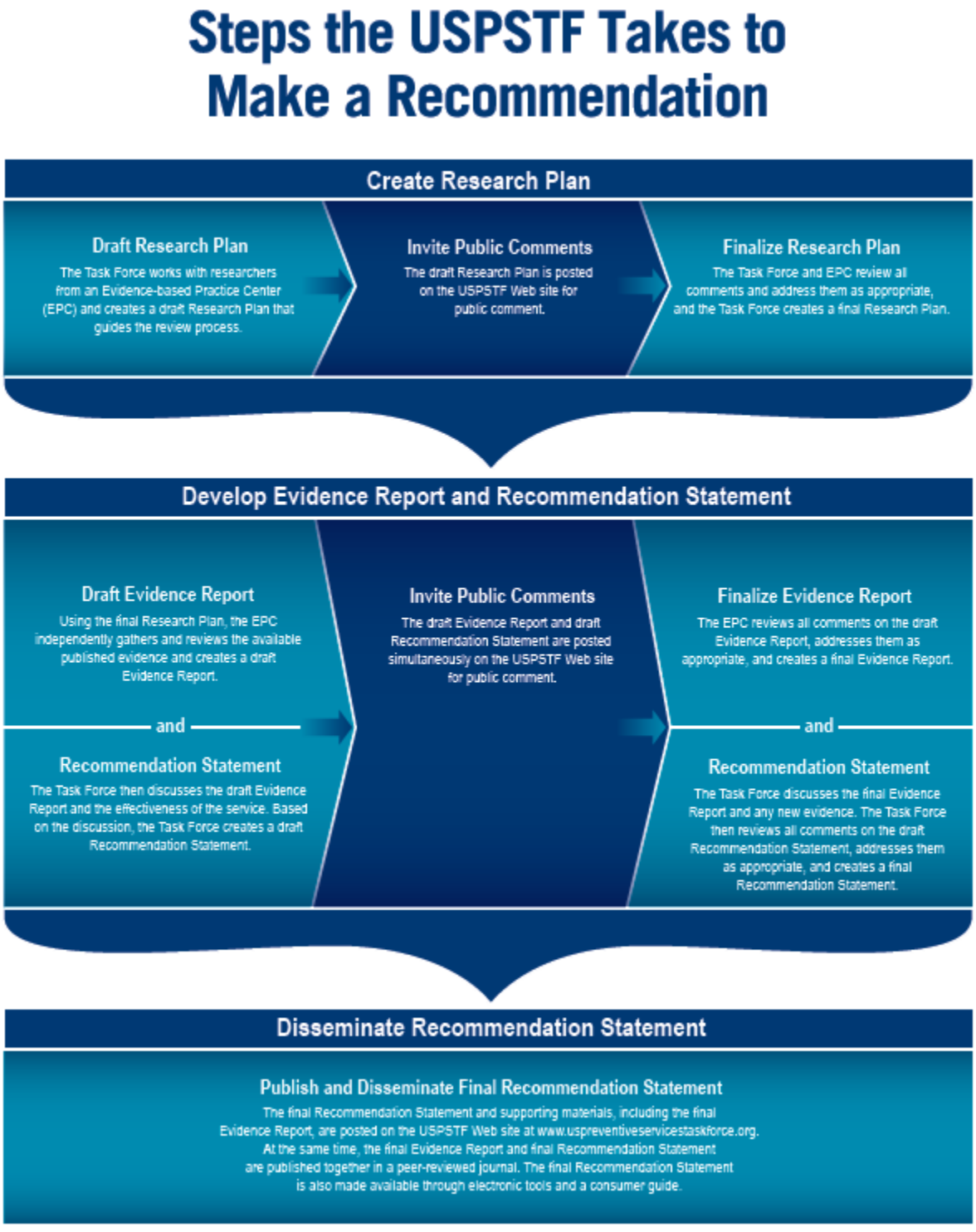


Figure 2. Steps the USPSTF Takes to Make a Recommendation



1.9.9 Development of Draft Recommendation Statement

While the draft evidence review is under review and revision, the Task Force topic leads discuss specific recommendations and the content of the Clinical Considerations section of the recommendation statement. The Task Force leads draft the recommendation statement with the AHRQ Medical Officer, which is presented to the entire Task Force at its next meeting.

1.9.10 USPSTF Vote on Draft Recommendation Statement

At the Task Force meeting, a representative from the EPC presents the expert-reviewed evidence review findings, and the Task Force topic 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 when the work plan is approved to when the peer-reviewed evidence review and draft recommendation statement are presented to the Task Force for a vote.

1.9.11 Public Comment on Draft Evidence Review and Draft Recommendation Statement

The draft evidence review and draft recommendation statement are typically posted together on the USPSTF Web site for public comment for a period of 4 weeks. During the comment period, any member of the public may submit comments on either or both of the documents. USPSTF partners are encouraged to submit comments.

1.9.12 Final Evidence Review

After receiving and reviewing all comments in the Draft Evidence Review from experts, partners, the public, the USPSTF (in particular, the topic leads), and the AHRQ Medical Officer, the EPC revises the evidence review. The EPC sends a summary of all comments received and the revised evidence review, indicating how the comments were addressed, to the AHRQ Medical Officer and made available to the Task Force. After the AHRQ Medical Officer has reviewed and approved the revised document, the review is considered final. At this point, the EPC may undertake preparation of a manuscript to be submitted for publication in a peer-reviewed journal. An effort is made to synchronize publication in the journal with the publication and/or release of the USPSTF final recommendation statement.

1.9.13 Development of Final Recommendation Statement

The Task Force leads working with the AHRQ Medical Officer propose revisions to the recommendation statement based on discussion at the meeting and all comments received from the public, experts, and partners. This revised recommendation statement is sent to all Task Force topic leads for approval.

1.9.14 Approval of Final Recommendation Statement

The final recommendation statement is then sent to all Task Force members for ratification, usually via email.

1.9.15 Release of Final Recommendation Statement and Final Evidence Review

An arrangement is made with appropriate journals to publish the final recommendation statement (which is published by the journal without substantive editing) and the manuscript derived from the EPC review. The desired timeline from USPSTF vote to recommendation release is 9 months. All final recommendation statements and supporting evidence are made available on the USPSTF Web site (www.uspreventiveservicestaskforce.org).

1.10 Procedures for Writing Papers and Documents

Task Force recommendations are usually published in a peer-reviewed journal. The Chair is listed as the author on behalf of the Task Force. Previous Task Force members who made significant contributions to the recommendation, such as leads on the topic workgroup who have since rotated off the Task Force, are also acknowledged. Members serving at the time of the recommendation's finalization are listed in an appendix to the publication.

Evidence summaries (articles summarizing evidence reviews produced by EPCs for each topic) are usually published in the same peer-reviewed journals as the corresponding recommendations. Authors include EPC staff contributors.

Additionally, the Task Force often disseminates its methods and processes through publication in peer-reviewed journals. When Task Force methods are shared through publication, clinicians and the general public can better understand the work of the Task Force and consider Task Force recommendations when making health care decisions.

Each individual designated as an author of a paper to be submitted to a journal 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 final approval of the version to be published. All three conditions must be met, and all who qualify for authorship should be listed. USPSTF 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 members or EPC staff 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 Task Force works under severe time constraints in producing its products. Accordingly, Task Force members and EPC staff 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).

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 www.icmje.org.

Section 2. Topic Selection, Prioritization, and Updating

The Task Force has a large library of current topics and frequently receives nominations for new topics. The overall goal for 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 population, type of service (screening, counseling, preventive medication), type of disease (e.g., cancer, endocrine disease), and size of project (e.g., update vs. new topic). The Task Force also prioritizes topics with the aim of updating topics every 5 years, in accordance with currency 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 of active topics (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. The processes for developing work plans, assessing evidence, and making recommendations for active topics are discussed in Sections 3−7.

2.1.1 Active Topic Types

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

Updated topics are topics reviewed in the past by the Task Force 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 Task Force 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 justify a change in the grade of the recommendation. Only recommendations with a current grade of A or D are considered for a reaffirmation evidence update. The procedure for a reaffirmation evidence update is discussed in Section 4.7.

Referred topics are topics in which the Task Force refers providers to another organization's recommendation. The Task Force originally made a recommendation on these topics and are retained as active in the Task Force library; however, the Task Force has determined that there is another organization (e.g., the CDC's Community Preventive Services Task Force, ACIP) with evidence-based methods that is better positioned to make accurate and timely recommendations for the topic. The procedure for referring to other organizations is discussed in Section 2.5.

2.1.2 Inactive Topics

Inactive topics are topics the Task Force has decided to not update or keep active for one or more reasons (go to Section 2.4 for more details).

2.2 Determination of Scope and Relevance of New Topic Nominations and Topic Selection

Anyone can nominate a new topic for Task Force consideration or request an update of an existing topic at any time online at www.uspreventiveservicestaskforce.org/Page/Name/nominating-recommendation-statement-topics.

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

The Topic Prioritization Workgroup first considers whether newly nominated topics are within the scope of the Task Force, using the following criteria:

- The focus population should be asymptomatic for the condition of interest
- The nominated topic should represent a clinical preventive service (e.g., screening test, preventive medication, counseling about healthful behaviors)

- The preventive service should meet the definition of primary prevention (i.e., avoid the development of disease) or secondary prevention (i.e., identify and treat an existing disease before it results in significant symptoms)
- The preventive service should be provided in or referable from 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.

The Task Force considers interventions that are delivered in primary care settings or are judged to be feasible for delivery in or referable from 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.

Topics that are within the scope of the Task Force are then assessed for relevance using the following criteria:

- Public health importance (i.e., burden of suffering and expected effectiveness of the preventive service to reduce that burden)
- 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)
- Balance of Task Force portfolio (i.e., does the nomination overlap with current or in-process Task Force recommendations; does the nomination balance the overall Task Force portfolio of recommendations by population, type of service, type of disease, and/or size of project)

Based on the above criteria, the Topic Prioritization Workgroup assigns each nomination to one of the following categories for consideration by the entire Task Force:

1. Not a potential new topic:
    a. Out of scope
    b. In scope, of less relevance
    c. In scope, already addressed
2. In scope, potential new topic

The entire Task Force ultimately votes on the selection of potential new topic nominations for inclusion in the Task Force portfolio. As new topics are selected, the Topic Prioritization Workgroup and full Task Force prioritize the potential new topics in comparison with existing new topic nominations. The Task Force maintains a list of one to three new topic nominations for possible review over the next 2 years. All potential new topics enter the yearly prioritization process (described in Section 2.3).

2.3 Prioritization and Selection of Active Topics

The Topic Prioritization Workgroup begins prioritization of an active group of topics approximately 3 years after their previous publication.

Step 1. A brief background paper on the topic is produced that includes the following information: previous recommendation statement, estimate of disease burden, relevance to prevention and primary care, recommendations of other guideline developers, existing controversy or gap between evidence and practice, 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 (go to Section 2.5 for the process of referring a topic and Section 2.4 for the process of inactivating a topic).

Step 3. A request for feedback on all active topics and potential new topics, is sent to Task Force members and partner organizations. Respondents are asked to categorize each proposed topic as high-, moderate-, or low-priority for review in the next 12 to 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 the prior recommendation
4. Need for a balanced portfolio of topics

Step 4. The feedback from Task Force members and partner organizations is considered by the Topic Prioritization Workgroup, along with the background paper, in assigning a tentative priority category for active topics. 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 (active, inactive, refer) and prioritization (high, moderate, low) becomes final after a vote of the full Task Force membership.

Steps 2 to 5 are repeated yearly for topics not selected for review in the preceding year (**Figure 3**).

AHRQ staff develops the work queue for the next 12- to 18-month cycle using the priority level determined by the Task Force. 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 for one or more of the following reasons:

1. Topic is no longer relevant to clinical practice because of 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 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 Task Force'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 Task Force 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

Recommendations for some topics in the Task Force 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 Task Force. An example is ACIP, a non-Federal panel of immunization experts convened by the CDC. In the past, the Task Force has referred recommendations on immunizations to ACIP. Another example is the CDC-supported Community Preventive Services Task Force, 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). 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. The Topic Prioritization Workgroup reviews the previous Task Force recommendation statement and evidence review.
3. The Topic Prioritization Workgroup reviews the recommendations and review methods of the chosen organization.
4. A brief summary is prepared that includes 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's, a new recommendation date, and a statement that the previous evidence review will not be updated.

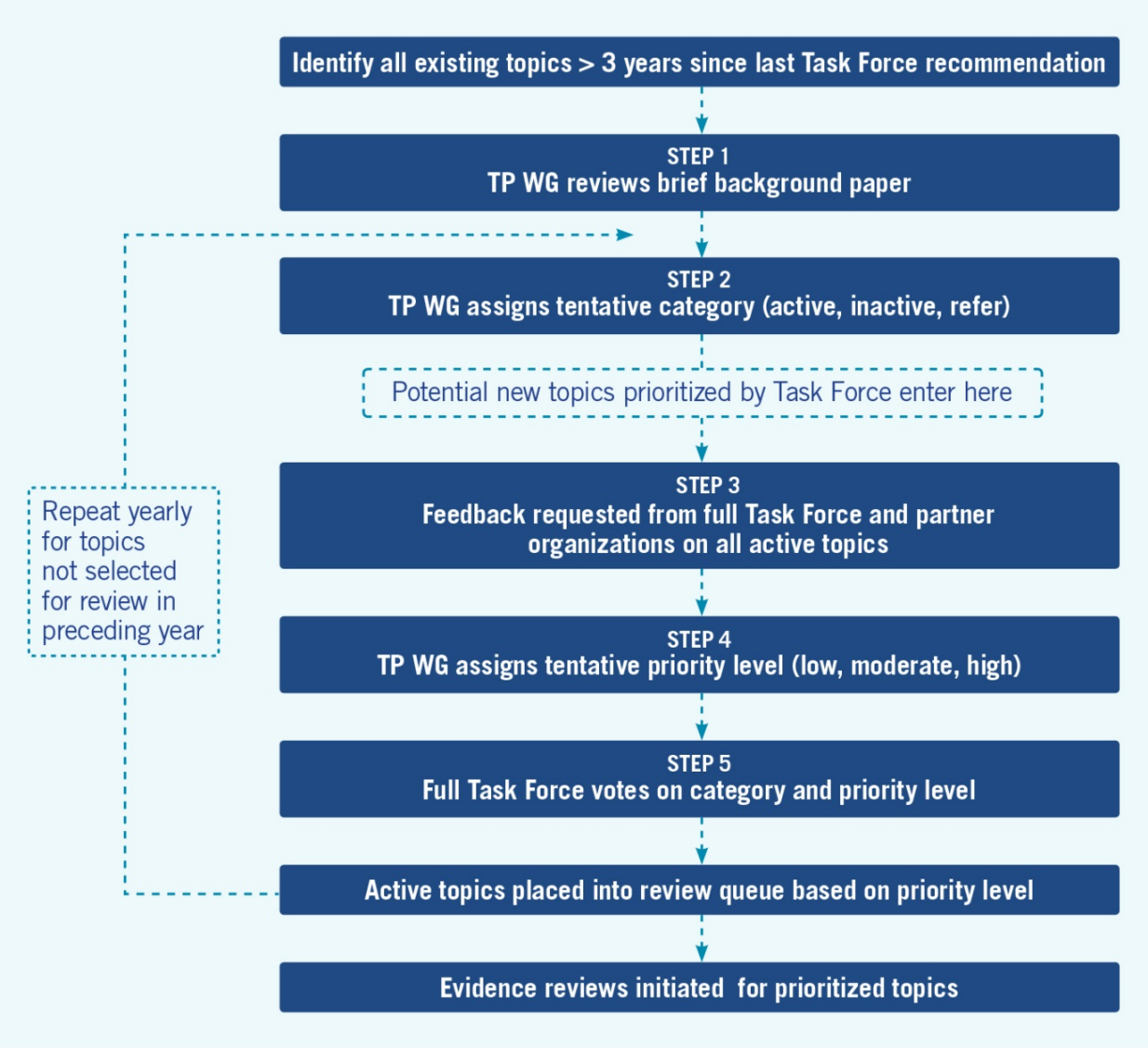
5. The Topic Prioritization Workgroup decides whether to proceed with a full Task Force discussion.
6. If the Topic Prioritization Workgroup decides to proceed, the summary is presented at a Task Force meeting for general discussion. The Task Force then votes on the decision to refer the topic to the specific organization.
7. A single summary paragraph is added to the USPSTF Web site that includes a link to the organization's recommendation.

The criteria for referring to another organization's recommendation are:

1. The organization has been identified by the Task Force 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

Referred topics may be re-activated through the usual new topic nomination process (described in Section 2.2).

Figure 3. Steps in Topic Prioritization



2.6 Consideration of an Early Topic Update

Occasionally a study will be published after a recommendation's release that may potentially affect the Task Force'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 Task Force by a number of sources, including the public, Task Force members, EPCs, professional organizations (including Task Force partners), and advocacy groups.

A regular audit of information sources is conducted to locate newly published research and/or guidelines that are relevant to topics in the Task Force portfolio. This LitWatch process is described 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 Topic Prioritization AHRQ staff member or another assigned Medical Officer completes a form with the following items:
    - Citation
    - Nominator and affiliation
    - Assigned Medical Officer
    - Brief summary/abstract of study
    - Number of criteria met (see below)
    - Recommendation of Medical Officer/Scientific 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)

    The Medical Officer proposes a disposition as to whether the new evidence should trigger an early review, based on the following criteria (order is not necessarily based on criteria weighting):
    - New evidence conflicts with current recommendation
    - Large-scale study may improve certainty of net benefit
    - New evidence has potential to change recommendation grade
    - Evidence focuses on a new intervention/strategy not previously considered
    - Study shows a change in magnitude of benefit or harm that might alter the Task Force's assessment of magnitude of net benefit
    - Evidence has the potential to fill a gap in the chain of indirect evidence
    - High level of existing controversy about the topic
    - High public health burden of the condition
    - High quality or relevance (e.g., a randomized, controlled trial [RCT] is published on a topic for which the current recommendation is based on observational evidence)
    - Published in a peer-reviewed journal
    - Study directly links the prevention strategy to the primary outcome of interest (i.e., direct evidence of health effect)
    - Study was identified by a reliable source (e.g., professional organization, Task Force member, advocacy group)
2. The form and the Medical Officer's recommendation are sent to the Scientific 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 Task Force 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 Task Force votes at its next meeting on whether the evidence update should be accelerated because of the new evidence. If the Task Force votes for an early topic update, the Task Force also assigns a priority level (high, moderate, low) based on the usual topic prioritization criteria.
6. If the Task Force decides to accelerate the update, the USPSTF Scientific Director at AHRQ places the topic in the review queue.
7. A brief notice from the Task Force Chair is sent to the nominator about the disposition.

Section 3. Topic Work Plan Development

When a topic is prioritized for review by the Task Force 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. This section applies to systematic reviews undertaken for a new topic or to update an existing topic. Work plan development for topic reaffirmation updates is described in Section 4.7.

A topic team is appointed for each prioritized topic before topic scoping begins and consists of Task Force leads (including one of the Task Force Chairs), at least one AHRQ Medical Officer, and the EPC review team. EPCs are scientific research centers tasked with conducting systematic evidence reviews that serve as the foundation for Task Force recommendations. Based on expertise and interest, several Task Force members are assigned to serve as leads for each topic. An AHRQ Medical Officer is assigned to oversee the topic and may be joined by the Task Force Scientific Director and/or Associate Scientific Director in overseeing that topic. A lead investigator is assigned by the EPC to lead the evidence review team.

Two integrated documents are developed during this phase of the systematic review: a work plan and a research plan. Both of these documents are revised and finalized through discussions with the Task Force leads and the AHRQ Medical Officer in an ongoing process that includes public comment on the research plan. The work plan is drafted by the EPC review team and captures the history, previous Task Force recommendations, and proposed approach to the topic. The purpose of the work plan is to establish the review perspective for the upcoming review. The template for the work plan is described below and in **Appendix V**.

Based on the draft work plan, a draft research plan that contains the analytic framework, key questions, and inclusion/exclusion criteria is created for public comment. After approval by the Task Force leads, the draft research plan is posted on the Task Force Web site for 4 weeks to allow public comment. All comments received during the public comment period are provided verbatim to the topic team, and the EPC review team summarizes major themes and makes suggested revisions to the research plan based on these comments.

The topic team discusses any major suggestions for revisions, the EPC review team incorporates final revisions into the research plan, and the Task Force leads approve the final research plan. For new topics, the work plan may be peer reviewed and presented to the entire Task Force at one of its regular meetings. Development of a work plan generally takes from 6 to 7 months, including public comment.

3.1 Determining Topic Scope and Review Approach

The Task Force has determined that using systematic reviews is the best method for organizing and evaluating the existing scientific evidence relevant to questions about a clinical preventive service. In order to answer the relevant questions about a clinical preventive service, the EPC review team usually undertakes a series of related systematic reviews to answer each of the key questions in the analytic framework.

3.1.1 Principles for Determining the Review Approach

During work plan development, the EPC review team considers the scope of the evidence needed for the Task Force to make its recommendation. For reviews undertaken to update existing Task Force recommendations, this process is based on:

1. Examination of the previous Task Force recommendation(s), including the populations and clinical preventive services addressed, to determine their fit with current questions about the clinical preventive service
2. Examination of the previous Task Force evidence review process for the topic and the review findings in order to identify established evidence, important review limitations, and evidence gaps
3. Determination of current contextual information (e.g., changes in understanding of the nature of the disease process or changes in diagnosis, therapeutics, or practice; controversy over any of these elements)

In order to facilitate the consistent development of the review approach across topics, the Task Force has developed a template to guide the development of the final work plan (**Appendix V**).

The work plan can be considered generally analogous to a protocol, such as those developed for an AHRQ Effective Health Care Program review or a Cochrane review. It is also an articulation of the rationale for the scope decisions made in framing the topic.

3.1.1.1 Primary Care Interventions Addressed by the Task Force

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)

The Task Force considers interventions that are delivered in primary care settings or are judged to be feasible for delivery in or referable from primary care. To be feasible in primary care, the intervention should target patients seeking care in primary care settings. Additionally, clinicians and/or related staff in the primary care setting should have (or could have) the skills necessary to deliver the intervention, or the intervention could be one generally ordered or initiated by a primary care clinician.

Task Force recommendations address primary or secondary preventive services. Primary preventive measures are those provided to persons in a clinical setting to prevent the onset of a targeted condition (e.g., aspirin for the prevention of colorectal cancer, counseling for a healthful diet), 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 (e.g., screening for colon cancer). Interventions that are part of the treatment and management of persons with clinical disease are usually considered tertiary prevention and are outside the scope of Task Force recommendations.

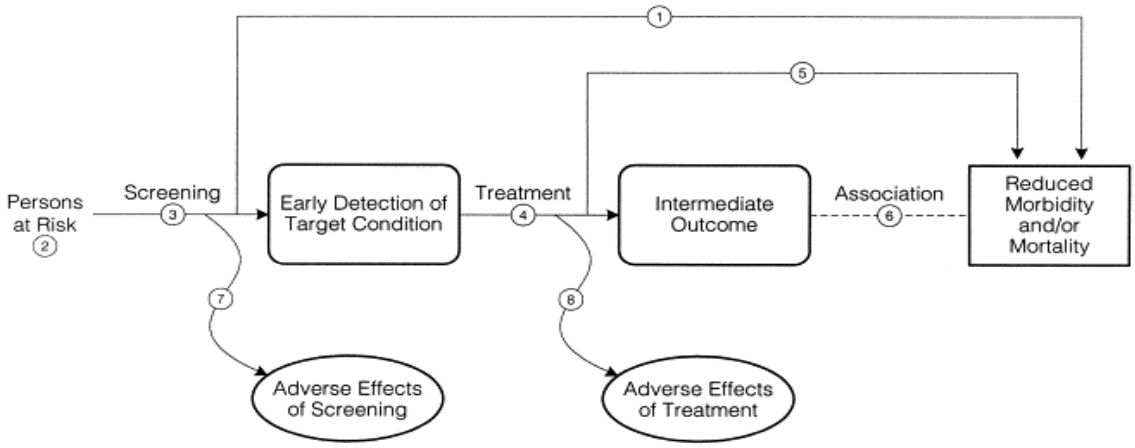
3.1.1.2 Incorporation of Subpopulation Considerations

The Task Force incorporates subpopulation-specific concerns when they may represent substantial heterogeneity in screening or preventive treatment effects. Data on the incidence/prevalence, complications, morbidity, and mortality of the condition of interest should be routinely summarized by age, race/ethnicity, sex, or other important topic-specific clinical characteristics. Additional details on the process for incorporating subpopulation considerations into systematic reviews for the Task Force and its evidence deliberations are under development and will be detailed in a future version of the Procedure Manual.

3.2 Methods Relevant to Work Plan Development

The work plan template (**Appendix V**) stimulates thinking and guides the systematic consideration of the factors that experience has shown are important in planning a review to update or issue a new Task Force recommendation. Since most reviews conducted for the Task Force are for updating previous TF recommendations, the work plan template was developed with that purpose in mind. However, the same template can be used to plan and guide the systematic review for a new topic; sections addressing the previous Task Force recommendation and previous review findings would not be included.

Figure 4 . Template of an Analytic Framework



3.2.1 Analytic Frameworks

The purpose of an analytic framework (**Figure 4**) is to clearly present in graphical format the specific questions that need to be answered by the literature review in order for the Task Force to evaluate the effectiveness and safety of the proposed preventive service. 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 1994 paper by Woolf et al.

3.2.1.1 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 4**).

The population appears at the left 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 "women at risk for cervical cancer." 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. Curved arrows lead to adverse effects of the action (i.e., "harms"). 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 the prostate-specific antigen test") 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. Overarching linkages generally represent studies in which the population is randomized to the clinical preventive service and health outcomes are measured. This is considered direct evidence. Each arrow is a "key question" that must be addressed by an evidence review. However, in the situation where there is robust direct evidence for the overarching linkage (such as multiple population-based screening trials), there may not be a need to address the multiple indirect evidence linkages through systematic review.

Outcomes are depicted using a rectangle; intermediate outcomes have rounded corners and health outcomes have squared corners. A health outcome that follows an intermediate outcome, which typically reflects the natural progression of disease (e.g., from "retinopathy" to "visual impairment"), is depicted by a dotted line (no arrowhead). Other important outcomes (e.g., societal/legal effects, non-disease benefits) can be included in an analytic framework if needed for the topic, and can be depicted as intermediate or health outcomes as defined below. Annotated footnotes are specific key questions that are associated with each linkage and that must be answered by the literature review. The key questions are written in detailed narrative below the analytic framework. Details required to interpret the key questions are further delineated in the inclusion/exclusion criteria for each review.

3.2.1.2 Analytic Frameworks Are Not Causal Pathways

Analytic frameworks as used by the Task Force are not intended to comprehensively depict all factors and variables that cause patients to receive the preventive service or are 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 the Papanicolaou test detects early disease and whether early detection reduces mortality) need not specify other covariables, such as the risk factors for cervical cancer, demographic characteristics of women who are more likely to be screened, etiological determinants of cervical cancer, or pathological progression of cervical cancer from the atypical cell stage to invasive disease.

Although the research plan is developed and established for an ongoing review, the components of an analytic framework are not static for a given topic, and may require revision for future reviews and recommendations as the scientific basis for the clinical preventive service advances and the current important clinical questions, populations, or outcomes change accordingly.

3.2.1.3 Analytic Frameworks Are Not Decision Trees, Clinical 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 analysis models, or to calculate their probabilities. Similarly, analytic frameworks do not guide clinical decision-making for an individual patient, nor do they depict every action in the sequence of services for a clinical preventive service. Instead, the analytic framework is a logic model of the minimal, sequential clinical assumptions that must be verified using empirical evidence in order to determine the net benefit of a preventive service.

3.2.1.4 Actions Versus Outcomes

Analytic frameworks used by the Task Force distinguish between actions (e.g., obtaining a screening test, treatment with a drug) and outcomes (e.g., detection of a disease, reduced morbidity and mortality, change in patients' behavior, adverse effects). The performance characteristics (e.g., sensitivity, specificity) of a screening test is not itself an outcome. Actions are depicted by arrows, whereas outcomes are depicted by rectangles (**Figure 4**).

3.2.1.5 Intermediate Outcomes Versus Health Outcomes

Analytic frameworks used by the Task Force distinguish between intermediate outcomes and health outcomes, and consider both beneficial and harmful outcomes (e.g., adverse effects of screening and treatment).

Health outcomes
Health outcomes are symptoms, functional levels, and conditions that patients can feel or experience and are defined by measures of physical or psychological well-being. A clinical "sign" is not a health outcome that is not sensed by the patient; a clinical sign is analogous to an abnormality on a blood test or radiologic exam (and therefore an intermediate outcome). Examples of health outcomes include visual impairment, pain or dyspnea, functional status, quality of life, impotence after prostatectomy, child development, and death.

Intermediate outcomes
Intermediate outcomes are outcomes that may be influenced by a preventive service but are not health outcomes in and of themselves. They are pathologic, physiological, psychological, social, or behavioral measures and other study endpoints related to a preventive intervention. Examples include blood pressure, serum cholesterol, vitamin levels, asymptomatic ductal carcinoma in situ (DCIS), asymptomatic carotid artery stenosis, weight, dietary intake , car crashes, improved educational achievement, reduced rate of psychiatric hospitalizations, and physical activity.
The USPSTF gives greater weight to evidence of an effect on health outcomes than 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.

At times the USPSTF may consider the evidence on societal (including caregiver) outcomes. The effect of an intervention may extend beyond the individual to society as a whole or to another individual. For example, reducing an individual's alcohol consumption decreases mortality related to car crashes not only for that person but also for others on the road. Screening for cognitive impairment may provide benefits to the caregiver beyond that to the individual. In addition, the USPSTF may consider outcomes that are not traditionally in the realm of health, such as educational attainment. These are not direct measures of health but are indicators of positive or negative effects on the larger society. When being considered in the context of an evidence review for the USPSTF, societal outcomes are represented in the analytic framework as an intermediate outcome or a health outcome, depending on the specific topic.

When data are available and relevant to decisions about the preventive service delivery, the Task Force considers data on both all-cause and cause-specific mortality in making its recommendations (go to Section 4.5 for discussion of these outcomes).

3.2.1.6 Revisions

Analytic frameworks can evolve with time and may appropriately differ when recommendations are updated because of changes in clinical questions or important uncertainties about the evidence. During the systematic review, it is sometimes necessary to revise an analytic framework to more clearly reflect the methods of the review. New key questions may be added when new interventions, outcomes, or logical arguments emerge during the course of the review. If these revisions only reflect improving the clear communication of the systematic review methods, they can be undertaken by the EPC review team. If there are any scientific ramifications to a potential analytic framework

revision, review and approval by the AHRQ Medical Officer and Task Force leads is expected. Such changes are reflected in the final review and manuscript resulting from the systematic review.

3.2.2 Key Questions

Key questions are an integral part of the approach to conducting systematic reviews the Task Force 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 searches, data abstraction, and analysis processes.

Key questions, in association with the analytic framework, establish the necessary steps in the clinical logic that must be demonstrated to evaluate the effectiveness and harms of a clinical preventive service in primary care. Key questions articulate the key aspects of the relevant populations, interventions, and outcomes—aspects that are essential in order to focus the review on a manageable and clinically relevant topic and to clearly communicate to readers what the review will address. In constructing key questions, the topic team must balance specificity of detail and readability; the detailed inclusion/exclusion criteria provide additional necessary details to understand how the key questions will be interpreted in the systematic review.

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 Task Force. In addition, there may be reason to focus on an overarching linkage (and the associated key question) in an analytic framework rather than the indirect linkages (and their associated key questions). All key questions are reviewed and approved by the Task Force leads and AHRQ Medical Officer in the process of assessing and refining the topic before the detailed literature review is conducted. Input is also obtained through public comment and from the full Task Force (for complex or new topics). Key questions addressed in a systematic review are listed in the Methods section and used to organize the results in the final review. Key questions are addressed using up-to-date systematic review methods, under the current guidance and methods of the Task Force. Each key question is addressed through a distinct literature search, if necessary, and reported separately in the Results section of the review.

Contextual questions represent issues in a review for which the Task Force needs a valid but not necessarily systematic summary of current research in order to provide the context for its deliberation and recommendation statement. Contextual questions may address a range of different types of informational needs, 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 which are part of the Task Force's considerations (e.g., screening interval, ages when screening should be stopped, newer technologies for screening and/or intervention); or 3) published modeling studies (when the Task Force has decided not to formally commission a modeling study). When formulating a work plan, issues in the background and introduction may emerge as candidates for formal contextual questions when the Task Force requires detailed and representative information to inform its consideration of the systematically reviewed evidence.

Although contextual questions are not necessarily addressed systematically, 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 published 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 the contextual questions is typically included as part of the Introduction or Discussion sections, and related as appropriate to the results of the systematic review.

3.3 Previous Task Force Review and Recommendations

To ensure that the current work plan builds coherently upon the Task Force's previous work on the topic, this part of the work plan succinctly summarizes the conceptual clinical framework and evidence foundation built by any previous USPSTF reviews and recommendation statements on the topic. The current Task Force recommendations are listed here verbatim, along with the analytic framework, key questions, 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 also listed.

3.4 Search for New Synthesized Evidence/Pending Studies

At the work plan development stage, the EPC librarian works with the 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). This is the first systematic search that will be incorporated into the overall searching done by the EPC for

the 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.

This synthesized evidence also provides background information that informs the approach to the topic and development of key questions. Background information that is typically collected from the synthesized evidence includes etiology and natural history, risk factors, screening strategies, interventions, current clinical practice, and prevalence and burden of disease/illness for the condition and for important subpopulations. Additional background information is the definition of "burden of suffering" of the condition in question. This burden is the ultimate target of implementing the preventive service. 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 (including societal or caregiver populations when relevant), is critical context for considering the potential population-level benefit of any clinical preventive service. The severity of the condition as measured by such metrics as prevalence and severity (e.g., number of life-years and quality-adjusted life-years lost in a population) is 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. The Task Force is also aware that implementation of various screening strategies can affect estimates of the burden of the disease, even in the absence of effective strategies, through lead, length, stage shift, and detection bias.

The following databases and Web sites are usually searched: Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database (United Kingdom), National Institute for Health and Care Excellence (United Kingdom), Institute of Medicine, PubMed® (using the systematic review search engine developed by the National Library of Medicine), and when appropriate, subject-specific databases (e.g., PsycINFO®). Searches are limited to literature published approximately 12 months prior to the last search of the previous review to the present.

In order to identify ongoing studies that could affect review scope and/or planning, the EPC librarian and/or topic team searches ClinicalTrials.gov, Current Controlled Trials (www.controlled-trials.com), Australian New Zealand Clinical Trials Registry (www.anzctr.org.au), and the World Health Organization's International Clinical Trials Registry (www.who.int/ictrp/en).

The EPC review team also checks to determine whether there is a finished, in-process, or planned Community Preventive Services 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 team.

3.5 Current Task Force Review Approach

3.5.1 Inclusion/Exclusion Criteria (Admissible Evidence)

The EPC review team, in consultation with the Task Force leads, clearly documents the criteria by which it will include evidence on a given key question. Such criteria might include study design (RCTs, cohort studies), 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 and key question basis, depending on the questions and the quality of the most applicable evidence anticipated being available. The goal is to identify the highest-quality evidence relevant to making an accurate determination of benefits and harms of delivering a preventive health service by primary care providers to persons living in the United States. All inclusion/exclusion criteria are posted for public comment, revised by the EPC, and approved by the Task Force leads.

One variable in the inclusion/exclusion criteria relates to the timeframe of the literature search. For a review to update a previous recommendation from the Task Force, the EPC review 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 search period covered by the previous review. If the EPC does not systematically re-review the evidence from a previous USPSTF review, it will synthesize and incorporate the results of the previous review into the current review in order to allow a comprehensive consideration of the evidence for a topic.

In addition, the EPC review team searches for other systematic reviews on the topic. If another systematic review is found that is rigorous and addresses the same key question, the topic team may choose to incorporate that review as appropriate rather than redoing all of the work already represented in a good-quality, existing systematic review.

3.5.2 Use of Topic Experts

By design, EPC review teams consist of generalist clinicians, researchers, methodologists, and staff with various levels of content expertise. When appropriate, the EPC review team engages content experts and specialists as consultants or co-investigators to advise about work plan formulation and operational decisions made during the conduct of the review. To allow continuity with the previous Task Force review, the EPC review team may intentionally engage previous review team members as consultants or members of the current review team. Conflict of interest considerations are taken into account when engaging all content experts and specialists.

3.6 Peer Review of Work Plan

The work plan for full systematic reviews for new topics is usually peer reviewed before it is finalized. Four to six peer reviewers are chosen to provide content expertise, specialty perspective, topical research experience, and relevant methodological or policy expertise as appropriate to the topic. Peer reviewer lists are drafted by the EPC review team and amended and approved by the AHRQ Medical Officer. The EPC review team 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 for final input and approval by the Task Force as a whole.

3.7 Public Review of Research Plan

Based on the full draft work plan, a draft research plan that contains the analytic framework, key questions, and inclusion/exclusion criteria is created for public comment (go to Section 9 for more detail on public comment processes). After approval by the Task Force leads, this document is posted on the Task Force Web site for 4 weeks to allow public comment and input on the research plan. All results from the comment period are provided verbatim to the topic team. The EPC review team summarizes major themes and makes suggested revisions based on these comments.

3.8 Task Force Approval of Final Research Plan

After the EPC review team incorporates revisions into the research plan, it is presented for final input and approval by the Task Force leads. The final approved research plan, including a section on Response to Public Comment, is posted on the Task Force Web site.

Section 4. Evidence Review Development

The evidence review development begins with finalization of topic scope, review approach, and research plan, as described above, and continues in the next stage with literature searches. The stages in the evidence review development are displayed in Figure 5.

4.1 Literature Retrieval and Review of Abstracts and Articles

4.1.1 Methods for Literature Searches

All literature searches are conducted using MEDLINE and the Cochrane Central Registry of Controlled Trials, using appropriate search terms to retrieve studies for all key questions that meet the inclusion/exclusion criteria for the topic. Other databases are included when indicated by the topic (e.g., PsycINFO for mental health topics). Searches are limited to articles published in the English language. For reviews to update recommendations, searches are conducted for literature published approximately 12 months prior to the last search of the previous review to the present. For new topics, the date range for the search is determined by the nature of the screening and treatment interventions for the topic, with longer time frames for well-established interventions that have not been the focus of recent research activity or topics with limited existing research and shorter time frames for topics with more recently developed interventions. The EPC review team supplements these searches with suggestions from experts and a review of reference lists from other relevant publications.

Search terms used for each key question, along with the yield associated with each term, are documented in an appendix of the final evidence review. A followup or “bridge” search to capture newly published data is conducted close to the time of completion of the draft evidence review, with the exact timing determined by the topic team.

4.1.2 Procedures for Abstract and Article Review

After literature searches are conducted, the EPC review team uses a two-stage process to determine whether identified literature is relevant to the key and contextual questions. This two-stage process is designed to minimize errors and to be efficient, transparent, and reproducible. First, titles and abstracts are reviewed independently by two reviewers by broadly applying a priori inclusion/exclusion criteria developed during the work plan stage of the review. When in doubt as to whether an article might meet the inclusion criteria, reviewers err on the side of inclusion so that an article is retrieved and can be reviewed in detail at the article stage. All citations are coded with an excluded or included code, which is managed in a database and used to guide the further literature review steps. Two reviewers then independently evaluate the full-text articles for all citations included at the title/abstract stage. Included articles receive codes to indicate the key question(s) for which they meet criteria and excluded articles are coded with the primary reason for exclusion, though additional reasons for exclusion may also apply.

4.1.3 Literature Database

For each systematic review, the EPC review team establishes a database of all articles located through searches and from other sources. The database is the source of the final literature flow diagram documenting the review process. Information captured in the database includes the source of the citation (e.g., search or outside source), whether the abstract was included or excluded, the key question(s) associated with each included abstract, whether the article was excluded (with primary reason 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.

4.2 Internal and External Validity Assessment of Individual Studies

By means of its analytic framework and key questions, the Task Force indicates what evidence is needed to make its recommendation. By setting explicit inclusion/exclusion criteria for the searches for each key question, the Task Force indicates what evidence it will consider admissible and applicable. 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 internal and external validity of individual studies is conducted by the EPC review team using the USPSTF criteria as a baseline and newer methods of quality assessment as appropriate (go to **Appendix VI** and **Appendix VII** for more detail on USPSTF criteria). Likewise, studies of interventions that require training or equipment not feasible in most primary care settings would be judged to have poor external validity and would not be admissible evidence.

4.2.1 Assessing Internal Validity (Quality) of Individual Studies

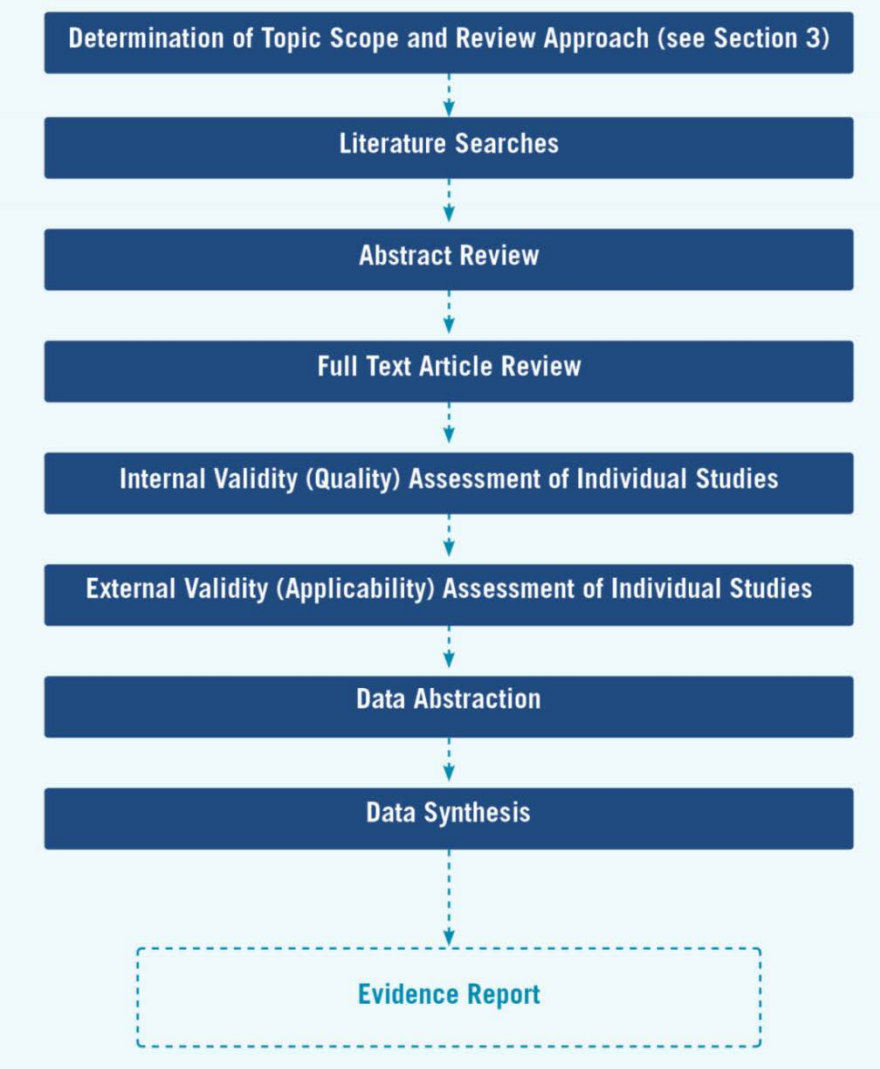
The Task Force recognizes that research design is an important component of the validity of the information in a study for the purpose of answering a key question. Although RCTs cannot answer all key questions, they are ideal for questions regarding benefits or harms of various interventions. Thus, for the key questions of benefits and harms, the Task Force currently uses the following hierarchy of research design:

I. Properly powered and conducted 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 analysis study
II-3. Multiple time-series, with or without the intervention; results from uncontrolled studies that yield results of large magnitude
III. Opinions of respected authorities, based on clinical experience; descriptive studies or case reports; reports of expert committees

Although research design is an important determinant of the quality of information provided by an individual study, the Task Force also recognizes that not all studies with the same research design have equal internal validity (quality).

To assess more carefully the internal validity of individual studies within research designs, the Task Force has developed design-specific criteria for assessing the internal validity of individual studies. The EPC may supplement these with the use of newer methods of assessing quality of individual studies as appropriate.

Figure 5. Stages of Evidence Review Development



These criteria (**Appendix VI)** provide general guidelines for categorizing studies into one of three internal validity categories: "good," "fair," and "poor." These specifications are not inflexible 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 at least one specified criterion, but has no known important limitation that could invalidate its results. "Poor" studies have at least one "fatal flaw" or multiple important limitations. A fatal flaw is due to 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 EPC, at its discretion, may include some poor-quality studies in its review. When studies of poor quality are included in the results of the systematic review, the EPC explains the reasons for inclusion, clearly identifies which studies are of poor quality, and states how poor-quality studies are analyzed with regard to good- and fair- quality studies. When poor-quality studies are excluded, the EPC identifies the reasons for exclusion in an appendix table.

4.2.2 Assessing External Validity (Applicability) of Individual Studies

Judgments about the external validity (applicability) of a study pertinent to a preventive intervention address three main questions:

1. Considering the *subjects* in the study, to what degree do the study's results predict the likely clinical results among asymptomatic persons 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 predict the likely clinical result in primary care practices in the United States?
3. Considering the *providers* who were a part of the study, to what degree do the study's results predict the likely clinical results among providers who would deliver the service in the U.S. primary care setting?

4.2.2.1 Criteria and Process

The criteria used to rate the external validity of individual studies according to the population, setting, and providers are described in detail in **Appendix VII**. As with internal validity, this assessment is usually conducted initially by the EPC review team, with input from Task Force members for critically important or borderline studies. This assessment is then used to answer the question, "If the study had been done with the usual U.S. primary care population, setting, and providers, what is the likelihood that the results would be different in a clinically important way?"

4.2.2.2 Population

Participants in a study may differ from persons receiving primary care in many ways. Such differences may include sex, ethnicity, age, comorbid conditions, 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 divergences between the study's results and what would be reasonably anticipated to occur in asymptomatic persons or those 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 inclusion/exclusion criteria that limit the study to persons most likely to benefit; other study features may affect the risk level of the subjects recruited to the study. The absolute benefit from a service is often greater for persons at increased risk than for those at lower risk.

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

4.2.2.3 Setting

When assessing the external validity of a study, factors related to the study setting should be considered in comparison with U.S. primary care settings. 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 study patients may not be attainable when the item must be purchased. Results obtained in a trial situation that ensures immediate access to care if a problem or complication occurs may not be replicated in a non-research setting, where the same safeguards cannot be ensured, and where, as a result, the risks of the intervention are greater. When considering the applicability of studies from international settings, the EPC often uses the United Nations Human Development Index to determine which settings might be most like the United States.

4.2.2.4 Providers

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 U.S. primary care. Studies may recruit providers selected for their experience or 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 Data Abstraction

Data is abstracted in abstraction forms or directly into evidence tables specific to each key question. Although the Task Force has no standard or generic abstraction form, the following broad categories are always abstracted from included articles:

- Study design
- Study period
- Inclusion/exclusion criteria
- Participant characteristics
- Participant recruitment setting and approach
- Number of participants who were recruited, randomized, received treatment, analyzed, and followed up
- Details of the intervention or screening test being studied
- Intervention setting
- Study results, with emphasis on health outcomes where appropriate
- Individual study quality information, including specific threats to validity

Information relevant to applicability is consistently abstracted (e.g., participant recruitment setting and approach, inclusion/exclusion criteria for the study). The EPC review team uses these general categories, and other categories if indicated, to develop an abstraction form or evidence table specific to the topic. For example, source of funding may be an important variable to abstract for some topics, and performance characteristics are abstracted for diagnostic accuracy studies.

The EPC review team abstracts only those articles that meet inclusion criteria. Abstractions are conducted by trained team members, and a second reviewer checks the abstracted data for accuracy, including data included in a summary table, a meta-analysis, or in calculations supporting a balance sheet/outcomes table. Initial reliability checks are done for quality control.

4.4 Data Synthesis

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

1. The individual study (discussed in Section 4.2)
2. The key question (discussed below)

The Task Force also assesses the adequacy of the evidence at the key question level (discussed in Section 6.2).

4.4.1 Quantitative Synthesis

When the evidence for a key question includes more than a few trials and there appears to be homogeneity in interventions and outcomes, meta-analysis is considered by the topic team. (Please see section 4.6 about how the EPC may incorporate published meta-analysis and systematic reviews into the Task Force review.) Meta-analysis provides the advantage of giving summary effect size estimates generated through a transparent process. The decision to pool evidence is based on the judgment that the included studies are clinically and methodologically similar, or that important heterogeneity among included trials can be addressed in the meta-analysis in some way, such as subgroup or sensitivity analyses. The EPC review team considers whether a pooled effect would be clinically meaningful and representative of the given set of studies. A pooled effect may be misleading if the trials clinically or methodologically differ to such a degree that the average does not represent any of the trials. Interpretations of pooled effect sizes should consider all sources of clinical and methodological heterogeneity. Similarly, the interpretation of pooled results takes into account the width of the confidence interval and the consequences of making an erroneous assessment, not simply statistical significance. Results of meta-analyses are usually presented in forest plot diagrams.

4.4.2 Qualitative Synthesis

If there are too few studies or data are too clinically or statistically heterogeneous for quantitative synthesis, the EPC review team qualitatively synthesizes the evidence in a narrative format, using summary tables to display differences between important study characteristics and outcomes across included studies for each key question.

4.4.3 Overall Summary of Evidence

The EPC review team provides an overall summary of the evidence by key question in table format (**Appendix XII**). The table includes the following domains:

- Key question
- Number of studies and observations for each study design
- Summary of findings (quantitative and qualitative findings for each important outcome, with some indication of its variability)
- Consistency/precision (the degree to which studies estimate the same type [benefit/harm] and magnitude of effect)
- Estimates of potential reporting bias (publication, selective outcome reporting, or selective analysis reporting bias)
- Overall study quality (combined summary of individual study-level quality assessments)
- Body of evidence limitations (qualitative descriptions of important limitations in body of evidence from what would have been desired to answer the overall key question)
- Applicability (descriptive assessment of how well the overall body of evidence would apply to the U.S. population based on settings, populations, and other intervention characteristics)
- Overall strength of evidence (brief explanatory text describing deficiencies in the evidence and stability of the findings)

Within key questions, it may be most informative to stratify the evidence by subpopulation or by type of intervention/comparison or outcome, depending on how the Task Force has conceptualized the questions for the particular topic. The EPC review team does not publish an actual grade for the strength of the evidence but rather synthesizes the issues in the bulleted list above for each key question to inform the Task Force's assessments of the adequacy of the evidence (Section 6).

4.5 Other Issues in Assessing Evidence at the Individual Study Level

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

The Task Force strongly prefers multiple large, well-conducted RCTs to adequately 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 other evidence is insufficient to determine benefits and/or harms, the Task Force encourages the research community to conduct large, well-designed and well-conducted RCTs.

Observational studies are often used to assess harms of preventive services. The Task Force also uses observational evidence to assess benefits. Multiple large, well-conducted observational studies with consistent results showing a large effect size that does not change markedly with adjustment for potential known confounders may be judged sufficient to determine the magnitude of benefit and harm of a preventive service. Also, large well-conducted observational studies often provide additional evidence even in situations when there are adequate RCTs. Ideally, RCTs provide evidence that an intervention can work (efficacy), and observational studies provide better understanding if these benefits exist across broader populations and settings.

4.5.2 Ecological Evidence

The Task Force rarely accepts ecological evidence alone as sufficient to recommend a preventive service. The Task Force is careful in its use of this type of evidence because substantial biases may be present. Ecological evidence is data that are not at the individual level but rather relate to the average exposure and average outcome within a population. Ecological 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 (i.e., cross-sectional ecological study); or 2) within a single population with changing exposure status over time (i.e., time-series ecological study). In either case, the potential for ecological fallacy is a major concern. Ecological fallacy is the bias or inference error that may occur because an association observed between variables at an aggregate level does not necessarily represent an association at an individual level. In addition, ecological data sets often do not include potential confounding factors; thus, one cannot directly assess the ability of these potential confounders to explain apparent associations. Finally, some ecological studies use data collected in ways that are not accurate or reliable.

The Task Force does not usually accept ecological evidence alone as adequate to establish the causal association of a preventive service and a health outcome because it is not possible to completely avoid the potential for making the ecological fallacy in these studies,. In some very unusual situations, ecological evidence may play the primary role in the Task Force's evidence review and subsequent recommendation (e.g. screening for cervical cancer) , but this is rare. The Task Force may use ecological evidence for background or to develop an understanding of the context for which the preventive service is being considered. In addition, a review of ecological evidence may be warranted when well-known ecological data are used as evidence by others to justify a recommendation for Task Force consideration. The Task Force only rarely considers ecological studies as part of its evidentiary assessment. These circumstances could include when evidence from other study designs is considered inadequate but high-quality ecological evidence, especially studies demonstrating a very large magnitude of benefit or harm, could add important information. When the Task Force critically appraises ecological studies for use to develop a recommendation, the following criteria are used to assess the quality of the studies: 1) the exposures, outcomes, and potential confounders are measured accurately and reliably; 2) known potential explanations and potential confounders are considered and adjusted for; 3) the populations are comparable; 4) the populations and interventions are relevant to a primary care population; and 5) multiple ecological studies are present that are consistent/coherent.

4.5.3 Mortality as Outcome: All-Cause Versus Disease-Specific Mortality

When available and relevant, the Task Force considers data on both all-cause and cause-specific mortality in making its recommendations, taking into account the real and methodological contributions to any discrepancies between apparent and true effect. When a condition is a common cause of mortality, all-cause mortality is the desirable health outcome measure. However, few preventive interventions have a measurable effect on all-cause mortality. When there is a discrepancy between the effect of the preventive intervention on all-cause and disease-specific mortality, this is important to recognize and explore.

Three situations can result in a discrepancy between the effect on disease-specific and all-cause mortality. First, when a preventive intervention increases deaths from causes other than the one targeted by the intervention, all-cause mortality may not decline, even when cause-specific mortality is reduced. This indicates a potential harm of the intervention for conditions 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 is small, the effect on all-cause mortality may be immeasurably small, even with very large sample sizes.

Third, when the preventive intervention is applied in a population with strong competing causes of mortality, the effect of the preventive intervention on all-cause mortality may be very small or absent, even though the intervention reduces cause-specific mortality. For example, preventing death due to hip fracture by implementing an intervention to decrease falls in 85-year-old 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.

Methodological issues can arise because of difficulties in the assignment of cause of death based on records. In the absence of detail about the circumstances of death, it may be attributed to a chronic condition known to exist at the time of death but which is not, in fact, the direct cause. Coding conventions for death certificates also result in deaths from some causes being attributed to chronic conditions routinely present at death. For example, it is conventional to assign cancer as the primary cause of death to persons with a mention of cancer on the death certificate. The result of these methodological issues is a biased estimate of cause-specific mortality when the data are obtained from death certificates, which may not reflect the true effect an intervention has on death from the targeted condition. Similar methodological issues may occur as a result of adjudication committees.

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.

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 net benefit. Thus, it often takes into consideration subgroup analyses of large studies or studies evaluating particular subgroups of interest. The Task Force examines the credibility of 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 were consistent within themselves; whether the subgroup analyses were prespecified; and whether the results are biologically plausible.

4.6 Incorporating Other Systematic Reviews in Task Force Reviews

Existing systematic reviews or meta-analyses that meet quality and relevance criteria can be incorporated into reviews done for the Task Force. Existing reviews can be used in a Task Force review in several ways: 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. Quality assessment of existing systematic reviews is a critical step and should address both the methods used to minimize bias as well as the transparency and completeness of reporting of review methods, individual study details, and results. The Task Force has specific criteria for critically appraising systematic reviews (**Appendix VI**). The EPC may supplement these criteria with newer methods of evaluating systematic reviews and meta-analyses. Relevance is considered at two levels: "Is the review or meta-analysis relevant to one or more of the Task Force key questions for this review?" and "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 or requires a bridge search.

4.7 Reaffirmation Evidence Update Process

Reaffirmed topics are topics kept current by the Task Force because the topic is within the Task Force's scope and 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 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 justify a change in the grade of the recommendation. Only recommendations with a current grade of A or D are considered for reaffirmation. The goal of this process is to reaffirm the previous recommendation. Therefore, the goal of the search for evidence in a reaffirmation evidence update is to find new and substantial evidence sufficient enough to change the recommendation.

1. The topic may be identified for a reaffirmation evidence update by the Topic Prioritization Workgroup and approved for a reaffirmation evidence update by the entire Task Force following the usual process for prioritization, including the annual request for feedback from USPSTF members, partner organizations, and stakeholders. Several Task Force members (one as the primary lead) are identified to take the lead on the topic and serve in the same lead role as on other topic teams, as described in **Appendix IV**.
2. The topic team (review team, AHRQ Medical Officer, and Task Force leads) reviews the previous recommendation statement, evidence review, and background document prepared for topic prioritization. The topic team confirms that the topic is appropriate for a reaffirmation evidence update and then further defines the scope of the literature search. The literature search scope is limited to key questions in the evidence review for the previous recommendation. If there is a need for additional key questions or other expansion beyond the original scope, the topic is referred back to the Topic Prioritization Workgroup for consideration for a systematic review. Any other concerns about whether the topic is appropriate for a reaffirmation evidence update are referred for discussion by the Topic Prioritization Workgroup.
3. The topic team consults experts in the field to identify important evidence published since the last evidence review.
4. The topic team identifies recommendations from other Federal agencies and professional organizations.
5. The topic team performs literature searches in PubMed and the Cochrane Central Registry of Controlled Trials on benefits and harms of the preventive service. Other databases are included when indicated by the topic (e.g., PsycINFO for mental health topics). The benefits and harms to be reviewed are predefined through consultation with the Topic Prioritization Workgroup. In general, the literature search uses the MeSH terms from the previous evidence review, searches for studies published since the last review (at least 3 months prior to the end date of the previous search), is limited to the English language, is limited to humans, and is limited to the journals in the abridged Index Medicus (i.e., the 120 "core clinical journals" in PubMed). These limits may be expanded or modified as needed. For the literature search on benefits, the search is limited to meta-analyses, systematic reviews, and controlled trials; for harms, the search includes meta-analyses, systematic reviews, controlled trials, cohort studies, case-control studies, and large case series. Additionally, the reference lists of major review articles or important studies are reviewed for potential studies to include.
6. The topic team defines the inclusion/exclusion criteria. Limits on the size or duration of studies may be used as exclusion criteria. Two independent reviewers select studies to be included based on consensus on whether they meet the inclusion/exclusion criteria. A third reviewer is consulted if consensus is not reached among the two reviewers.
7. If substantial new evidence is identified or if the topic team discovers that the evidence base for the prior review may not support a reaffirmation evidence update, the issue is discussed with the Topic Prioritization Workgroup, who decides whether the topic should be addressed with a systematic review.

8. The topic team prepares a summary of the findings of the evidence update. The format of this document depends on whether the summary will be submitted to a journal for publication.
9. The results of the evidence update, expert discussion, and draft recommendation statement are presented to the leads and then to the entire Task Force for approval.

Section 5. Modeling

In addition to systematic reviews, the Task Force uses modeling to inform the recommendation process. The Task Force uses modeling only when there is direct evidence of the benefit of a preventive service on health outcomes or when there is evidence for each of the linkages in the analytic framework. While the latter situation is considered indirect evidence, it is sufficient basis for using modeling, subject to the considerations described below.

Topics meeting the direct evidence criterion and for which modeling has previously been applied include cervical cancer screening (very strong ecological evidence of mortality benefit), colorectal cancer screening, lung cancer screening, and breast cancer screening (RCT evidence of screening and mortality benefit). Other examples of USPSTF recommendations based on direct evidence are limited, but include topics such as abdominal aortic aneurysm screening and breast cancer chemoprevention. Candidates for decision modeling are generally A, B, and some C recommendations. While models may integrate sufficient evidence across an analytic framework, they should not be used to bridge a gap in the analytic framework where evidence is insufficient by using assumptions or unreliable data in order to provide the basis for a USPSTF recommendation. However, decision models can be critical to the USPSTF when there is insufficient empirical evidence to determine when to start and how long to continue delivering a clinical preventive service, how frequently to repeat the service, and the best or equally appropriate choices among different ways of delivering the service. Thus, decision modeling is primarily warranted when: 1) there are outstanding clinical questions about how best to target the clinical preventive service at the individual and the program level; and 2) it is highly unlikely that the systematic review can confidently determine magnitude of net benefit, particularly as it varies among important prespecified subpopulations.

Models used by the Task Force include simple calculations using probabilities for each sequential step from screening to health outcomes (historically known as “outcomes tables”) and more formal decision models. The purpose of either of these is to estimate the population impact of a given preventive service.

Outcomes tables derived from systematic reviews apply evidence from the systematic review alone to estimate the number of persons in a hypothetical population who would be affected in specific ways from implementation of the preventive service over a given time horizon (often 5 to 10 years). Further details about systematic review–derived outcomes tables are provided in **Appendix VIII**. A decision model is a more formal exercise using complex mathematical simulation to project the health outcomes that result from alternative interventions for screening, diagnosis, prevention, and treatment. Decision models incorporate evidence about natural history, disease-specific epidemiology, and other topic-related factors in addition to evidence from the systematic review.

There are often limitations in the evidence on important aspects of clinical preventive services. Critical aspects such as starting and stopping ages or preventive service delivery intervals are almost never directly addressed in trials or studies. How to define the best (or comparable) options among different approaches to preventive service delivery, particularly in the context of a screening program, may also have limited empirical evidence. There is often limited evidence to determine when a new or different screening test can be substituted for one that was initially recommended based on RCT evidence; available evidence on new screening tests is usually about test performance characteristics. When necessary, a decision model can link natural history, population characteristics, and screening or treatment effectiveness to estimate the relative impact and comparative effectiveness of varying screening or treatment approaches. As the USPSTF methods on modeling evolve details will be published and made available on the Task Force Methods Webpage.

Similarly, the usual information available through systematic review may not adequately summarize net benefit with the level of precision required for the USPSTF recommendation process. This is particularly likely when the clinical preventive service is one characterized by multiple benefits and harms, such as when the benefits and harms vary substantially between subpopulations; when there is moderate certainty of net benefit, but the magnitude is difficult to determine; or when the empirical evidence provides an estimate based on intermediate outcomes or using a limited time horizon that is clearly insufficient to determine net health benefits.

Inputs into decision models include best estimates from evidence on the linkages between screening and outcomes of interest and explicit assumptions about natural history, epidemiology, and other topic-related factors. It is important that modelers make their assumptions explicit so that Task Force members can assess the applicability of the model and its results to Task Force deliberations.

While acknowledging the demonstrated value of decision modeling for selected USPSTF recommendations, the Task Force also recognizes that there are opportunity costs and resource implications.

To maintain consistency and transparency across topics, the Task Force has developed a framework and criteria for determining whether decision modeling will be added to topics that could benefit from decision modeling along with

the systematic review. This framework is based on the following six sequential questions that the Task Force considers, to the extent possible, during the scoping phase of an eligible topic's systematic review.

1. *Has benefit for this clinical preventive service been established?*

Based on past experience and its evidence-based philosophy, the USPSTF intends to primarily apply decision modeling in cases where the benefit of the clinical preventive service is satisfactorily established, yet decisionmakers could benefit from further specification. As stated above, for the USPSTF, this implies that there is either: 1) direct evidence of a health benefit from a randomized (or otherwise very robust) comparison of delivery of the preventive service versus not delivering this service; or 2) evidence of benefit that can be achieved through strong linkages between multiple separate bodies of evidence.

2. *Which of the primary reasons for adding decision modeling are important to address for this clinical preventive service?*

In planning for decision modeling, the USPSTF specifies one or more of these primary reasons to add decision modeling to the systematic review in its recommendation development: 1) to assess appropriate starting or stopping ages or compare alternative intervals of preventive service delivery; 2) to compare alternative technologies, such as different screening tests; 3) to assess the impact of a newly developed substitute test in an established screening program; 4) to quantify net benefit more precisely or specifically than can be done based on systematic evidence review alone; 5) to extend time horizons beyond those available from studies; 6) to assess net benefit for population groups that represent combinations of higher and lower risk for benefits and/or harms; and 7) to assess net benefit stratified by sex or other demographic characteristics, such as race or ethnicity.

3. *Is the information gained from modeling or reviewing existing models likely to be worth the opportunity cost of modeling*?

This question represents a judgment and offers a potential stopping point in the process of considering the addition of decision modeling to a topic. Clearly articulating the expected advantages of using a decision model can enhance transparency and consistency and focus efforts on the most important topics for potential decision modeling, given that resource constraints are likely to limit its incorporation in all instances where it may be desired. Deciding to not commission a decision model at this stage can also reduce effort that would be required to complete steps 4, 5, and 6 below.

The Task Force is continuing to determine a full set of clear criteria, using its past experience with decision modeling. Until that work is complete, the following considerations may help to determine whether to commission a decision model.

1. When decision modeling will primarily assist in specifying screening program components (i.e., stopping and starting ages, intervals, comparative modalities or programs) for an effective clinical preventive service, the usual criteria of public health burden will apply. In addition, the Task Force may want to consider the potential for enhancing benefit through extending the preventive service to unstudied populations, reducing harms, or enhancing efficiency through more targeted service delivery.
2. When decision modeling is primarily intended to assist in net benefit or subpopulation determinations, USPSTF members offer the following questions to help project expected gains from incorporating modeling:
    a. Would the model provide important information for addressing uncertainty when making a recommendation for this preventive service?
    b. If one could predict the ultimate recommendation based on the systematic review evidence, would the results of a decision model inform one's deliberation on a grade (e.g., from an I statement to a letter grade) or change the messaging (i.e., to enhance public understanding of the USPSTF recommendation, especially if it differs from those of other organizations or community practice)?
    c. Would modeling enhance appropriate uptake of a recommendation by reducing under- or over-utilization in a subpopulation?
    d. Would modeling enhance appropriate uptake of a recommendation by providing clarity for measurement within quality improvement activities?

4. *Can the desired modeling approach be clearly outlined at this point, or is it contingent on additional information not known at the outset of the systematic review*?

There is tension between the need to commence modeling as early as possible to coordinate the modeling and systematic review work, and wasting effort when it is too early in the process to determine an appropriate focus for the decision model. The USPSTF must determine when it is possible in the overall process to define the decision problem/objective, the decision-important modeling outputs, and any approaches to be compared through modeling.

The “Decision Problem/Objective” is a template approach (**Appendix XIII**) that outlines the modeling objectives for the purposes of the recommendation statement, using the PICOTS (population, intervention, comparator, outcome, timing, and setting) framework to specify the scenarios to be modeled, required outcome measures, and the relevant modeling time horizon. If the main constructs of relevance cannot be defined, it is premature to proceed with decision modeling. Instead, if possible, the USPSTF delineates what modeling-relevant information is still uncertain and at what point in the systematic review process it will likely become clearer, in order to plan when to reconsider if a decision model is appropriate.

5. What is the decision problem/objective to be addressed through decision modeling?

Taking into consideration the prior USPSTF recommendation statement (and its evidence review approach), public response to the previous recommendation, current issues for clinical practice, and the expected state of the science, the USPSTF defines the parameters of the decision framework relevant to this topic, including: the rationale for decision modeling, the desired approach (e.g., type of model, one model or comparative models), scenarios to be modeled, populations and settings to be targeted and/or compared, desired outcome measures, and modeling time horizon needed. When searching for existing decision models, recency and setting should be considered.

6. What is the most expedient approach for needed decision modeling?

The most efficient mechanism for considering decision models alongside systematic reviews would be to use an already published analysis. While this is an attractive concept, it is uncommon that an existing decision model will match the decision problem outlined by the USPSTF well enough to be of use to the USPSTF. With more defined USPSTF guidance, however, opportunities for using existing models may increase. Further, iterative exchange between modelers and decisionmakers has been proposed as a critical element of the process by those with prior experience. Nonetheless, identifying existing models may be important if only to identify existing modeling groups and candidate modelers for collaboration.

Searching for existing decision models is most efficient if it is focused on the specific decision problem outlined by the USPSTF. Unlike a systematic review search, which is highly sensitive to avoid missing any relevant primary research, searches for existing models should emphasize specificity, since the purpose is to find one or a few excellent candidates for use or adaptation. The proposed process is outlined in **Appendix XIV**.

These six questions frame the considerations around adding a decision model to the systematic review for a specific topic, but do not make that decision. In the case of multiple competing priorities and resource constraints, the USPSTF Chairs make the final decision about whether to add decision modeling to a specific topic. The decision of the USPSTF Chairs is based on the findings from these six questions and their judgment about relative needs and priorities, considering the overall USPSTF portfolio.

Section 6. Methods for Arriving at a Recommendation

6.1 Overview

The preceding sections have described the processes for systematic evidence reviews and related reports that serve as the foundation for Task Force recommendations. This section briefly describes the specific work of the Task Force in examining and judging the cumulative evidence and making recommendations. Further details about the Task Force's methods can be found in a series of published articles (Section 10).

The Task Force's steps to arrive at a recommendation include:

1. Assessing the adequacy of evidence at the key question level
2. Assessing the adequacy of evidence at the linkage level
3. Estimating the magnitude of benefit and harm of the preventive service
4. Evaluating the certainty of the evidence for net benefit for the preventive service
5. Estimating the magnitude of the net benefit of the preventive service
6. Developing a recommendation grade for the preventive service in the relevant population, based on the above parameters

Once the admissible evidence has been gathered and the internal and external validity of individual studies has been assessed by the EPC review team, the Task Force evaluates the adequacy of evidence that the studies provide toward answering the key questions and addressing the linkages in the analytic framework. The evidence addressing a linkage in an analytic framework may come from more than one key question. For example, to assess the benefit of an intervention (linkage), key questions may need to be answered about the 1) effect of different intervention types (e.g., behavioral counseling, pharmacotherapy), 2) effect of intervention timing, and 3) health outcomes in different subpopulations. The USPSTF often uses a tool structured by key question to help facilitate its discussions and organize its assessment of the adequacy of the evidence and the certainty and magnitude of the net benefit (**Appendix XI**).

Direct evidence linking clinical preventive services to outcomes is often inadequate either because of a lack of direct evidence or because of limitations in the direct evidence that is available. Thus, the Task Force often needs to use indirect evidence in making its recommendations. This requires assessing the evidence related to the linkages in the chain of evidence between the preventive service and outcomes. For example, the linkage in the chain of evidence for screening usually includes key questions about the accuracy of screening tests, the efficacy and harms of early treatment, and the association between changes in intermediate measures due to specific interventions and changes in health outcomes.

The Task Force's process for determining the adequacy of evidence for a key question or across a linkage in the analytic framework involves answering six critical appraisal questions about the admissible evidence (**Table 2**).

Table 2. Factors Considered for Evaluating Adequacy of Evidence for Key Questions (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 sufficient quality (i.e., what is the internal validity)?
3. To what extent are the results of the studies generalizable to the general U.S. primary care population of interest to the intervention and situation (i.e., what is the applicability)?
4. How many and how large are the studies that address the key question(s)? Are the results precise?
5. How consistent are the results of the studies?
6. Are there additional factors that assist us in drawing conclusions (e.g., fit within a biological model)?

6.2 Assessing Evidence at the Key Question and Linkage Levels

6.2.1 Ratings of Adequacy for Key Questions

The Task Force rates the body of evidence for each key question as convincing, adequate, or inadequate.

6.2.1.1 Internal and External Validity

In making a determination of adequacy, the Task Force considers the aggregate internal and external validity of all studies across each of the key questions. The determination is based on a careful consideration of the studies that are judged as fair or good quality, using criteria based on each type of study design, in a body of evidence. The EPC

develops its evidence appraisal for each key question based only on studies with strong internal and external validity by including studies that are potentially relevant to the U.S. primary care population, settings, and providers (**Appendix VII**). The EPC displays factors related to applicability in its summary of evidence table for use by the USPSTF.

6.2.1.2 Other Factors to Consider

The Task Force also considers other important factors in addition to internal and external validity to judge the adequacy of evidence for each key question: the number and heterogeneity (statistical and/or clinical) of studies for each key question, the consistency and precision of reported outcomes, and other factors that appear to strengthen inferences about causal relationships.

6.2.1.3 Criteria for Ratings

Evidence for a key question may be deemed "convincing" when there are sufficient well-conducted studies of appropriate design that demonstrate consistent and precise results focused on outcomes and generalizable to the intended U.S. primary care population and setting. The consistency of or a large number of individual studies and lack of heterogeneity in pooled results strengthens the case for the evidence to be deemed convincing. The Task Force assesses the consistency of the evidence addressing a key question by examining the degree to which studies demonstrate similar directionality and magnitude of results. When statistical or clinical heterogeneity is present, the Task Force evaluates the comparability of studies with regard to study design, patients/subjects, interventions, comparators, settings, and outcomes to determine the coherence of evidence.

Evidence for a key question may be deemed as "adequate" when the evidence is sufficient to answer a key question, but is less convincing because of one of more significant limitations in factors, such as the appropriateness of study design, quality of studies, applicability of results, overall precision, and/or heterogeneity of evidence.

Evidence for a key question may be deemed as "inadequate" when evidence is insufficient to answer a key question because of a complete lack of evidence or a fatal flaw in one or more of the following factors: consistency of results, precision, applicability, and/or study quality and design. Inadequate evidence (for either benefits or harms) may create a critical gap in the evidence chain.

6.3 Dealing With Intermediate, Secondary, and Composite Outcomes

6.3.1 Intermediate Outcomes

The need to evaluate the evidence for the key question linking intermediate to health outcomes (Key Question (KQ) 6 in Figure 4) in order to make a recommendation will depend on the overall body of evidence. The Task Force may consider the evidence linking intermediate and health outcomes (KQ 6) when there is inadequate direct evidence (KQ1) and inadequate evidence linking earlier treatment to health outcomes (KQ5). The process for determining the adequacy of the evidence for the key question (KQ6) on the link between intermediate and health outcomes is similar to the process for other key questions, as described above in sections 6.2 and 6.4, but additionally requires special considerations, as discussed below. These considerations are necessary, given the unique characteristics of relationships between intermediate and health outcomes and because of the added potential threat to the certainty of the evidence that this additional link creates.

As discussed in 6.2 and 6.4, the adequacy of the evidence depends on the availability of well-conducted studies of appropriate design that demonstrate consistent and precise results focused on outcomes that are generalizable to the intended U.S. primary care population and setting. The consistency of, or a large number of, individual studies using the same intermediate and health outcomes strengthens the case for the evidence to be deemed convincing. As with other types of key questions and linkages, the Task Force assesses the consistency of the evidence by examining the degree to which studies demonstrate similar directionality and magnitude of results.

The ultimate goal for assessing the evidence for this key question is to determine precisely a consistent relationship between the direction and magnitude of change in an intermediate outcome with a predictable resultant direction and magnitude of change in the health outcomes. Acceptable evidence may come from post-trial data that follows subjects over time to determine the effect of an intervention or treatment on intermediate and health outcomes. When trial data are not available to answer this key question, the Task Force may consider observational evidence that provides epidemiological support for causation. In addition, cohort studies that follow subjects over time and report changes in intermediate and health outcomes may be considered. Added strength may come from evidence that shows similar magnitude of effect on multiple intermediate and health outcomes. When statistical, methodological, or clinical heterogeneity is present in the body of evidence for this link, the Task Force evaluates the comparability of

studies with regard to study design, patients/subjects, interventions, comparators, settings, and outcomes to determine the coherence of evidence. This is of particular importance in this key question.

The TF will exercise great caution when making a recommendation that depends in large part on the evidence linking intermediate and health outcomes. Due to the inherent limitations of the evidence that will be used to link intermediate and health outcomes, it is very unlikely that the evidence for this key question will be deemed convincing. This is due to the likely need to depend on observational evidence and the high potential for confounding. It is important to emphasize that strong associations between intermediate and health outcomes that are based only on cross-sectional studies would likely not be deemed adequate evidence for this key question. Evidence for this key question may be deemed as "adequate" (but not convincing) when the evidence is sufficient to answer the key question, but is less convincing because of one or more significant limitations, such as the appropriateness of study design, quality of studies, number and size of studies, applicability of results, overall precision, and/or heterogeneity of evidence (as discussed above).

6.3.2 Dealing With Secondary and Composite Outcomes

The Task Force adopted a policy of critically appraising all the endpoints (outcomes) of trials in a similar manner, following the six critical appraisal questions in **Table 2**. 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 three 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?

6.4 Determining the Adequacy for Benefits and Harms Linkages

After assessing the adequacy of the evidence at the key question level, the USPSTF assesses the adequacy of the evidence across the linkages. A linkage is represented by an arrow or a combination of arrows in the analytic framework that links the population on the left side of the framework to the beneficial or harmful health outcomes on the right side of the framework. Each linkage may be evaluated by one or more key questions. For example, in a cancer screening recommendation, the "benefit linkage" connects the population of asymptomatic adults without known cancer to the potential benefit of a reduction in mortality. This linkage could be evaluated by the following key questions, in combination: 1) Does screening reduce mortality? 2) Does early treatment of asymptomatic adults reduce early markers? 3) Does early treatment of asymptomatic adults reduce mortality? and 5) Do changes in these early markers lead to improvements in health outcomes and/or reduced mortality? Evidence from key questions on risk assessment and/or detection may also inform the assessment of the evidence on the benefit (or harm) linkage. The evidence on an overall linkage is classified into one of three categories: convincing, adequate, or inadequate.

Adequacy of the linkage for benefits or harms is not a simple summation of the adequacy for the key questions, but is determined by consideration of the six critical appraisal questions (**Table 2**), the coherence across all the evidence for the linkage, and other considerations, as described below.

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 benefit of a preventive service in an asymptomatic U.S. primary care population. It includes an assessment of the concordance between populations, interventions, and outcomes in the studies reviewed. Several studies of a preventive service may find different results (and thus be inconsistent), but the results may still be understandable (and thus coherent) in terms of the populations studied or the interventions used.

As part of coherence, USPSTF members assess the applicability of the body of evidence to populations, situations, providers, and settings as one of the components of the overall linkage. Judgment about applicability considers these factors but also involves synthesis of the evidence from the individual studies across the key questions and for the overall body of evidence for a linkage. The goal of the assessment is to judge whether there are likely to be clinically important differences between the observed body of evidence and the results expected when the intervention is implemented among asymptomatic patients in U.S. primary care settings, populations, and providers.

The following concepts are used to understand the applicability of the body of evidence to preventive interventions in an asymptomatic, primary care population in the United States:

- Inferences from the evidence that the intervention has effectiveness for U.S. primary care populations, situations, and providers

- The magnitude of benefit or harm observed in individual studies that comprise the body of evidence compared with the expected magnitude in the U.S. primary care setting
- The applicability of the information on benefits and harms in individual studies that comprise the body of evidence to understanding the expected benefits and harms in U.S. primary care settings and populations
- Information on the acceptability, feasibility, and availability of the studied intervention in U.S. primary care populations and settings
- Biological plausibility and clinical relevance of extrapolating from the body of evidence to large populations of asymptomatic persons in a primary care setting

Extrapolation is used by the Task Force to make inferences across the analytic framework to complete a chain of evidence connecting the intervention with health benefits. The Task Force extrapolates from existing evidence only when the case for doing so is strong. 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 epidemiological evidence, logic and biological plausibility. Two factors—logic and biological plausibility—play the greatest role in the decision about extrapolation.

When extrapolation is used to reach a letter grade recommendation, the scientific rationale for the recommendations and the methods used to review and judge the evidence are explicitly stated along with the recommendations. Examples of such consideration include whether the Task Force can: 1) extrapolate evidence about intermediate outcomes to health-related outcomes, 2) infer long-term health outcomes based on shorter-term outcomes, and 3) infer the effects of population-based screening based on RCTs of treatment of selected patients identified through case-finding.

6.4.1 Other Considerations for Determining the Adequacy of Evidence for Linkage: Benefits

The adequacy of the evidence for benefits may be further limited or strengthened by the following considerations:

- The availability of evidence on the effectiveness of early treatment compared with later treatment (or when the condition is asymptomatic vs. clinically apparent) of the subtype of the condition that would cause health problems
- Evidence on the prevalence or natural history of the target condition, or for heterogeneous conditions, evidence on the prevalence of the subtype of the condition that would cause important health problems
- For screening, the sensitivity and expected positive predictive value of the screening test (i.e., the degree to which the test will detect the subtype of the condition that would potentially cause health problems)
- Evidence showing a statistically significant effect but limited or unknown clinical importance
- Lack of evidence on important health outcomes
- Applicability of included studies to the screen-detected, asymptomatic, primary care, U.S. population
- Inability to ascertain the precise combination of factors and the risk threshold necessary to identify the target population for the preventive intervention; this happens when recommendations are made for high-risk populations that are identified largely based on one or more risk factors for a condition, but no multivariable risk prediction tools are available
- Evidence on well-established, evidence-based benefits for key questions (often addressed in previous USPSTF reviews) (i.e., “foundational evidence”)

6.4.2 Other Considerations for Determining the Adequacy of Evidence for Linkage: Harms

The adequacy of the evidence for harms may be further strengthened or limited by the following considerations:

- Information on the severity of harms, including patients’ perspectives on the acceptability of specific harms
- What is known about the number of false-positives, the invasiveness of the diagnostic workup, and the expected amount of overdiagnosis and overtreatment
- Prevalence and severity of disease
- Timing between screening test and confirmatory/diagnostic testing and/or treatment
- Evidence on well-established, evidence-based harms for key questions (often addressed in previous USPSTF reviews)

After careful deliberation, the Task Force determines how all the evidence and considerations are coherent; that is, it “make sense” in assessing if there is enough information to determine an overall benefit or harm and whether the Task Force can determine its magnitude (see below for more information about assessing magnitude). Therefore, the overall benefit linkage can be determined to be inadequate by the Task Force even though there is adequate evidence for the individual key questions related to benefits because of the limitations discussed above. One example

is screening for cognitive impairment, in which there was adequate evidence that some interventions resulted in improvements in some measures of beneficial outcomes, but the overall evidence for the benefit linkage was determined to be inadequate because 1) there was limited evidence that these changes resulted in clinically significant benefits, and 2) there was a lack of data on important outcomes (i.e., decision-making by patients and their families). These situations can occur because coherence, based on the six critical appraisal questions (**Table 2**) and other considerations, strengthens or limits the adequacy of the evidence above or below the actual evidence for individual key questions. How the Task Force uses the impact of these considerations on evidence adequacy to "bound" the magnitude of benefits or harms is discussed below.

6.4.3 Reaffirmations

The USPSTF may reaffirm a previous grade A or D recommendation. These are well established, current, evidence-based practices in primary care for which only a very high level of evidence would justify a change in the grade of the recommendation (e.g., screening for hypertension). In determining the adequacy of evidence for the benefits and harms linkages, the USPSTF considers whether the new evidence is of sufficient strength and quality to change its previous assessment of the certainty of the evidence. If the USPSTF does not find evidence of sufficient strength and quality to change its previous assessment, the USPSTF may vote to reaffirm the previous recommendation, using its usual voting procedure. A draft reaffirmation statement is prepared for consideration that includes a summary statement of the recommendation and evidence, the rationale, updated clinical considerations, and a brief summary of the systematic review or evidence update, with references to both the current evidence update and the previous systematic review. The draft reaffirmation statement is posted for public comment following the usual process. The newly dated reaffirmation statement, a link to the previous evidence review and recommendation statement, and the summary of the evidence are made available on the USPSTF Web site following usual processes.

6.5 Assessing Magnitude of Benefit

6.5.1 Definitions of Magnitude Ratings and Criteria

In situations where the evidence is adequate or convincing for benefit, the Task Force considers all the admissible evidence to determine the magnitude of benefit that would be expected from implementing the preventive service in a defined population. The magnitude of benefit is categorized as substantial, moderate, small, or zero. If the evidence is deemed inadequate for the assessment, the magnitude of benefit rating is not applicable. The Task Force uses the evidence to estimate the size of the population that would benefit from implementation of the preventive service over a given time horizon (appropriate to the service under consideration) and over the expected time to benefit. Specific health benefits might include such outcomes as overall mortality reduction, clinically meaningful improvements in health-related quality of life, or avoidance of specific disease events (e.g., cardiovascular events, cancer incidence and mortality, visual impairment, complications from alcohol use).

6.5.2 Determining Magnitude of Benefit Across the Analytic Framework

For some services, benefits can be estimated directly from large well-conducted RCTs of preventive services, with specific health benefits as prespecified outcomes and conducted in participants representative of the population under consideration. More commonly, however, this direct evidence is not available and the Task Force must assess the evidence across the key questions and the linkages in the analytic framework. For example, if an RCT of screening is not available, evidence may be assessed related to the accuracy of the screening test and the benefits of treatment in a screened population, and may include studies examining intermediate or final health outcomes. This indirect method of determining magnitude of benefit requires more assumptions and thus is associated with greater uncertainty than when direct evidence is available. In general, evidence derived from well-conducted trials evaluating a preventive service is likely to have more certainty than evidence from indirect assessment across key questions and linkages between intermediate and final health outcomes.

The Task Force examines both relative risk reduction (RRR) and absolute risk reduction (ARR) from intervention studies. It generally prioritizes ARR over RRR. That is, it places less emphasis on a large RRR in situations of low ARR; it remains interested in an intervention with a low RRR if its ARR is high. Even a low ARR may be important for critical outcomes (e.g., mortality).

6.5.3 General Considerations for Determining Limits on Magnitude of Benefit

Estimates of magnitude of benefit are meant to describe 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 preventive interventions, the population benefit may be further limited by such issues as the following:

1. The prevalence and incidence of the target condition
2. For heterogeneous conditions, the prevalence of the condition subtype that would cause important health problems
3. The sensitivity of the screening test (i.e., the degree to which the test or a given threshold to define abnormality of the screening test will detect the subtype of the condition that would potentially cause health problems; sensitivity is rarely 100%)
4. The comparative effectiveness of early treatment of asymptomatic disease relative to later treatment of symptomatic disease of the subtype of the condition that would cause health problems (rarely 100%)

6.5.4 Conceptual Confidence Limits

As previously noted, estimates of magnitude of benefit are intrinsically more uncertain when direct evidence is limited or absent or restricted to select populations or clinical scenarios. In these cases, the Task Force may place conceptual upper or lower bounds on the magnitude of benefit as applied to the population targeted in the recommendation. Considerations such as baseline risk of study participants and the clinical setting in which the studies were conducted also factor into the bounds of estimates of magnitude of benefit. For example, if magnitude of benefit is estimated only from studies of an intervention conducted by highly trained clinicians using specialized equipment for persons at considerably increased risk, this estimate might be considered the upper bound for benefit that might reasonably be anticipated for a general population. In other situations, the Task Force may also logically judge the lower bounds of the benefit, particularly when estimating the anticipated benefits in a population with a lower prevalence of disease than the study population in which the estimate of the benefit was derived.

Screening for abdominal aortic aneurysm is an example of the Task Force's use of conceptual confidence intervals. The benefits observed in screening studies of male smokers that were conducted in academic centers with optimal diagnostic and surgical treatment capabilities were judged to likely represent the upper bounds of benefit if these services were to be provided more generally in community-based settings. A lower conceptual bound of potential benefit was judged when extrapolating these studies in a high-risk population (male smokers) to populations at lower risk (male nonsmokers and female smokers and nonsmokers).

6.5.5 Outcomes Tables and Decision Modeling for Determining Magnitude of Benefit

One way to determine the magnitude of benefit is to use an outcomes table based on the systematic evidence review or, when available, outputs from a decision model. An outcomes table can demonstrate how many or the proportion of persons likely to benefit—and in what ways—from implementation of the preventive service. Estimates from direct and indirect evidence may be included in outcomes tables in order to provide the range of expected magnitude of specific beneficial outcomes (**Appendix VIII**).

6.6 Assessing Magnitude of Harm

6.6.1 Definitions of Magnitude Ratings and Criteria

The Task Force starts with the conceptual notion that screening, counseling, or use of preventive medications are intended for asymptomatic individuals in order to prevent or delay future health problems. The burden of proof that the benefits exceed the harms prior to recommending implementation of screening or other preventive services is thus higher than it is for diagnosis or treatment of symptomatic conditions. As such, assessment of the magnitude of harm is critically important. As with the magnitude of benefit, in situations where the evidence is adequate or convincing for harm, the magnitude of harm is assessed using the following categories: substantial, moderate, small, or zero. If the evidence is deemed inadequate for the assessment, the magnitude of harm rating is not applicable.

The Task Force uses the evidence to estimate the size of the population that would be harmed from implementation of the preventive service over a given time horizon (appropriate to the service under consideration) and over the expected time to be harmed and the duration/severity of the harm. Assessment of the magnitude of harm may be more difficult than assessment of benefit for many reasons. The broad range of potential harms is often less well identified or reported than potential benefits. At times severe harms occur at a relatively infrequent rate compared to benefits and require larger sample sizes than those studied in RCTs designed to evaluate benefits. Unlike fairly discrete benefits that the preventive service is intended to provide, harms are often varied and complex, occur at several stages in the screening cascade (including at earlier times than for benefits), may persist, and may be poorly recognized. Furthermore, for many, understanding that screening and preventive tests and procedures can cause harm is conceptually difficult.

As with benefit, the magnitude of harm might be determined directly from the reported results of large well-conducted RCTs of a preventive intervention, but more often also requires an assessment across the key questions and the linkages in the analytic framework (even when RCTs are available). Nonrandomized studies are often considered a more reliable source of detecting and determining the magnitude of harm (especially rare but serious harms) than for assessment of benefit. Data on harms may be inadequate for an assessment of magnitude, even when there is adequate data to characterize benefit because of the variability in the reporting of harms and the fact that many studies are not statistically powered or designed to detect some harms.

6.6.2 General Types of Harm for Consideration

The Task Force starts with the assumption that nearly all preventive interventions have the potential to result in some magnitude of one or more harms to patients. For screening-based recommendations, the Task Force looks for harms of the screening test, the subsequent diagnostic tests resulting from screening, and early treatment of screen-detected asymptomatic disease. For recommendations that involve preventive medications and behavioral interventions, the Task Force looks at the magnitude of harm from these interventions.

Harms of screening may include psychological harm from labeling, the harms of diagnostic studies to confirm the presence of the condition, and overdiagnosis of screen-detected conditions. Because screening and other preventive interventions are implemented in asymptomatic persons with the goal of preventing future disease, the Task Force places a high priority on the effects of overdiagnosis and overtreatment, whereby the preventive service has the unintended consequence of creating “disease” that often leads to unnecessary and ineffective treatment. Harms of early treatment and overdiagnosis may accrue to patients whose condition 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 that would not have occurred in the absence of screening.

Harms may also be considered in the form of opportunity costs for both patients and providers. The Task Force may consider the time and effort required by both patients and the health care system to implement the preventive care service. If the time and effort are judged to be substantially greater than other preventive services delivered in the primary care setting, these factors are also considered in the harms category. The Task Force usually derives qualitative, rather than precise, estimates of opportunity costs.

Although opportunity costs may be considered in the determination of Task Force recommendation grades, financial costs are not. Financial costs are also not considered in the decision models used for Task Force recommendations.

6.6.3 Conceptual Confidence Intervals in Face of Inadequate Direct Evidence of Harms

Although there is often less evidence about potential harms than about potential benefits, the Task Force may draw general conclusions from evidence on 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%, the positive predictive value may be low and false-positive test results will be expected. If the diagnostic workup is invasive or otherwise carries clinically important potential for harm, the Task Force can infer that at least some harms will result from implementation of the screening program, because some persons with false-positive screening tests will undergo an invasive diagnostic protocol for no possible benefit.

Similarly, if overdiagnosis (and therefore overtreatment) is common, and if the treatment has some adverse effects, the Task Force may infer that implementation of routine screening will cause at least some incremental harms, even in the absence of studies that characterize harms. This approach does not require an exact estimate of the magnitude of harm, 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 diagnostic workup, and the expected amount of overtreatment. Care should be taken to call attention to the estimate’s lack of precision.

6.6.4 Presentation of Harms in Outcomes Tables

As with the magnitude of benefit, the magnitude of harm may be informed by an outcomes table based on the systematic review or, when available, outputs from a decision model. When outcomes tables are used to present benefits, estimates for harms will also always be presented.

6.7 Assessing Certainty of Evidence for the Entire Analytic Framework

6.7.1 Overview

The 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 the benefits minus the harms 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.

The recommendation grade for a preventive service is derived from separate assessments of the certainty of evidence for a service and the magnitude of net benefit the service offers to persons when performed.

Assessing the certainty of evidence requires a complex synthesis of all evidence across the entire analytic framework. Ultimately, the Task Force rates the certainty of the evidence as one of three categories: high, moderate, or low.

The goal of this assessment is to judge whether the results observed in the individual studies that comprise the body of evidence would be expected when the intervention is delivered to asymptomatic persons by providers in U.S. primary care settings.

6.7.2 Assessment of Certainty of Net Benefit

The Task Force uses multiple sources of information to rate certainty of net benefit, using the criteria listed in **Table 3**. The evidence review and tables provide much of the data, but the Task Force also uses the information on adequacy derived from the assessment of key questions and linkages (**Appendix XI**).

The evidence is evaluated for both the direct linkage (generally for key question 1) and indirect linkages (involving multiple key questions). The direct pathway is typically derived from RCTs of the targeted screening or preventive intervention that adequately measure the desired health outcomes in the population(s) of interest. This type of evidence is generally associated with higher level of certainty. If certainty for net benefit cannot be derived from the *direct* pathway, then the Task Force determines if the evidence is sufficient across the key questions and linkages in the *indirect* pathway to determine overall certainty. If there is a clear gap in the chain of evidence in the indirect pathway (e.g., insufficient evidence for treatment or a screening test), then the certainty across the entire framework is categorized as “low” (Section 6.7).

If there is at least adequate evidence for either the direct or indirect pathway, the Task Force addresses how the body of evidence within the analytic framework fits together to provide an accurate and coherent estimate of the expected magnitude of net benefit (i.e., benefits minus harms) that would be realized from widespread implementation of the preventive service either in the general population or in specific subpopulations. To achieve moderate or high certainty, the Task Force requires that the body of evidence is applicable to asymptomatic persons in the U.S. primary care population, and that the services can be feasibly delivered in U.S. primary care settings or referred to outside resources.

6.7.3 Conceptual Confidence Intervals to Define Certainty Levels

Certainty may also be thought of as the width of the conceptual confidence interval given by the evidence to estimate the magnitude of net benefit. This is not a quantitative calculation, but rather a judgment based on the six critical appraisal questions and how the evidence fits together to complete the linkages from the left side of the analytic framework (population) to the right side (health outcomes). A wide conceptual confidence interval can be due to: lack of evidence about one or more key questions or inadequate evidence to support the linkages; limitations in study design (including inadequate power or poor internal or external validity); too few studies; inconsistency or incoherence of results across studies; or other aspects of the studies that cloud the interpretation of the magnitude of net benefit. When the conceptual confidence interval is wide, the magnitude cannot be estimated with sufficient confidence, and the entire body of evidence is categorized as having low certainty.

When the evidence satisfies most of the six critical appraisal criteria and fits together well enough to make the connections across the analytic framework, the conceptual confidence interval is considered to be narrower. In this case, there is a better (although not precise) estimate of the magnitude of benefit, harm, and net benefit. This type of body of evidence is categorized as having moderate certainty.

When the evidence satisfies each of the six critical appraisal criteria across the analytic framework and the evidence fits together well, the conceptual confidence interval is narrow there is a precise estimate of the magnitude of benefit, harm, and net benefit. In this case, the body of evidence is categorized as having high certainty. **Table 3** defines the three certainty levels of the overall evidence.

Table 3. USPSTF Levels of Certainty Regarding Net Benefit

Level of Certainty	Description
High	The available evidence usually includes consistent results from a multitude of well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on the desired health outcomes. Because of the precision of findings, this conclusion is therefore unlikely to be strongly affected by the results of future studies. These recommendations are often based on direct evidence from clinical trials of screening or behavioral interventions. High-quality trials designed as "pragmatic" or "effectiveness" trials are often of greater value in understanding external validity.
Moderate	The available evidence is sufficient to determine the effects of the preventive service on targeted health outcomes, but confidence in the estimate is constrained by factors such as: • The number, size, or quality of individual studies in the evidence pool • Some heterogeneity of outcome findings or intervention models across the body of studies • Mild to moderate limitations in the generalizability of findings to routine primary care practice. 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.
Low	The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of: • The very limited number or size of studies • Inconsistency of direction or magnitude of findings across the body of evidence • Critical gaps in the chain of evidence • Findings are not generalizable to routine primary care practice • A lack of information on prespecified health outcomes • Lack of coherence across the linkages in the chain of evidence More information may allow an estimation of effects on health outcomes.

The Task Force is careful to assess the certainty of the evidence and the magnitude of benefit, harm, and net benefit separately. 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 net benefit. The Task Force may also have moderate certainty of the evidence and determine that there is a substantial magnitude of net benefit.

6.7.4 Implementation Considerations When Grading Certainty

The Task Force seeks to make recommendations based on projections of what would be expected from widespread implementation of the preventive service in primary care practice settings across the United States. For this reason, the Task Force carefully considers the applicability to clinical practice of "efficacy" trials, which measure the effects of the preventive care service under ideal circumstances. The Task Force ultimately seeks to weight its recommendations with "effectiveness" trials, which measure the effects of widespread implementation under usual practice circumstances. Such studies are not always available. Therefore, the Task Force attempts to estimate the likelihood that the benefits and harms reported in efficacy studies could be replicated in clinical practice and non-study populations. Additionally, the Task Force carefully considers the real-world feasibility of interventions specified in efficacy studies. Some practices have greater support and more resources than others to implement recommended services.

6.8 Assessing Magnitude of Net Benefit

6.8.1 Definitions and Criteria Used to Assess the Magnitude of Net Benefit

To specify the magnitude of the effect of a preventive service, the Task Force separately assesses the magnitude of benefit and harm and then combines these into an assessment of net benefit. The Task Force defines net benefit as the magnitude of the benefit of the service minus the magnitude of the harm. 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 (e.g., diagnostic workup). Furthermore, preventive services are offered to asymptomatic persons. The majority of persons do not benefit from the service, and if they do, the benefit is in the future, often requiring many years to realize. Yet these persons are still subject to harms (often immediate), including opportunity costs.

This initial assessment is first developed by the topic leads after assessing the benefits and harms for each of the key questions and linkages. It is then presented to the full Task Force at the time the draft recommendation statement is

deliberated. The Task Force requires the certainty of evidence to be either moderate or high in order to make an assessment on the magnitude of net benefit. If the certainty of the evidence is low, the Task Force is unable to assess the magnitude of net benefit of the preventive service.

The Task Force rates net benefit as substantial, moderate, small, or zero/negative. “Substantial” net benefit indicates that the benefits substantially outweigh the harms, whereas “zero/negative” net benefit indicates that the harms equal or outweigh the benefits.

6.8.2 Metrics and Data Used to Assess the Magnitude of Net Benefit

Weighing the balance of benefits and harms can be challenging, because these outcomes are often measured in different metrics and over different time frames. Benefits are often quantified in terms of lives extended, quality of life improved, or illness events averted. Many of these often take years to achieve and may only accrue to a small percentage of individuals. Harms are often measured using metrics such as the false-positive screening test rate, overdiagnosis, diagnostic tests and their complications, or adverse effects of treatment, some of which are ineffective or unnecessary. Many of these occur more frequently and earlier than benefits. When the body of evidence does not use a single metric common to both benefits and harms, the assessment of net benefit is inherently subjective.

The Task Force attempts to quantify the magnitude of benefit and harm 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” (i.e., NNT, the number of persons who would need to be treated for some defined period of time to prevent one adverse health event) or “number needed to screen” (i.e., NNS, the number of persons who 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” (i.e., NNH, the number of persons needed to be treated or screened for a defined period of time to cause one adverse health event). Because of the uncertainty and variability in the evidence used to make these estimates, the Task Force does not have a defined threshold for NNT, NNS, or NNH for assessing the magnitude of net benefit.

When results from decision modeling conducted for the Task Force are available, the outputs may specifically inform the assessment of magnitude of net benefit (Section 5). Similarly, an outcomes table generated from a systematic review also outlines the tradeoffs in terms of projected benefits and harms in a population. The Task Force has standardized this outcomes table to the extent possible (**Appendix VIII**), but there will always be some variation, depending on the topic.

The Task Force does not use specific criteria to differentiate levels of net benefit. Net benefit, as used by the Task Force, is often assessed as 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) would be relieved from society by implementing the preventive service. This criterion applies even if the target condition is rare.
2. A large amount of the burden of suffering would be relieved from society (minus the additional burden caused by the preventive service) by implementing the preventive service.

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 the harms of the intervention are zero or small. Thus, both the magnitude of harm and the magnitude of benefit are each critical factors in determining net benefit.

Section 7. Formulation of Task Force Recommendations

7.1 General Principles for Making Recommendations

7.1.1 All recommendations are based on a body of scientific evidence that is derived from systematic evidence reviews and can use modeling to inform the process and make decisions after full consideration of the certainty and magnitude of net benefit (Section 5).

Evidence may come from indirect evidence in the analytic framework, but ultimately the complete chain (linking populations with health outcomes) must be supported by acceptable evidence.

Inferences about supporting evidence can include generalizations from one population to other subgroups when there are acceptable grounds to assume the evidence is applicable to both.

The Task Force invites and considers the opinions of the public and experts throughout the recommendation development process, including the draft evidence review and the draft recommendation statement. The Task Force is particularly interested in receiving comments on the sufficiency of the systematic review process and interpretation of the body of evidence. However, expert opinion and clinical experience cannot substitute for the body of evidence that the Task Force reviews through a systematic process.

Recommendations describe services that should or should not be routinely offered based on scientific evidence, although it is recognized that in clinical practice and public policy, concerns other than scientific evidence (e.g., feasibility, public expectations) may take precedence.

7.1.2 When making recommendations, the Task Force considers most strongly patient-oriented health benefits and harms.

In assessing health benefits, outcomes that patients can feel or care about (e.g., pain, quality of life, disease-specific death, overall mortality) receive more weight than intermediate outcomes.

In judging the magnitude of benefit, absolute reductions in risk matter more than relative risk reductions.

Evidence for service effectiveness is considered as valuable as, 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. 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. The quality of evidence for harms need not be as strong as that for benefits because of the ethical imperative to do no harm, especially when caring for asymptomatic persons. Physical, psychological, and social harms are considered.

Judgments about tradeoffs between benefits and harms are generally made at the population level. For interventions where the relationship between benefits and harms is influenced heavily by personal preferences, the Task Force advocates that providers and patients engage in shared decision-making.

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).

7.1.3 The USPSTF does not consider the financial costs of providing a service in its assessment of the balance of benefits and harms, but may provide contextual information regarding costs for use by providers, including cost-effectiveness studies.

7.1.4 Recommendations apply only to persons without signs or symptoms of the condition for which the preventive service is intended.

Persons living in the United States are the target population for all recommendation statements. The evidence reviews and recommendations may be useful in other countries, but may not apply to populations with markedly different epidemiology and health care system design.

7.1.5 Recommendations apply only to preventive services that are delivered in or are referable from the primary care setting to a specialist or community resource.

The evidence for preventive services delivered outside the primary care context (e.g., programs at schools, worksites, public health sites) is usually out of scope unless these services are linked to primary care.

7.2 Recommendation Grades

The Task Force applies grades to all of its recommendations and may issue multiple grades on a topic to address specific subpopulations. The Task Force can issue a grade of A, B, C, or D, as described in **Table 4**. When evidence is insufficient to make a recommendation, the Task Force issues an "I statement."

Table 4. How to Interpret Task Force Recommendation Grades

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or 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 or provide this service.
C	The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small.	Offer or provide this service for selected patients depending on individual circumstances.
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 the Clinical Considerations section of the USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

After full consideration and decision on both certainty and magnitude of net benefit, the topic leads discuss the appropriate grade for the service in the targeted population, using the scoring matrix in **Table 5**.

Table 5. U.S. Preventive Services Task Force Recommendation Grade Grid: Certainty of Net Benefit and Magnitude of Net Benefit

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

The Task Force values consistency in our process for determining grades. Changes in the grade when updating a previously published recommendation should have a strong rationale that stems directly from our process of determining grades (i.e. there is a difference in certainty or magnitude that warrants a change in grade). After the leads discuss the adequacy of the evidence on calls leading to the vote at the TF meeting, the leads identify any grade changes and discuss the rationale for proposed grade change.

A grade may result in a change from a previous Task Force recommendation because of one or more of the following: 1) a change in methods and/or analytic framework since the last recommendation statement; 2) a change in the definition of a grade (i.e. change in C grade definition); 3) evidence has increased or decreased and results in a change in the certainty or magnitude of net benefit, or has made the issuance of a grade less relevant. This may occur when there is a change in our understanding about the applicability of older evidence or international evidence; 4) new methods and/or new evidence regarding subpopulations. The TF strives to avoid a narrow "I" grade for a subpopulation when there is a grade for the overall population and no strong rationale exists that the subpopulation would be different from the larger population. Grade changes may also result from changes in context (clinical context, societal values for specific outcomes, and context of intervention and treatment). In this case, while the analytic frame work is largely similar to the prior framework, something has changed in the contextual issues. It is important that the Task Force communicate in its recommendation statement how the changes in the above factors or context affect our rating of certainty and magnitude and why this results in a grade that is different than a previously published grade.

Before the grading discussion, the Task Force is provided with an oral presentation summarizing the evidence to supplement the full evidence review provided by the EPC. Following clarification of any questions regarding the evidence, the Task Force then hears from the topic leads regarding their proposal for a grade. After full debate and consideration of grading options, the Task Force Chair calls for a motion for a draft recommendation grade (go to Section 7.4 for voting procedures). The leads refine the draft recommendation with final language before it is released for public comment (go to Section 9 for more information on public comment).

To help readers better 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 provides its rationale and statements about clinical considerations in the recommendation statement. While an "I statement" is considered a statement and not a recommendation, these topics are accompanied by the same type of rationale and clinical considerations as grade A, B, C, or D recommendations.

For clarity, consistency, and usability, Task Force recommendations follow a standard, structured format.

Each recommendation statement is also accompanied by a one-page clinical summary, which provides a table of key information about the recommendation, including the population of interest, recommendation, risk assessment, screening or intervention of interest, treatment, balance of benefits and harms, and other relevant USPSTF recommendations.

A fact sheet for each recommendation is also prepared for consumers. The Task Force also produces additional fact sheets, summary tables, infographics, and videos when appropriate to further explain recommendations to diverse audiences.

7.3 Process for Public Comment on Task Force Documents

To increase the clarity, transparency, and utility of its recommendation statements to primary care providers and the public, the Task Force shares drafts of its research plans, evidence reviews, and recommendation statements for public comment. The comments are considered in finalizing the documents. The procedures for posting draft materials for public comment are described in Section 9.

All comments received through the public comment process are shared with the topic leads for their review and consideration before finalizing the document. All Task Force members have access to the full text of all comments; a disposition table summarizing the comment themes and the proposed Task Force response; and the revised research plan, evidence review, or recommendation statement.

7.4 Voting

Formal votes are taken for major procedural and methodological decisions, for draft recommendations before posting, for final recommendations, and for statements about clinical practice. Votes may be taken for other decisions at the discretion of the Chair(s).

7.4.1 General Voting Procedures

All motions on recommendations (at any stage) requiring a vote are passed when two thirds of the current Task Force membership vote "yes." Votes are taken by voice, hand, or email, without secret ballots.

Motions on procedural, methodological, and other decisions requiring a vote are passed when a majority of current Task Force membership votes "yes."

Votes are submitted as yes, no, abstain, or absent.

Members recused by reason of potential conflict of interest are recorded as recused and do not vote.

In votes that are less than unanimous, there are no minority reports.

The result of a vote is recorded in the meeting minutes, though the count of "yes," "no," and "abstain" votes is not recorded.

7.4.2 Voting on Draft Recommendations

At a meeting of the full Task Force (usually in person), the presiding Chair accepts motions for draft recommendations. A "yes" vote from two thirds of the current Task Force membership is needed to pass the motion. After the meeting, the topic leads draft the full recommendation statement, and it is posted for public comment.

7.4.3 Motion to Reconsider a Draft Recommendation Already Voted Upon

A vote to reconsider a motion on a draft recommendation is required if the topic leads or any other Task Force member would like to request consideration for a change in the grade of the draft recommendation that was voted by the full Task Force. In this case, the individual member or the primary topic lead speaks with the Chairs, requests a vote to reconsider, and, if passed (requiring a "yes" vote from two thirds of the current membership), makes a new motion. The Chair then calls for a vote (which may take place via conference call or email after several days of reflection and discussion). A "yes" vote from two thirds of the current Task Force membership is needed to pass the motion on the new draft recommendation. If the motion to approve a reconsidered recommendation fails to pass, the approval reverts back to the originally approved recommendation.

7.4.4 Voting on Final Recommendations

After consideration of public comments, the topic leads puts forward a new motion for consideration by the full Task Force for the final recommendation. If the final recommendation statement is similar to the posted draft, debate is limited, and the full Task Force votes via email. A "yes" vote from two thirds of the current Task Force membership is needed to pass the motion and ratify the final recommendation.

If, as a result of the comment process or new evidence identified during the public comment period, any member of the Task Force believes that a change in the recommendation grade is warranted, he or she can request that the topic leads make a motion to the Task Force. At that point, any new evidence is reviewed by the topic leads with help from AHRQ and the appropriate EPC staff. The AHRQ Medical Officer and Scientific Director facilitate this process. The topic leads present their motion and any important new evidence to the full Task Force (most often via conference call or Webinar), followed by time for discussion. The Chair then calls for a vote on the motion (which may take place via email after several days of reflection and discussion). This approach recognizes that the vote on the final recommendation is a different motion than the vote on the draft (hence two separate motions and votes). A "yes" vote from two thirds of the current Task Force membership is needed to pass the motion and ratify the final recommendation.

7.5 Dissemination Process and Products

Task Force recommendations are widely disseminated to professional audiences in professional peer-reviewed journals, in an electronic tool (Prevention TaskForce) available online or as a mobile application, in print through the "Guide to Clinical Preventive Services," and as reprints in peer-reviewed journals, such as *American Family Physician*. Tools for clinicians, including Prevention TaskForce, are available on the Task Force Web site at https://uspreventiveservicestaskforce.org/uspstf/recommendation-topics/tools-and-resources-for-better-preventive-care.

AHRQ's Center for Evidence and Practice Improvement is committed to improving the health of people nationwide by working to make sure that *everyone* in the health care system—clinicians, consumers, providers, and payers—knows about and uses these evidence-based clinical preventive services.

To achieve this goal, AHRQ's Center for Evidence and Practice Improvement 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.

Section 8. Workgroups of the USPSTF

Several standing and ad hoc workgroups are committed to ensuring that the Task Force's methods and processes are up to date and implemented consistently and transparently.

The **Methods Workgroup** reviews and updates Task Force methods and processes to follow best practices for guideline-setting bodies and incorporate methodological advances. This workgroup identifies issues that need further consideration, recommends the creation of new workgroups as needed to address these issues, and incorporates input from all other workgroups into Task Force methods and processes.

The **Topic Prioritization Workgroup** develops procedures for prioritizing the portfolio of USPSTF topics and reviews and prioritizes nominations for new topics and suggestions for reconsidering or updating existing topics from the public. It also proposes a determination of the status of all topics (active, inactive, and referred to others) and prioritization of the active queue of topics each year for consideration by the full Task Force.

The **Subpopulation Workgroup** assesses methods for using evidence from published studies on the differential effects of clinical preventive services within relevant population subgroups defined by race/ethnicity, sex, age, and other clinically relevant characteristics. It also suggests processes for incorporating this evidence into the Task Force's deliberations and recommendations.

The **Older Adults Workgroup** helps the Task Force assess the applicability of its recommendations to older adults by offering guidance on the benefits and harms of clinical preventive services at older ages.

The **Child and Maternal Health Workgroup** provides specialized knowledge to inform the work of the USPSTF and develop new methods and procedures for making recommendations for child and maternal health. Activities of the workgroup include publishing articles on USPSTF methods related to child and maternal health, addressing methodological issues such as the challenges of identifying meaningful health outcome measures for children and adolescents, and serving as consultants on relevant Task Force projects and topics.

The **Conflict of Interest Workgroup** is an ad hoc committee that reviews and updates Task Force policy on reporting and addressing Task Force members' conflicts of interest in regard to Task Force topics.

The **Behavioral Counseling Intervention Workgroup** makes recommendations related to the standards of evidence for behavioral counseling interventions, relevant measures and metrics, coordination with the Community Preventive Services Task Force, knowledge gaps, and other methodological issues related to behavioral counseling interventions.

The **Modeling Workgroup** identifies opportunities to further inform the recommendation process through the use of decision models as a complement to systematic evidence reviews.

The **Dissemination and Implementation Workgroup** helps the Task Force better communicate with clinicians and members of the public about its recommendations, and also writes the Task Force's annual report to Congress.

The Task Force also occasionally convenes groups of experts to advise on a particular topic. For example, the Task Force has convened groups of experts to discuss methods related to behavioral counseling interventions and prevention for older adults.

Section 9. Engagement With the Public, Stakeholders, and Partners

The Task Force is committed to making evidence-based recommendations about clinical preventive services that are valid, reliable, and useful to clinicians, patients, and family members. The Task Force is also committed to making the recommendation development process clear and transparent, and there are several opportunities for the public and other stakeholders to engage in the recommendation process.

9.1 Engagement With the Public

The Task Force engages with the public in many steps throughout its recommendation making process. Currently, through the Task Force website, anyone can:

- Nominate new members to serve on the Task Force
- Nominate new topics for Task Force consideration or request an update of an existing topic
- Provide comments on draft research plans
- Provide comments on draft evidence reviews
- Provide comments on draft recommendation statements

Anyone who is interested in the Task Force's work can sign up for email alerts on the USPSTF Web site.

9.2 Engagement With Liaisons and Partners

Since its inception, the Task Force has worked with a group of standing Federal liaisons and Dissemination and Implementation Partners.

Federal liaisons provide input from national scientific experts and keep the Task Force apprised of major Federal initiatives that may produce new evidence or duplicate the Task Force's efforts on a given topic. In the case of a potential duplication of effort, the Task Force may choose to refer the recommendation topic to another group.

Dissemination and Implementation Partners are major national organizations representing primary care clinicians, consumers, and other stakeholders involved in the delivery of primary care. They help the Task Force to ensure that its recommendations are meaningful to the groups they represent and help put the recommendations into practice. They are also a powerful vehicle for ensuring that America's primary care workforce remains up to date on Task Force recommendations.

Both Federal liaisons and Dissemination and Implementation Partners are invited to observe Task Force meetings. Engagement with partners also includes email updates, Webinars, and opportunities for public comment and dissemination. Partner organizations may choose to promote opportunities for public comment among their membership and assist their members with the implementation of final recommendation statements.

9.3 Participation in the USPSTF Member and Topic Nomination and Recommendation Processes

Members of the public can participate in the USPSTF process by nominating new members to the Task Force. The USPSTF Web site has a page where anyone can nominate candidates for consideration or self-nominate. Nominations must be received by March 15 of a given year to be considered for an appointment that will begin in January of the following year.

Members of the public can also participate in the recommendation process itself. The Task Force provides opportunities for public comment at four stages of the recommendation process:

- Topic nomination
- Draft research plan
- Draft evidence review
- Draft recommendation statement

Anyone can nominate a new topic at any time. Public comment periods for draft research plans, evidence reviews, and recommendation statements last for 4 weeks on the USPSTF Web site. The public comment period helps to ensure that final recommendations are valid, reliable, and useful to clinicians, patients, and family members.

9.4 Dissemination of USPSTF Recommendations and Processes

The Task Force disseminates its research plans, methods, evidence reviews, and recommendation statements through its Web site. Different dissemination activities, described below, are conducted for researchers; clinicians; members of the public, including consumers and patients; and the media.

9.4.1 Researchers

All Task Force research plans, evidence reviews, and recommendation statements are posted on the USPSTF Web site. Task Force final evidence reviews and final recommendation statements are also published in peer-reviewed journals. Journal articles are sometimes accompanied by videos summarizing their key "take home messages". Further details about Task Force procedures for writing papers and documents are available in Section 1.10.

9.4.2 Clinicians

Health care professionals have access to the full library of Task Force recommendations and evidence reviews on the USPSTF Web site.

The Task Force also shares its recommendations through Prevention TaskForce, an application for smartphone, mobile, and tablet devices designed to help primary care clinicians identify clinical preventive services that are appropriate for their patients. Users can search the tool for recommendations by patient age, sex, pregnancy status, and risk factors. Prevention TaskForce is available at https://www.uspreventiveservicestaskforce.org/apps/.

Additionally, to help primary care clinicians learn about its recommendations and put them into practice, the Task Force has posted resources about its role in preventive medicine and its process for developing evidence-based recommendations. These resources are available at https://uspreventiveservicestaskforce.org/uspstf/about-uspstf/task-force-resources.

9.4.3 The Public

In 2012, the Task Force began producing plain-language consumer fact sheets for each of its draft and final recommendations to help members of the public, including consumers and patients, understand what each recommendation means for them.

The consumer fact sheets, posted on the USPSTF Web site, contain links to resources for learning more about each topic and encourage individuals to have informed discussions about clinical preventive services with their health care provider.

In addition to breaking down the main points of each recommendation, consumer fact sheets for draft recommendation statements explain how to offer feedback to the Task Force about the draft recommendation statement.

Further, in response to common questions posed to the USPSTF, the Task Force produced a series of materials explaining its mission, composition, and processes, including an introductory slide show called "USPSTF 101," available to view and download at https://uspreventiveservicestaskforce.org/uspstf/about-uspstf/task-force-resources.

Finally, the Task Force engages with the public through MyHealthFinder, a mobile and Web-based application similar to Prevention TaskForce and geared to a consumer audience. Users may search Task Force recommendations by sex, age, pregnancy status, and risk factors at http://healthfinder.gov. The application also incorporates recommendations from ACIP, Bright Futures, and the Institute of Medicine's Committee on Preventive Services for Women.

9.4.4 The Media

The Task Force engages with the media in several ways to disseminate information about research plans, evidence reviews, and draft and final recommendation statements. These include news bulletins, release of materials under embargo in advance of publication, and interviews with Task Force members.

News bulletins are developed for each draft and final recommendation statement. The purpose of news bulletins is to organize key information about a Task Force recommendation for the media. The news bulletin may include a summary of the recommendation and supporting science as well as quotes from a Task Force spokesperson. News bulletins are intended to streamline the release of information concerning the Task Force's work. The USPSTF Web site also has a Newsroom page (https://uspreventiveservicestaskforce.org/uspstf/news) to provide reporters with background information on the Task Force and media contact information.

Section 10. References

Butler M, Talley KM, Burns R, Ripley A, Rothman A, Johnson P, et al. Values of Older Adults Related to Primary and Secondary Prevention. Evidence Synthesis No. 84. AHRQ Publication No. 11-05154-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2011.

Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, Editors. *Primary Care: America's Health in a New Era. Committee on the Future of Primary Care, Division of Health Care Services*. Institute of Medicine. National Academy Press. Washington, D.C. 1996: p. 31.

Eckstrom E, Feeny DH, Walter LC, Perdue LA, Whitlock EP. Individualizing cancer screening in older adults: a narrative review and framework for future research. *J Gen Intern Med*. 2013;28(2):292-8.

Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. AHRQ Publication No. 13-EHC128-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

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;147(2):117-22.

Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al; 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;20(3 Suppl):21-35.

Kane RL, Talley KM, Shamliyan T, Pacala JT. Common Syndromes in Older Adults Related to Primary and Secondary Prevention. Evidence Report/Technology Assessment No. 87. AHRQ Publication No. 11-05157-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2011.

Leipzig RM, Whitlock EP, Wolff TA, Barton MB, Michael YL, Harris R, et al; U.S. Preventive Services Task Force Geriatric Workgroup. Reconsidering the approach to prevention recommendations for older adults. *Ann Intern Med*. 2010;153(12):809-14.

Lin JS, Whitlock EP, Eckstrom E, Fu R, Perdue LA, Beil TL, et al. Challenges in Synthesizing and Interpreting the Evidence From a Systematic Review of Multifactorial Interventions to Prevent Functional Decline in Older Adults. Rockville, MD: Agency for Healthcare Research and Quality; 2012.

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. *Lancet*. 1999;354(9193):1896-900.

Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt T; U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. *Ann Intern Med*. 2009;150(3):199-205.

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;147(12):871-5.

Smetana GW, Umscheid CA, Chang S, Matchar DB. Methods guide for authors of systematic reviews of medical tests: a collaboration between the Agency for Healthcare Research and Quality and the *Journal of General Internal Medicine*. *J Gen Intern Med*. 2012;27(Suppl 1):S1-3.

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al; for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*. 2000;283(15):2008-12.

U.S. Preventive Services Task Force. Grade Definitions. Accessed at http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions on 13 January 2015.

Whitlock EP, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence-based approach. *Am J Prev Med*. 2002;22(4):267-84.

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. AHCPR Pub. No. 95-0009. Rockville, MD: Agency for Health Care Policy and Research; 1994. pp. 105-13.

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. One of the duties of AHRQ is to convene the U.S. Preventive Services Task Force (42 U.S.C. §299b–4 (a) as amended by Public Law 106-129 (1999) and Public Laws 111-148 and 111-152 (2010), Sec. 4003):

1. ESTABLISHMENT AND PURPOSE.—The Director shall convene an independent Preventive Services Task Force (referred to in this subsection as the "Task Force") to be composed of individuals with appropriate expertise. Such 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, to be published in the Guide to Clinical Preventive Services (referred to in this section as the "Guide"), for individuals and organizations delivering clinical services, including primary care professionals, health care systems, professional societies, employers, community organizations, non-profit organizations, Congress and other policy-makers, governmental public health agencies, health care quality organizations, and organizations developing national health objectives. Such recommendations shall consider clinical preventive best practice recommendations from the Agency for Healthcare Research and Quality, the National Institutes of Health, the Centers for Disease Control and Prevention, the Institute of Medicine, specialty medical associations, patient groups, and scientific societies.
2. DUTIES.—The duties of the Task Force shall include—
   (A) the development of additional topic areas for new recommendations and interventions related to those topic areas, including those related to specific sub-populations and age groups;
   (B) at least once during every 5-year period, review[1] interventions and update[2] recommendations related to existing topic areas, including new or improved techniques to assess the health effects of interventions;
   (C) improved integration with Federal Government health objectives and related target setting for health improvement;
   (D) the enhanced dissemination of recommendations;
   (E) the provision of technical assistance to those health care professionals, agencies and organizations that request help in implementing the Guide[3] recommendations; and
   (F) the submission of yearly reports to Congress and related agencies identifying gaps in research, such as preventive services that receive an insufficient evidence statement, and recommending priority areas that deserve further examination, including areas related to populations and age groups not adequately addressed by current recommendations.
3. ROLE OF AGENCY.—The Agency shall provide ongoing administrative, research, and technical support for the operations of the Task Force, including coordinating and supporting the dissemination of the recommendations of the Task Force, ensuring adequate staff resources, and assistance to those organizations requesting it for implementation of the Guide's recommendations.
4. COORDINATION WITH COMMUNITY PREVENTIVE SERVICES TASK FORCE.—The Task Force shall take appropriate steps to coordinate its work with the Community Preventive Services Task Force and the Advisory Committee on Immunization Practices, including the examination of how each task force's recommendations interact at the nexus of clinic and community.
5. 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].
6. INDEPENDENCE.—All members of the Task Force convened under this subsection, and any recommendations made by such members, shall be independent and, to the extent practicable, not subject to political pressure.
7. AUTHORIZATION OF APPROPRIATIONS.—There are authorized to be appropriated such sums as may be necessary for each fiscal year to carry out the activities of the Task Force.

Sec. 2713 of the Affordable Care Act requires private insurers to cover preventive services recommended by the USPSTF with a grade of A or B, along with those recommended by ACIP, Bright Futures, and HRSA's guidelines for women's health. The Affordable Care Act requires insurers to cover these services with no cost-sharing (i.e., no deductible and no co-pay).

Sec. 4105 of the Affordable Care Act authorizes Medicare to expand its existing coverage of preventive services consistent with USPSTF recommendations. Services covered by Medicare prior to the Affordable Care Act, and which have received a grade of A, B, C, or I from the USPSTF, must still be covered. However, Sec. 4105 authorizes Medicare to not pay for services that have received a grade of D from the USPSTF.

Finally, Sec. 4106 of the Affordable Care Act requires Medicaid to cover preventive services recommended by the USPSTF with a grade of A or B, as well as those recommended by ACIP.

SEC. 2713. COVERAGE OF PREVENTIVE HEALTH SERVICES.
''(a) IN GENERAL.—A group health plan and a health insurance issuer offering group or individual health insurance coverage shall, at a minimum provide coverage for and shall not impose any cost sharing requirements for—
''(1) evidence-based items or services that have in effect a rating of 'A' or 'B' in the current recommendations of the United States Preventive Services Task Force;
''(2) immunizations that have in effect a recommendation from the Advisory Committee on Immunization Practices of the Centers for Disease Control and Prevention with respect to the individual involved; and
''(3) with respect to infants, children, and adolescents, evidence-informed preventive care and screenings provided for in the comprehensive guidelines supported by the Health Resources and Services Administration.
''(4) with respect to women, such additional preventive care and screenings not described in paragraph (1) as provided for in comprehensive guidelines supported by the Health Resources and Services Administration for purposes of this paragraph.
''(5) for the purposes of this Act, and for the purposes of any other provision of law, the current recommendations of the United States Preventive Service Task Force regarding breast cancer screening, mammography, and prevention shall be considered the most current other than those issued in or around November 2009.

Nothing in this subsection shall be construed to prohibit a plan or issuer from providing coverage for services in addition to those recommended by United States Preventive Services Task Force or to deny coverage for services that are not recommended by such Task Force.

(b) INTERVAL.—
(1) IN GENERAL.—The Secretary shall establish a minimum interval between the date on which a recommendation described in subsection (a)(1) or (a)(2) or a guideline under subsection (a)(3) is issued and the plan year with respect to which the requirement described in subsection (a) is effective with respect to the service described in such recommendation or guideline.
''(2) MINIMUM.—The interval described in paragraph (1) shall not be less than 1 year.

SEC. 4105. EVIDENCE-BASED COVERAGE OF PREVENTIVE SERVICES IN MEDICARE.
(a) AUTHORITY TO MODIFY OR ELIMINATE COVERAGE OF CERTAIN PREVENTIVE SERVICES.—Section 1834 of the Social Security Act (42 U.S.C. 1395m) is amended by adding at the end the following new subsection:
''(n) AUTHORITY TO MODIFY OR ELIMINATE COVERAGE OF CERTAIN PREVENTIVE SERVICES.—Notwithstanding any other provision of this title, effective beginning on January 1, 2010, if the Secretary determines appropriate, the Secretary may—
''(1) modify—
''(A) the coverage of any preventive service described in subparagraph (A) of section 1861(ddd)(3) to the extent that such modification is consistent with the recommendations of the United States Preventive Services Task Force; and
''(B) the services included in the initial preventive physical examination described in subparagraph (B) of such section; and
''(2) provide that no payment shall be made under this title for a preventive service described in subparagraph (A) of such section that has not received a grade of A, B, C, or I by such Task Force.''
(b) CONSTRUCTION.—Nothing in the amendment made by paragraph (1) shall be construed to affect the coverage of diagnostic or treatment services under title XVIII of the Social Security Act.

SEC. 4106. IMPROVING ACCESS TO PREVENTIVE SERVICES FOR ELIGIBLE ADULTS IN MEDICAID.
(a) CLARIFICATION OF INCLUSION OF SERVICES.—Section 1905(a)(13) of the Social Security Act (42 U.S.C. 1396d(a)(13)) is amended to read as follows:
''(13) other diagnostic, screening, preventive, and rehabilitative services, including—
''(A) any clinical preventive services that are assigned a grade of A or B by the United States Preventive Services Task Force;
''(B) with respect to an adult individual, approved vaccines recommended by the Advisory Committee on Immunization Practices (an advisory committee established by the Secretary, acting through the Director of the Centers for Disease Control and Prevention) and their administration (…)

Appendix II. USPSTF Conflict of Interest Disclosure Form

Thank you for taking the time to complete the USPSTF Conflict of Interest Disclosure Form. In your role as a member of the USPSTF, the Task Force Chairs and AHRQ require full disclosure of all possible conflicts of interest. Please note that disclosure of potential conflicts of interest is part of the process to eliminate bias and ensure transparency of the process. Disclosing potential conflicts of interest does not necessarily disqualify you as a participant. For the purposes of this disclosure form, individuals should disclose all interests that apply to all USPSTF topics in development. For those topics for which you have disclosed interests on a previously submitted Disclosure Form, please update any new interests since the last disclosure was provided. For any new topics, please provide a full disclosure.

Full Name

Insert Task Force member name

U.S. Preventive Services Task Force Disclosure Form

Disclosure of Financial Interests: Task Force Members

For the purposes of this Disclosure Form, individuals should disclose all interests that apply to all USPSTF topics that will be covered during the upcoming meeting as well as all topics in development. For "Type of Financial Interest," please select one of A - H according to the following definitions:

a) Ownership or owning stock (stock, options, warrants) or holding debt or other significant proprietary interests or investments in any third party that could be affected by a USPSTF decision on a specific topic
b) Having an employment, independent contractor, or consulting relationship or other contractual arrangements, whether written or unwritten, with an entity that could be financially affected by a Task Force decision
c) Receiving a proprietary research grant or receiving patents, royalties, or licensing fees from such an organization
d) Participating on a company's proprietary governing boards
e) Participating in speakers bureaus
f) Receiving honoraria
g) Receiving payment as an expert witness for a plaintiff or a defendant
h) Receiving remuneration for services with respect to transactions involving parties with a financial interest in the outcome of a USPSTF decision. This may include clinical specialty practice

	Type of Financial Interest	Amount ($)	Description	Date Potential COI Occurred (mm/yy)
Task Force Member Name				
Task Force Member Name				

Disclosure of Financial Interests: Immediate Family Members and Close Personal Relationships

For the purposes of this Disclosure Form, individuals should disclose all interests that apply to all USPSTF topics that will be covered during the upcoming meeting as well as all topics in development. For "Type of Financial Interest," please select one of A - H according to the following definitions:

a) Ownership or owning stock (stock, options, warrants) or holding debt or other significant proprietary interests or investments in any third party that could be affected by a USPSTF decision on a specific topic
b) Having an employment, independent contractor, or consulting relationship or other contractual arrangements, whether written or unwritten, with an entity that could be financially affected by a Task Force decision
c) Receiving a proprietary research grant or receiving patents, royalties, or licensing fees from such an organization
d) Participating on a company's proprietary governing boards
e) Participating in speakers bureaus
f) Receiving honoraria
g) Receiving payment as an expert witness for a plaintiff or a defendant
h) Receiving remuneration for services with respect to transactions involving parties with a financial interest in the outcome of a USPSTF decision. This may include clinical specialty practice.

	Friend or Family Member	Relationship	Type of Financial Interest	Amount ($)	Description	Date Potential COI Occurred (mm/yy)
Task Force Member Name						
Task Force Member Name						

Disclosure of Nonfinancial Interests

For the purposes of this Disclosure Form, individuals should disclose all interests that apply to all USPSTF topics that will be covered during the upcoming meeting as well as all topics in development. For "Role," please select one of A - E according to the following definitions:

a) Public comments and testimony
b) Leadership role on a panel
c) Substantial career efforts/interests in a single topic area
d) Previously published opinions
e) Advocacy or policy positions

	Organization	Role	Type of Participation (Paid/Unpaid)	Description	Date Potential COI Occurred (mm/yy)
Task Force Member Name					
Task Force Member Name					

To the best of your ability, please respond yes or no to each of the questions listed below. If the answer is yes for any question, please include details or references that may be helpful in evaluating the potential influence of each relationship or personal belief. A "yes" answer will not necessarily disqualify you from participating in Task Force activities.

	Response (Required)	If Yes, Explain
A. Do you have strongly held beliefs related to the topic area that would make it difficult for you to work in an unbiased manner on any new or ongoing Task Force topics?		
B. Have you ever authored, coauthored, or publicly provided an opinion related to any new or ongoing Task Force topics?		
C. Could your institution benefit or be harmed based on whether the Task Force finds benefit, harm, or no difference in outcomes?		
D. Would the support you would receive from your institution (or primary mentor) change if your work on the Task Force generated a strong negative reaction from peers outside your institution?		
E. To the best of your knowledge, do you work for, or are you a member of, an organization with a stated position (e.g., position statement, Blog, editorial, legislature or legal testimony, or related document) related to any new or ongoing Task Force topics?		
F. Are you involved in formulating/voting for positions in any organization with a stated position related to any new or ongoing Task Force topics?		
G. Could this recommendation statement conflict with policies you have promoted or are obliged to follow?		

In the space below, please describe any nonfinancial interests for your immediate family members and close personal relationships related to all new and continuing topics.

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

I have listed all personal financial interests from the past 36 months in the Disclosure Form (including equity positions, consulting agreements, or employment arrangements with an entity that could be financially affected by a Task Force decision) for myself, my immediate family members, and close personal relationships for all new topics covered during this meeting, all topics in development, as well as any changes in my situation since this form was last completed for continuing topics. Period of disclosure is 36 months prior to the meeting and continues until the final reports are completed. I have declared any other real or perceived nonfinancial conflict(s) of interest for myself, immediate family members, and close personal relationships in the Disclosure Form related to the subject matter of

all new topics covered during this meeting, all topics in development, as well as any changes in my situation since this form was last completed for continuing topics.

I understand and agree to the above two items.

If for any reason you feel you cannot sign this statement as worded or if you have further questions, please contact the chairs.

Appendix III. USPSTF LitWatch Process

The USPSTF LitWatch is a regular audit of information sources to locate newly published research and/or guidelines that are relevant to topics in the USPSTF portfolio. A list of included articles and guidelines is compiled in a LitWatch newsletter every 2 months and distributed to the USPSTF Scientific Director, the USPSTF Topic Prioritization Workgroup, USPSTF members, EPC staff, AHRQ Medical Officers, and other related staff.

Articles and guidelines for consideration address primary or secondary preventive interventions in the general primary care setting. Topics of interest include screening, preventive services, effectiveness of early treatment of screen-detected disease, new technologies, and methodologies for care delivery. Special attention is given to topics currently or previously reviewed by the USPSTF. Articles on vaccinations (except those for human papillomavirus or a new breakthrough), community interventions, or general review articles (except systematic reviews) are not included. Articles are considered for inclusion in the LitWatch if they meet the following criteria: 1) possible impact on past USPSTF recommendations, 2) new evidence, and/or 3) importance to a current USPSTF evidence review.

The tables of contents of the following journals are reviewed for potential articles:

- *American Journal of Epidemiology*
- *American Journal of Health Promotion*
- *American Journal of Preventive Medicine*
- *American Journal of Public Health*
- *Annals of Family Medicine*
- *Archives of General Psychiatry*
- *Annals of Internal Medicine*
- *Archives of Disease in Childhood*
- *Archives of Internal Medicine*
- *Archives of Pediatric and Adolescent Medicine*
- *British Medical Journal*
- *Canadian Medical Association Journal*
- *Journal of the American Medical Association*
- *Journal of General Internal Medicine*
- *Journal of Medical Screening*
- *Journal of the National Cancer Institute*
- *Journal of Pediatrics*
- *Lancet*
- *Morbidity and Mortality Weekly Report*
- *New England Journal of Medicine*
- *Obstetrics and Gynecology (The Green Journal)*
- *Pediatrics*
- *Preventing Chronic Disease*

Guidelines issued by the following USPSTF partner organizations and other relevant groups are identified by reviewing the organization's Web pages during the last week of the month prior to distribution of the newsletter:

- Centers for Disease Control and Prevention
- Centers for Medicare & Medicaid Services
- Department of Defense Military Health System
- Department of Veterans Affairs Center for Health Promotion and Disease Prevention
- National Institutes of Health (including Physician Data Query®)
- Institute of Medicine
- American Academy of Family Physicians
- American Academy of Pediatrics
- American Congress of Obstetricians and Gynecologists
- American College of Physicians
- American College of Preventive Medicine
- American Osteopathic Association
- American Cancer Society
- American Diabetes Association
- American Heart Association
- Canadian Task Force on Preventive Health Care

Regular email updates from the National Guideline Clearinghouse™ are also reviewed for any guidelines of relevance, as well as *MedPage Today* for relevant news stories based on a recent publication.

Appendix IV. Roles and Responsibilities of USPSTF Members Serving as a Topic Lead

Each topic team (Section 1.9) includes the AHRQ Medical Officer, a Task Force Chair or Co-Vice Chair, representatives from the EPC conducting the systematic evidence review, and several Task Force members, known as "leads." One of the Task Force leads serves as the primary lead for that topic.

AHRQ staff solicits volunteers for the position of primary lead from among assigned topic leads during work plan development. When selecting a primary lead, an effort is made to choose an individual whose tenure on the Task Force will extend throughout the life of the topic.

The leads' role on a topic begins once they have received a draft work plan on the topic. A call is then scheduled to discuss the draft work plan.

Responsibilities of the Task Force topic leads include:

- Attending calls on the topic
- Attending in-person Task Force meetings where the topic is to be discussed
- Providing input on the draft research plan, reviewing public comments and proposed changes, and approving the research plan as final
- Providing input on the draft evidence review and reviewing the final evidence review in preparation for drafting the recommendation statement
- Assessing the evidence on each key question as convincing, adequate, or inadequate
- Assessing the certainty of evidence and magnitude of net benefit across all key questions
- Proposing a grade and supporting discussion of the grade at an in-person Task Force meeting
- Contributing to the drafting of a recommendation statement and reviewing the public comments received
- Contributing to revisions of the draft recommendation statement following the public comment period
- Approving the final recommendation statement for ratification by the full Task Force

Additional responsibilities of the primary lead (formerly known as the "lead lead"), beyond those of the other leads, include:

- Liaising with the AHRQ Medical Officer and the EPC and answering methodological questions as needed
- Reviewing every comment received from the public on the draft research plan and bringing other leads to consensus on revisions needed
- Facilitating discussion on calls and bringing other leads to consensus on key questions, certainty of evidence, magnitude of net benefit, and grade of the recommendation
- Presenting the draft recommendation statement at the in-person Task Force meeting
- Reviewing every comment received from the public on the draft recommendation statement and bringing other leads to consensus on revisions needed
- Potentially serving as spokesperson for the media, as needed
- Approving a one-page clinical summary for posting along with the recommendation statement

At the start of, and throughout a topic, the following applies:

- Scheduling of topic team calls will take the primary lead's schedule into account first.
- Calls may be scheduled if and only if the primary lead plus at least one other Task Force topic lead can attend the call.
- Other members of the topic team who cannot attend a scheduled call can direct comments prior to or after a call to the primary lead, all parties (e.g., via email), or the AHRQ Medical Officer.

Appendix V. Work Plan/Research Plan Template

For each topic, a preliminary draft of the work plan following this template is circulated prior to the first topic conference call with the Task Force leads, EPC review team, and AHRQ Medical Officer. This document is then revised based on the initial topic call, and the research plan (section IV of this template) is posted for public comment. After public comment, the entire work plan is finalized by the EPC review team and approved by the Task Force leads as the "final work plan." The final research plan is then posted on the Task Force Web site.

Instructions: This template is to be used for the final work plan, which includes the research plan for public posting. The text 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:

EPC Project Lead Investigator:

EPC Project Staff:

Task Force Leads:

Section I. Purpose and Background

Purpose

This report will be used by the USPSTF to:

What is the history of this Task Force recommendation (i.e., new, update from 1996, update from another 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 subpopulations? (If case distribution varies significantly by subpopulation, consider whether background questions about high-prevalence groups need to be addressed here, and also whether [and how] issues related to prevalence in subpopulations will be handled in this review as part of the "scope" section below.)

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 below.)

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 below.)

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 below.)

Intervention/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 below.)

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? (A table showing the recommendations of other groups should be included here, detailing any recommendations for specific subpopulations.) (If current practice issues seem important, consider whether background questions about current clinical practice need to be addressed here.)

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 subpopulations:

Do a disproportionate number of clinical cases come from one or a few subpopulations?

Natural history:

Is there "pseudo-disease" present among the apparently diseased population?
How long does it take for latent disease to become symptomatic?
How is the disease understood to come about, from a physiological perspective? What organs or systems are involved and what normal functions or mechanisms are aberrant when the disease is present?

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? (This is the "discriminatory" value of a risk calculation. Even though a set of risk factors may double or triple risk,, these risk factors may still be poor at discriminating those persons who will have the clinical condition from those who will not if the initial risk is small. If risk factors do not discriminate, they may have little clinical use.)

Detection/screening:

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

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 synthesizing the literature?

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 that are presumed to be in effect for the treatment to be effective?

Section II. Previous Review and USPSTF Recommendations

(This section is applicable to update topics only.)

Previous Task Force Recommendation(s)

"In (year of recommendation), the Task Force concluded (statement of the evidence) to recommend (recommendation statement). (Recommendation grade)"

Include any subpopulation-specific recommendations. Also list all the recommendation language in any 1996 or earlier Task Force topics, even if not a graded recommendation statement.

Previous Task Force Conclusions

List all conclusions made by the Task Force in the prior recommendation and rationale statement, including any evidence gaps identified by the Task Force.

Previous Analytic Framework and Key Questions

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

Previous Review Findings

Insert summary of evidence table with overall quality assessment for each key question.

Previous Review Conclusions

List all conclusions made by the authors of the prior evidence review. 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

Identify limitations cited in the prior evidence review and/or recommendation statement. Identify and list scope or method limitations identified 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," and other specific topics (such as "Potential Preventable Burden," "Current Practice").

Section III. Scan of Evidence Since Previous Recommendation

Existing Synthesized Evidence

Organize, summarize, and cite new evidence 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. Clearly delineate how systematic reviews have handled subpopulation considerations.

Previously Identified Pending Studies

Provide followup data on previously identified pending studies.

Newly 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 evidence review. Which key questions in the analytic framework will be addressed? Which key questions will not be updated, as their evidence is viewed as "established"? Are there areas that will be updated nonsystematically (i.e., contextual questions)?

The analytic framework, key questions, contextual questions, and inclusion/exclusion criteria will be copied from this section into the research plan template for public comment posting, so it should be written with that purpose in mind.

Analytic Framework and Key Questions

Analytic framework. Insert the analytic framework.

Key questions. Insert key questions.

Contextual questions. Insert contextual questions (if applicable).

Scope of Review

Using the inclusion/exclusion criteria table template (**Appendix Table 1**), specify who and what will be addressed in terms of populations, screening and treatment interventions, comparisons, outcomes, setting, study design, and quality. Other categories for which inclusion/exclusion criteria may be defined include study aim, disease/condition, timing of outcome assessment, intervention duration, publication date, and language. The descriptions below may help guide completion of the table. An introductory paragraph describing the general inclusion/exclusion criteria may be included.

Appendix Table 1. Inclusion and Exclusion Criteria

Category	Inclusion	Exclusion
Populations		
Interventions		
Comparisons		
Outcomes		
Setting		
Study Design		
Study Quality		

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

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 the 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 the 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? What role does quality assessment play in study eligibility?

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

Study quality. Specify that fair- and good-quality studies based on USPSTF criteria will be included and poor-quality studies will be excluded.

Exclusions

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.

Data Analysis

Briefly describe how, in addition to qualitative synthesis, quantitative synthesis will generally be conducted.

Timeline

The timeline is negotiated with AHRQ after work plan conference calls and is submitted with the final work plan as part of the deliverable. The timeline includes major milestones, including expected date for presentation at a future Task Force meeting. Since the length of the research plan review process may affect the timeline, the timeline should not be included until the final work plan is submitted.

Use of Outside Experts

Provide information on peer review of the work plan, if planned in addition to public comment, as appropriate. If known, describe the use or nonuse of previous review team members to provide continuity.

References

Section V. Research Plan

Use the following template for the research plan. The analytic framework, key questions, contextual questions, and inclusion/exclusion criteria should be copied from the appropriate section above. The “Response to Public Comment” section is only included in the final research plan.

USPSTF Draft Research Plan

Insert title of project.

Proposed Analytic Framework

Insert analytic framework.

Proposed Key Questions to Be Systematically Reviewed

Insert key questions.

Proposed Contextual Questions

“Contextual questions will not be systematically reviewed and are not shown in the Analytic Framework.”

Insert contextual questions.

Proposed Research Approach

“The Proposed Research Approach identifies the study characteristics and criteria that the Evidence-based Practice Center will use to search for publications and to determine whether identified studies should be included or excluded from the Evidence Review. Criteria are overarching as well as specific to each of the key questions.”

Insert inclusion/exclusion table from work plan.

Response to Public Comment

“The draft Research Plan was posted for public comment on the USPSTF Web site from [date] to [date].”

Insert a summary of the comments received by the USPSTF, how they were addressed in revisions to the research plan, and/or how they will be addressed during preparation of the systematic review.

Appendix VI. Criteria for Assessing Internal Validity of Individual Studies

The USPSTF Methods Workgroup 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, 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 exclusion criteria 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 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 team justifies such exclusion decisions if there could be reasonable disagreement about this step. These criteria are meant for those studies that pass this initial filter.

Presented below are a set of minimal criteria for each study design and 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. Recognizing that the methodology of systematic reviews are continuously evolving, the USPSTF allows the EPC to use newer methods of assessing quality of individual studies.

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.

Systematic Reviews

Criteria:

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

Definition of ratings based on above criteria:

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 above criteria:

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; accurate diagnostic procedures and measurements applied equally to cases and controls; and appropriate attention to confounding variables

Fair: Recent, relevant, and without major apparent selection or diagnostic workup bias, but response rate less than

80 percent or attention to some but not all important confounding variables

Poor: Major selection or diagnostic workup bias, response rate less than 50 percent, or inattention to confounding variables

RCTs 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 followup or overall high loss to followup
- 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 (followup ≥80%); reliable and valid measurement instruments are used and applied equally to all groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, intention-to-treat analysis is used for RCTs.

Fair: Studies are 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 followup; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is used for RCTs.

Poor: Studies are 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 equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. Intention-to-treat analysis is lacking for RCTs.

Diagnostic Accuracy Studies

Criteria:

- Screening test relevant, available for primary care, and adequately described
- Credible reference standard, performed regardless of test results
- Reference standard interpreted independently of screening test
- Indeterminate results handled in a reasonable manner
- Spectrum of patients included in study
- Sample size
- 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; assesses reliability of test; has few or handles indeterminate results in a reasonable manner; includes large number (>100) of 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; has moderate sample size (50 to 100 subjects) and a “medium” spectrum of patients

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

Appendix VII. Criteria for Assessing External Validity (Generalizability) of Individual Studies

Each study that is identified as providing evidence to answer a key question is assessed according to its external validity (generalizability), using the following criteria.

Study population: The degree to which a study's subjects constitute a special population—either because they were selected from a larger eligible population or because they do not represent persons 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 persons because of the inclusion/exclusion criteria, nonparticipation, or other reasons.
- Harms: The harms observed in the study could be greater or less than what might be expected in asymptomatic persons.

The following are features of the study population and the study design that may cause a participant's experience in the study to be different from what would be observed in the U.S. primary care population:

- Demographic characteristics (i.e., age, sex, ethnicity, education, income): The criteria for inclusion/exclusion or nonparticipation do not encompass the range of persons who are likely to be candidates for the preventive service in the U.S. primary care population.
- Comorbid conditions: The frequency of comorbid conditions in the study population does not represent the frequency likely to be encountered in persons 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 not representative of the U.S. primary care population.
- Refusal rate (i.e., ratio of included to not included but eligible participants): The refusal rate among eligible study subjects is high, making the study population not representative of the U.S. primary care population, even among eligible enrollees.
- Adherence (i.e., run-in phase, frequent contact to monitor adherence): The study design has features that may increase the effect of the intervention in the study more than would be expected in a clinically observed population.
- Stage or severity of disease: The selection of subjects for the study includes persons at a disease stage that is earlier or later than would be found in persons who are candidates for the preventive service.
- 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.

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

- Health care system: The clinical experience in the system in which the study was conducted is not likely to be the same as that experienced in other systems (e.g., 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 that in the United States (e.g., services available in the United States are not widely available in the other country or vice versa).
- Selection of participating centers: The clinical experience in which the study was conducted is not likely to be the same as in offices/hospitals/settings where the service is delivered to the U.S. primary care population (e.g., the center provides ancillary services that are not generally available).
- 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.

Study 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: Providers in the study are given special training not likely to be available or required in U.S. primary care settings.
- Expertise or skill to implement the intervention: Providers in the study have expertise and/or skills at a higher level than would likely be encountered in typical settings.

- Ancillary providers: The study intervention relies 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 U.S. primary care population/setting/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 U.S. primary care setting.

External validity is rated “fair” if:

- The study differs from the U.S. primary care population/setting/providers in a few ways that have the potential to affect the outcome in a clinically important way; it is moderately probable (50% to 89%) that the clinical experience with the intervention observed in the study will be attained in the U.S. primary care setting.

External validity is rated “poor” if:

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

Appendix VIII. Standardization of Outcomes Tables

Although it is not possible to completely standardize outcomes tables because of the diversity of issues among preventive services, the following are some preferred entries for authors to use in outcomes tables when possible.

Time frame:

- For services with an extended time frame: 10 years (5 years has typically been used, although 5 years is a short time for many consequences of screening and prophylactic interventions. Since these calculations require some assumptions, extrapolating to 10 years seems reasonable.). Alternatives: 5 years, lifetime.
- For services with a short time frame (e.g., pregnancy): 1 year.

Population:

- Express this number as per 1,000 persons targeted (e.g., per 1,000 women ages 40 to 49 years).
- Rationale: Preventive services with a large magnitude of effect should have substantial numbers of outcomes when expressed per 1,000 persons; those preventive services with less than one outcome averted will clearly be interpreted as having relatively small effect.

Interventions: Interventions should be shown in columns and described. For repeated services (e.g., annual fecal occult blood testing), the number of services should be identified.

Parameter estimates: Important parameters should be provided, as appropriate:

- Screening results (sensitivity, specificity)
- Prevalence of condition
- Adherence (to screening, treatment)
- Effectiveness
- Intermediate outcomes
- Number identified (with and without the condition)
- Number treated

Outcome measures (harms and benefits):

- Deaths (where relevant)
- Important health outcomes (e.g., strokes averted or cancers caused)
- Quality-adjusted life years (when possible)
- Adverse events/states

Number needed to screen/treat/counsel:

- Express in outcome terms (e.g., number needed to screen to avert one death).
- Number needed to counsel to achieve change in behavior should only be provided if it is also provided for a health outcome.

Appendix IX. Factors to Consider When Recommending Starting and Stopping Times for Screening

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 who are easily identified by one or more of the following factors: age, sex, ethnicity, particular behaviors (e.g., sexually active), and/or comorbid conditions 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.
4. Whether the feasibility, efficacy, and/or harms of treatment of the risk factor or target condition differ in particular subgroups.
5. Whether available research on the items above is limited to particular subgroups, especially if there is biological or epidemiological knowledge suggesting that the risk for disease, the accuracy of the screening test, and/or the efficacy of the treatment may vary significantly across a particular subgroup.

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

1. Whether the attributable risk and potential burden of the targeted condition is absent or decreased significantly in subgroups who are easily identified by one or more of the following factors: age, sex, ethnicity, particular behaviors (e.g., not sexually active), and/or biological or physical factors (e.g., surgical removal of the 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 available screening tests differs or is uncertain in particular subgroups.
4. Whether the feasibility, efficacy, and/or harms of treatment of the risk factor or target condition differ in particular subgroups.
5. Whether available research on the items above is limited to particular subgroups, especially if there is biological or epidemiological knowledge suggesting that the risk for 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 X. Factors to Consider When Recommending Screening Intervals

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

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 persons assigned to different screening intervals.

Appendix XI. Tool to Assess Adequacy and Certainty of Evidence for a Task Force Recommendation

	Population or Group 1	Population or Group 2
Detection ▪ **KQ 2.** What is the test performance of screening instruments to detect this disorder in community-dwelling primary care patients?	KQ2: *(convincing, adequate, or inadequate)*	KQ2: *(convincing, adequate, or inadequate)*
Benefits – KQ Evidence ▪ **KQ 1.** Does screening for this disorder in adults in primary care relevant settings improve outcomes compared to later diagnosis? ▪ **KQ 4.** Does early treatment of this screen-detected disorder improve outcomes compared to later treatment?	KQ1: KQ4: *(convincing, adequate, or inadequate)*	KQ1: KQ4: *(convincing, adequate, or inadequate)*
Benefits – Linkage Coherence	*(convincing, adequate, or inadequate)*	*(convincing, adequate, or inadequate)*
Benefits – Magnitude	*(substantial, moderate, small, zero, or cannot be determined)*	*(substantial, moderate, small, zero, or cannot be determined)*
Harms – Evidence ▪ **KQ 3.** What are the harms of screening for this disorder? ▪ **KQ 5.** What are the harms of early treatment for this disorder?	KQ3: KQ5: *(convincing, adequate, or inadequate)*	KQ3: KQ5: *(convincing, adequate, or inadequate)*
Harms – Linkage Coherence	*(convincing, adequate, or inadequate)*	*(convincing, adequate, or inadequate)*
Harms – Magnitude	*(substantial, moderate, small, zero, or cannot be determined)*	*(substantial, moderate, small, zero, or cannot be determined)*
Overall Certainty	*(high, moderate, low)*	*(high, moderate, low)*
Magnitude of Net Benefit *(Net Benefit = Benefits – Harms)*	*(substantial, moderate, small, zero, negative, cannot be determined)*	*(substantial, moderate, small, zero, negative, cannot be determined)*
GRADE	*(A, B, C, D, or I)*	*(A, B, C, D, or I)*

Appendix XII. Summary of Evidence Table for Evidence Reviews

The approach to the summary of evidence for the USPSTF should transparently represent the body of evidence at the key question level and support the application of the USPSTF's six critical appraisal questions to determine the adequacy of the evidence (convincing, adequate, or inadequate).

Summary of evidence tables created by different EPC teams for the USPSTF should be consistent in the methodological assessment of the body of evidence and the definitions of the information displayed; however, the format of the content may vary by the first, second, and subsequent stratification approaches required for a specific body of evidence (**Appendix Table 2**).

Appendix Table 2. Summary of Evidence Table

A. Key Question	B. Separate Populations or Interventions (1st-order stratification)	C. No. of Studies (k), No. of Participants (n), Study Design (2nd-order stratification)	D. Summary of Findings by Outcome (3rd-order stratification, if needed)	E. Consistency/ Precision	F. Reporting Bias	G. Overall Risk for Bias/ Quality	H. Body of Evidence Limitations	I. EPC Assess-ment of Strength of Evidence for Key Question	J. Applicability

A. Summary of evidence tables are organized by **key question** to reflect the linkages in the analytic framework.
B. Within the key questions, it can be most informative to stratify the body of evidence by **subpopulation** (e.g., by age or clinical group, such as pregnant women) or **type of intervention** (e.g., psychotherapy, specific medications), depending on the topic. This choice should not be rote, but should reflect the way the USPSTF has conceptualized the topic and key questions; the EPC should also consider the most informative approach for summarizing the available evidence given consistency and applicability issues within the body of evidence. The first-order stratification will generally result in a separate row for the entire subbody of evidence for that key question, particularly when the stratified data may be the basis for considering a subpopulation-specific recommendation or clinical considerations.
C. Within the first-order stratification, it may be necessary to organize the body of evidence by a second-order variable, such as **type of intervention** or **study design** (e.g., RCT vs. observational study). The number of studies (k) and number of participants (n) for each study design should be described within this level of stratification.
D. There may be a requirement for a third-order stratification, most likely for large bodies of evidence with pooled data available for different **types of outcomes**. To the degree made possible by the body of evidence, this summary should display the quantitative findings (pooled point estimates with 95% confidence intervals, heterogeneity measures, and predictive intervals, if warranted) or qualitative findings for each important outcome, with some indication of its variability. For qualitative or quantitative summaries at the outcome level, the number of contributing studies, number of events, and the combined sample size should also be reported. The consideration of the strength of evidence for the key question should be outcome-specific when multiple critical outcomes are measured and differ in any of the following domains (i.e., consistency/precision, reporting bias, overall risk for bias/quality).
E. **Consistency** is the degree to which contributing studies estimate the same direction of effect (i.e., consistently suggest benefit or harm); when there is consistency, confidence intervals overlap and statistical tests suggest low heterogeneity. Consistency can be rated as reasonably consistent, inconsistent, or N/A. Inconsistent results may indicate subgroup effects. **Precision** is the degree to which contributing studies estimate the same magnitude of effect (i.e., precisely suggest the magnitude of benefit or harm); when there is precision, point estimates are close and confidence intervals are narrow, without concerns about insufficient sample size, low event rates, or estimates that could suggest different clinical actions would be appropriate at the upper and lower bounds of the confidence interval. Precision can be rated as reasonably precise, imprecise, or N/A. Imprecise results may suggest the need for further research.
F. **Reporting bias** is the degree to which contributing studies may be limited by publication bias, selective outcome reporting bias, or selective analysis reporting bias. Reporting bias can be difficult to document (suspected, undetected, or N/A).
G. Within the appropriate level of stratification, a combined summary of individual study (or outcome-specific) **quality assessments** (or **risk for bias**) should be presented as good, fair-to-good, fair, fair-to-poor, or N/A (for no evidence). While the overall USPSTF quality rating occurs at the individual study level, EPC teams consider that threats to validity may apply differently to benefits and harms in the same study. Outcome-specific threats to validity may be reported when there are sufficient data and outcomes are of critical importance.

H. Important **limitations in the body of evidence** from what is desired to answer the overall key question are qualitatively described so the USPSTF might keep them in mind. These limitations might represent issues that led to low individual- or outcome-level study quality, such as concerns about populations selected and whether they adequately address racial/ethnic or other vulnerable subpopulations, lack of replication of interventions, or nonreporting of patient-important outcomes.

I. Using definitions from the EPC Program, the EPC provides a tentative **strength of evidence assessment** for each stratum for internal use by the USPSTF in its independent process of assessing the evidence. Strength of evidence assessments are labeled with the assessed grade (high, moderate, low, or insufficient), followed by language from the grade's definition (**Appendix Table 3**) that describes the critical appraisal issues leading to that grade. For example, a "high" strength of evidence assessment may state: "We are moderately confident that the estimate of effect lies close to the true effect; however, the body of evidence is still fairly small and not broadly representative of primary care settings, so some doubt remains."

J. **Applicability** is a descriptive assessment of how well the overall body of evidence would apply to the U.S. population based on settings; populations; and intervention characteristics, including accessibility, training, or quality assurance requirements.

Appendix Table 3. EPC Strength of Evidence Grades and Definitions

Grade	Definition
High	**We are very confident that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has few or no deficiencies. We believe that the findings are stable (i.e., another study would not change the conclusions).
Moderate	**We are moderately confident that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
Low	**We have limited confidence that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
Insufficient	**We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome.** No evidence is available or the body of evidence has unacceptable deficiencies, precluding us from reaching a conclusion.

Appendix XIII. Template for Scoping the "Decision Problem" to Address Through Decision Modeling

Date:

Prepared by:

Version:

Rationale: A priori articulation of the decisions to be addressed and their desired components is considered a best practice before constructing a model. Equally important is a clear conceptualization of the disease process(es) that must be modeled in order to make the desired modeling-based decisions. These interrelated issues establish the framework for locating relevant existing models, selecting modeling groups for commissioned models, and preparing decision modeling work plans by modeling groups commissioned to undertake modeling for use by the USPSTF.

1. What aspect(s) of evidence-based decisionmaking for the recommendation statement will require modeling?
   - Screening program details, such as stop and start times or screening intervals
   - Comparison of available modalities to determine equivalent or best approaches
   - Specification of high risk or other population selection approaches
   - Specification of net benefit for interventions with multiple, varying benefits and harms
   - Other (please specify)
2. Denote critical PICOTS inclusion or exclusion criteria to help determine the relevance of existing models (**Appendix Table 4**):
   - **Population:** Consider geography, patient characteristics (including sociodemographic information and comorbid conditions), disease stages, spectrum of disease, and other factors (family size, family impacts); important subpopulations and the characteristics that define them
   - **Setting:** In which countries would models based on their practices and policies be relevant to the United States?
   - **Intervention/Comparison/Strategy:** Consider base case and how well it represents current U.S. practice and policy; important alternative strategies (and their variations) that should be addressed, critical service components (including services that might precede or follow the intervention or strategy and affect its effectiveness or assumptions of intervention quality), and any service variations required for important subpopulations
   - **Health outcomes:** Deaths, quality-adjusted life years, disability-adjusted life years, incident disease cases, and disease-related health events (benefits and harms)
   - **Time horizon:** Minimal time horizon to capture relevant differences across strategies (this may or may not be critical at the scoping stage)
   - **Intermediate outcomes:** Intermediate disease or process outcomes necessary for validation or to determine net impact, such as diagnostic/treatment burden (these may or may not be critical at the scoping stage)
   - **Perspective of analysis and policy:** Perspective of analysis (i.e., medical sector, societal, health plan) undertaken, funder of model, original intended audience, and original development purpose (these may or may not be critical at the scoping stage)
3. Based on items 1 and 2, provide a clear written statement of the decision problem/objective and scope that includes disease spectrum, analysis perspective, target population, alternative interventions, health and other outcomes, and time horizon.
4. If possible, state the key questions that the desired model would address.

Appendix Table 4. PICOTS Table

Category	Inclusion Criteria	Exclusion Criteria
Condition definition		
Model approach		
Population		
Interventions/strategies		
Comparators/strategies		
Health outcomes		
Intermediate outcomes		
Time horizon		
Model perspective		
Setting		

Appendix XIV. Decision Framework to Assess and Guide the Need for Searches of Existing Decision Models

Background

Comprehensively identifying and evaluating models can be a time consuming process, particularly if it is not carefully structured in terms of process and timing. At the point that the USPSTF topic leads have determined a model is likely to be needed, a well-structured, noncomprehensive scanning process to identify existing models may prove effective as well as efficient.

Objectives

Define a topic-specific scanning process to identify readily available, published decision modeling studies and to determine their apparent relevance to the USPSTF recommendation statement being developed.

Outline of Process

1. The USPSTF topic leads, AHRQ Medical Officer, and EPC team (together comprising the topic workgroup) work through questions about the rationale for incorporating decision modeling, considering the extent of the prior USPSTF recommendation, public response to the prior USPSTF recommendation, prior use of modeling, current issues for clinical practice, and current state of the science.
2. The USPSTF topic leads scope the decision problem to be addressed by modeling, briefly articulating the modeling objective (for the purposes of the recommendation statement), the scenarios to be modeled, required outcome measures, and modeling time horizon needed to guide the search for existing models. The main constructs of relevance should be defined, including a time frame to define how recent models should be, and any setting considerations, including policy and practice context. If there are acceptable constraints on modeling approaches (i.e., perspectives of the analysis, type of model, or other), these should be specified. The optimal timing for this activity may vary by topic, as these components of the decision problem become clear.
3. A draft template to guide the articulation of the key questions, model objectives, and model scope is attached. This template can also serve to articulate the approach when commissioning a new or adapted model.
4. An appropriate party (to be determined*) conducts a scanning search in MEDLINE to identify readily available existing recent models. Assessing the quality and completeness of existing models is beyond this scanning exercise.
    a. Based on critical PICOTS factors, are there existing relevant models?
    b. Do these models address the key questions/decision needs for this recommendation statement?
    c. Is there more than one modeling group represented by the existing relevant models?
5. A summary document is prepared and distributed to the topic leads and resource allocation decision-makers to further inform the availability of relevant models and a potential approach (i.e., use existing model or models as they are, commission USPSTF-specific modeling from existing groups, or commission de novo modeling), if modeling is determined to be a priority in addition to the systematic review.

*The appropriate party needs to be determined based on timing and extent of work required for the scanning activity. This may be the topic-specific EPC team, a separate decision modeling support team, or a commissioned modeling group. If decision modeling is already commissioned, International Society For Pharmacoeconomics and Outcomes Research best practices suggest that the modelers search for previously published modeling analyses of the same or similar problems, in order to discuss their model with respect to others. Having commissioned modelers take on this scanning activity would also support this best practice.

Appendix XV. Specialist Expert Reviewer Nomination and Selection

The USPSTF values the opinions and input of specialists who are experts on the topics undergoing systematic review. The USPSTF has implemented a process for conducting outreach to specialty organizations to ask for nominations of expert peer reviewers of systematic evidence reviews. The organizations are solicited for nominations of scientific reviewers when the draft research plan is posted for public comment. They are asked to include a brief explanation of how the nominated individual meets the USPSTF's criteria and a current curriculum vitae. All supporting materials and conflicts of interests are reviewed by the USPSTF Scientific Program.

Once a nominee is selected as a scientific expert reviewer, they are asked to provide scientific feedback on the draft systematic evidence review. Reviewers are requested to provide feedback based on their individual opinions and expertise, not on behalf of the organization that nominated them. This does not preclude organizations from also submitting comments about the draft evidence review, as they have the opportunity to comment during the 4-week public comment period.

All expert reviewers are required to submit a conflict of interest form and sign a nondisclosure agreement. The draft systematic evidence review is considered confidential and should not be shared. All expert reviewers are given the option to be acknowledged as a reviewer in the draft systematic evidence review.

Organization Criteria

Organizations need to meet the following criteria to be included in this process:

- Based in the **United States**
- Operate on a **national** level (i.e., not a state or community level)
- Issue **preventive** guidelines (i.e., not treatment guidelines)

Note: USPSTF Dissemination and Implementation Partners and Federal Liaisons already review the systematic evidence review; they do not need additional outreach.

Expert Reviewer Criteria

Organizations should consider the following criteria when nominating expert reviewers:

- Experience in evidence synthesis and skills in evidence-based medicine
- Content expertise in prevention, screening, diagnosis, and treatment of the topic
- Specific expertise in critical aspects of the field such as populations at increased risk, evaluation of large clinical trials, and risk stratification
- Willingness to disclose any significant conflicts of interest and any preconceived position that would prevent objective review
- Familiarity with USPSTF methods