Chapter 2. Methods

Scope and Purpose

This systematic review will provide updated evidence regarding the effectiveness of one-time and repeated screening for AAAs, the associated harms of screening, and the benefits and harms of available treatments for small AAAs (aortic diameter of 3.0 to 5.0 cm) identified through screening. The USPSTF will use this review to update its 2014 recommendation for primary care practices.[93] This review included all trials from the previous review[105] that met current inclusion and exclusion criteria as well as newly identified studies.

Key Questions and Analytic Framework

Using the USPSTF's methods (detailed in **Appendix A**), we developed an analytic framework (**Figure 1**) and five Key Questions (KQs).

The KQs include:

1. What are the effects of one-time screening for AAA on health outcomes in an asymptomatic population age 50 years or older?
   a. Do the effects of one-time screening for AAA vary among subpopulations (i.e., by age, sex, smoking status, family history, or race/ethnicity)?
2. What are the effects of rescreening for AAA on health outcomes or AAA incidence in a previously screened, asymptomatic population without AAA on initial screening?
   a. Do the effects of rescreening for AAA vary among subpopulations (i.e., by age, sex, smoking status, family history, or race/ethnicity)?
   b. Do the effects of rescreening for AAA vary by the time interval between screenings?
3. What are the harms of one-time and repeated screening for AAA?
   a. Do the harms of one-time and repeated screening for AAA vary among subpopulations (i.e., by age, sex, smoking status, family history, or race/ethnicity)?
4. What are the effects of treatment (pharmacotherapy or surgery) on intermediate and health outcomes in an asymptomatic, screen-detected population with small AAAs (i.e., aortic diameter of 3.0 to 5.4 cm)?
   a. Do the effects of treatment of small AAAs vary among subpopulations (i.e., by age, sex, smoking status, family history, or race/ethnicity)?
5. What are the harms of treatment in an asymptomatic, screen-detected population with small AAAs (i.e., aortic diameter of 3.0 to 5.4 cm)?
   a. Do the harms of treatment of small AAAs vary among subpopulations (i.e., by age, sex, smoking status, family history, or race/ethnicity)?

Data Sources and Searches

In addition to considering all studies from the previous review for inclusion in the current review, we performed a comprehensive search of MEDLINE, PubMed (publisher supplied only), the Database of Abstracts of Reviews of Effects, and the Cochrane Central Register of Controlled Trials for studies published between January 2013 and September 14, 2018. A research librarian developed and executed the search, which was peer reviewed by a second research librarian (**Appendix A**).

In addition, we examined the reference lists of other previously published reviews, meta-analyses, and primary studies to identify additional potential studies for inclusion. We supplemented our searches with suggestions from experts and articles identified through news and table-of-contents alerts. We also searched ClinicalTrials.gov (https://ClinicalTrials.gov/) for ongoing trials (**Appendix H**). We imported the literature from these sources directly into EndNote® X7 (Thomson Reuters, New York, NY).

Study Selection

Two reviewers independently reviewed the titles and abstracts of all identified articles to determine whether studies met inclusion and exclusion criteria for design, population, intervention, and outcomes (**Appendix A Table 1**). Two reviewers then independently evaluated the full-text article(s) of all potentially included studies against the complete inclusion and exclusion criteria. Disagreements in the abstract and/or full-text review were resolved by discussion and consultation with a third reviewer if necessary. Excluded studies and reasons for exclusion are listed in **Appendix D**.

We developed an a priori set of criteria for inclusion and exclusion of studies based on our understanding of the literature (**Appendix A Table 1**). For KQ1 and KQ2, examining the effectiveness of one-time and repeated screening, we considered RCTs and large cohort studies (n ≥1,000) of asymptomatic adult populations. For KQ4, examining the effectiveness of treating small AAAs, we considered only RCTs of asymptomatic adult populations with AAAs identified as being small (aortic diameter of 3.0 to 5.4 cm). For KQ3 and KQ5, examining the harms of screening for and treating small AAAs, we were more inclusive and considered RCTs, observational studies, and registry data related to surgical harms. For KQ5, we considered only adult populations with asymptomatic, small aneurysms. For all KQs, the only screening modality we considered was ultrasound. We did not consider physical examinations due to literature reporting unfavorable sensitivity and specificity of this diagnostic method.[87] Further, we did not consider CT or magnetic resonance imaging (MRI) screening, as these modalities are not readily available in primary care. For KQ2, we accepted targeted screening defined as screening based on one or more patient risk factors or screening based on prediction/prognostic modeling. For KQs related to the treatment of small AAAs, we considered surgical intervention (open repair or EVAR) or pharmacotherapeutic interventions (statins, angiotensin-converting enzyme [ACE] inhibitors, beta-blockers, or antibiotics) compared to surveillance, usual care, or placebo. We limited our included studies to those published in English and those that were deemed good- or fair-quality by using items from the Newcastle-Ottawa Scale[106] and USPSTF quality rating

standards.[107] The outcomes that were reviewed are fully listed in **Appendix A Table 2**.

Quality Assessment and Data Abstraction

Two reviewers applied USPSTF design-specific criteria and items from the Newcastle-Ottawa Scale[106, 107] to assess the methodological quality of all eligible studies (**Appendix A Table 2**). We assigned each study a quality rating of "good," "fair," or "poor." Discordant quality ratings were resolved by discussion or by a third reviewer and adjudicated as needed. Studies rated as poor quality were excluded from the review.

Good-quality RCTs were those that met all or nearly all of the specified quality criteria (e.g., comparable groups were assembled initially and maintained throughout the study, followup was 90% or higher, assessment procedures were described and blinded if they involved direct interview, randomization methods were described, and allocation was concealed), whereas fair-quality studies did not meet all these criteria but did not have serious threats to their internal validity related to design, execution, or reporting. Intervention studies rated as poor quality generally had several important limitations, including at least one of the following risks of bias: very high attrition (generally >40%); differential attrition between intervention arms (generally >20%); lack of baseline comparability between groups without adjustment; or problematic issues in trial conduct, analysis, or reporting of results (e.g., possible selective reporting, inappropriate exclusion of participants from analyses, or questionable validity of allocation or assessment procedures).

Good-quality observational studies had an unbiased selection of the nonexposed cohort and adequate ascertainment of exposure. These studies addressed a population without the outcome of interest at the beginning of the study, and they had reliable outcome measures, blinded assessment, low attrition, adjustment for potential confounders, and no other important threats to internal validity. Observational studies were downgraded to fair if they were unable to meet the majority of good-quality criteria. Poor-quality observational studies had multiple threats to internal validity and were excluded from the review.

One reviewer extracted data from all included studies rated as fair- or good-quality directly into summary tables (Microsoft Word®, Microsoft Corporation, Redmond, WA), and a second reviewer checked the data for accuracy. Elements abstracted included population characteristics (e.g., baseline demographics, concurrent conditions, family history of AAA, smoking status, and CVD risk factors), as well as study design elements (e.g., recruitment procedures, inclusion/exclusion criteria, followup and population adherence), intervention characteristics (including postscreening management), and relevant outcomes.

Health outcomes included the number of participants experiencing an event and incidence rates. For KQs 1, 2, and 3 (efficacy and harms of screening), we abstracted the reported incidence and prevalence of AAA, incidence of ruptured aneurysms, and mortality (all-cause, AAA-related, and operative mortality). In addition, we extracted information on the number and circumstance (i.e., emergency or elective) of surgical interventions reported in each study and any adverse events related to screening (e.g., changes in quality of life or anxiety) that were reported. In our

previous review, we included the shorter-term outcomes from the screening trials, but in the current report we only report outcomes from the longest-term followup. We also did not include the Viborg Vascular (VIVA) trial mortality data due to our inability to capture the independent contribution of AAA screening within the multicomponent screening program. We did include AAA prevalence and data related to AAA procedures from this trial. For KQs 4 and 5 (efficacy and harms related to treating small AAAs), we abstracted data related to the dose and duration of the pharmaceutical intervention, surgical details (if reported), AAA growth rate, the number and circumstance (i.e., emergency or elective) of surgical interventions, incidence of aneurysm rupture, and mortality (all-cause, AAA-related, and operative mortality). For adverse events, we extracted all that were reported but specifically looked for incidences of reinterventions, endoleaks, device migration, conversion to open surgery, and hospital readmission within 30 days of surgery.

Data Synthesis and Analysis

We synthesized data separately for each KQ. Specifically, we provide a narrative summary of the included studies regarding study design and setting, internal validity and major factors threatening the interval validity, and important characteristics about patients and interventions.

For KQ1, we examined all-cause mortality, AAA-related mortality, rupture, and emergency surgeries for the comparison of screening vs. no screening. We pooled calculated risk ratios and used the DerSimonian and Laird[108] random-effects model as the primary analysis for all-cause mortality, since statistical heterogeneity was very low (I^2=0%; τ^2 = 0.0). Because of the relatively small number of trials being pooled, we also conducted a sensitivity analysis using the restricted maximum likelihood method, which tends to result in more conservative (larger) estimates of τ^2 when there are fewer than five to 10 trials being pooled. As expected, results showed slightly larger CIs and were consistent with the DerSimonian and Laird model with respect to statistical significance, so only DerSimonian and Laird analyses are reported. For AAA-related mortality, rupture, and emergency surgeries, we used the Peto method to pool ORs. ORs were calculated based on the numbers of events and participants in each study arm. The Peto method was chosen because events were very rare (occurred in <1% of participants in most study arms) and trials had a similar number of participants in both study arms.[109] All statistical testing was two-sided and we considered 0.05 as significant. We examined statistical heterogeneity across trials with the I^2 statistic and chi-square test of heterogeneity.

We did not conduct meta-analysis of the rescreening studies included in KQ2 because of substantial differences in patient population, length of followup, and outcomes reported. We provide a narrative summary of results and reported outcomes including incidence of large AAA, AAA ruptures, AAA procedure data, and AAA-related mortality and all-cause mortality.

To analyze the harms of screening vs. no screening in KQ3, we examined 30-day mortality after elective surgery, 30-day mortality after emergency surgery, overall operations, elective operations, emergency operations, and quality of life measures. The 30-day mortality after elective surgery and 30-day mortality after emergency surgery outcomes were not pooled since there were only two trials reporting these outcomes. The Peto method was used to pool overall

operations, elective operations, and emergency operations, as described under KQ1. All statistical testing was two-sided, and we considered 0.05 as significant. We examined heterogeneity across trials with the I^2 statistic and chi-square test of heterogeneity. Because of the substantial difference in quality of life measurements and insufficient reporting of data (e.g., lack of variation parameters), we were unable to pool these data in the studies of screening vs. no screening.

We narratively describe the treatment study results for KQs 4 and 5 and present the data in tables. We did not conduct meta-analysis of the treatment trials due to the small number of studies of each intervention type.

Subpopulation Methods

We prespecified subpopulations of interest in the KQs. These populations were selected based on analysis of subpopulation considerations in the previous review and recommendation, established characteristics associated with the development of AAA, and feedback received from three Key Informants during the scoping phase. During the data abstraction phase, we catalogued the availability and characteristics of subgroup analyses (i.e., whether analyses were a priori, post hoc, or unclear) for each subpopulation of interest for each trial and subsequently audited these results. Using this audit, formal subgroup analyses were prioritized based on the number of contributing studies and the credibility of subgroup analyses, with subpopulation-specific trials and a priori analyses given more weight. This process was aided by subpopulation credibility ratings conducted by our team based on the guidance from Whitlock et al.[110]

We then entered data from subgroup analyses into summary tables for the prioritized analyses of age, sex, smoking status, family history, and race/ethnicity. In addition to outcomes, subgroup summary tables included information relevant to the credibility of each trial's subgroup analyses, such as the interaction testing for heterogeneity of treatment effect if it were available. Direct evidence from within-study comparisons was emphasized over across-study comparisons, which can be confounded by differences in populations and their risk factors.

Based on a limited number of contributing studies for subgroup analyses, we did not pool results but analyzed them qualitatively.

Grading the Strength of the Body of Evidence

We graded the strength of the overall body of evidence for each KQ. We adapted the Evidence-based Practice Center approach,[111] which is based on a system developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group.[112] Our method explicitly addresses four of the five Evidence-based Practice Center required domains: consistency (similarity of effect direction and size), precision (degree of certainty around an estimate), reporting bias (potential for bias related to publication, selective outcome reporting, or selective analysis reporting), and study quality (i.e., study limitations). We did not address the fifth required domain—directness—as it is implied in the structure of the KQs (i.e., pertains to

whether the evidence links the interventions directly to a health outcome).

Consistency was rated as reasonably consistent, inconsistent, or not applicable (e.g., single study). Precision was rated as reasonably precise, imprecise, or not applicable (e.g., no evidence). Reporting bias was rated as suspected, undetected, or not applicable (e.g., when there was insufficient evidence for a particular outcome). Study quality reflects the quality ratings of the individual trials and indicates the degree to which the included studies for a given outcome have a high likelihood of adequate protection against bias. The body of evidence limitations field highlights important restrictions in answering the overall KQ (e.g., lack of replication of interventions or nonreporting of outcomes important to patients).

We graded the overall strength of evidence as high, moderate, or low. "High" indicates high confidence that the evidence reflects the true effect and that further research is very unlikely to change our confidence in the estimate of effects. "Moderate" indicates moderate confidence that the evidence reflects the true effect and that further research may change our confidence in the estimate of effect and may change the estimate. "Low" indicates low confidence that the evidence reflects the true effect and that further research is likely to change our confidence in the estimate of effect and to change the estimate. A grade of "insufficient" indicates that evidence is either unavailable or does not permit estimate of an effect. At least two independent reviewers rated the overall strength of evidence for each intervention type. We resolved discrepancies through consensus discussion involving more reviewers.

Expert Review and Public Comment

A draft Research Plan for this review was available for public comment from August 10, 2017, to September 6, 2017. In response to comments, the USPSTF expanded the scope of the evidence review to include cardiovascular events (e.g., myocardial infarction and stroke) and mortality related to CVD to more fully evaluate the benefits and harms of treatment of small AAAs (i.e., pharmacotherapy such as statins and antihypertension medications). The USPSTF made other minor modifications as appropriate, such as clarifying that surveillance alone would be included as a comparator for KQ 4. The draft version of this report was reviewed by experts and USPSTF Federal Partners and posted for public comment on the USPSTF website from June 25, 2019, to July 23, 2019. Comments received during any period were reviewed, considered, and addressed as appropriate; however, no changes were made to the evidence or to our conclusions.

USPSTF Involvement

We worked with four USPSTF members at key points throughout this review, particularly when determining the scope and methods for this review and developing the Analytic Framework and KQs. After revisions reflecting the public comment period, the USPSTF members approved the final Analytic Framework, KQs, and inclusion and exclusion criteria. The Agency for Healthcare Research and Quality funded this review under a contract to support the work of the USPSTF. An Agency Medical Officer provided project oversight, reviewed the draft report, and assisted in the external review of the report.