Evidence Based Benefit Review Advisory Committee: Methods Manual

March 2024

Center for Evidence-based Policy
Oregon Health & Science University
3030 S Moody, Suite 250
Portland, OR 97201
Phone: 503-494-2182
Fax: 503-494-3807
http://centerforevidencebasedpolicy.org/

Table of Contents

1. Purpose

This manual provides an overview of how the Center for Evidence-based Policy (Center) staff prepares and presents reports for consideration at meetings of the Evidence Based Benefit Review Advisory Committee (EBBRAC), which provides recommendations about coverage requests for potential benefits for New York's Medicaid program, in coordination with New York State Department of Health (the Department) staff. Sources of topics include the following:

- Requests for detailed evidence review from the Internal Benefit Review Committee (IBRC)
- Submissions to the ebbrac@health.ny.gov inbox monitored by the department staff (e.g., requests for coverage from manufacturers or members of the public)
- Health technology or service topics noted by the Department as timely to review given potential advances in evidence, current policy climate, or concerns about meeting emerging needs of the NYS Medicaid population

Proposed EBBRAC topics for consideration either represent a material change in coverage for the NYS Medicaid program or a new health technology assessment or medical evidence review. Center staff meets regularly with the Department staff to determine the scope of report topics, provide updates on report progress, and coordinate preparation for EBBRAC meetings to ensure that the findings present research, policy, and clinical practice guidelines to support EBBRAC decision making.

EBBRAC is tasked with the following[1]:

> The committee shall provide advice and make recommendations regarding coverage of health technology or service for purposes of the medical assistance program. The commissioner shall consult such committee prior to any determination made regarding the coverage status of a particular item, health technology or service based on procedures established in subdivision five of this section under the medical assistance program. For purposes of this section, "health technology" means medical devices and surgical procedures used in the prevention, diagnosis and treatment of disease and other medical conditions. For purposes of this section "services" means any medical or behavioral health procedure.

The credibility of the decisions made by EBBRAC depends on the transparency of their decision-making process. To that end, the purpose of this manual is to allow readers to understand how the evidence was gathered, assessed, and synthesized into the findings presented to the EBBRAC. The public deliberation provides transparency into how the EBBRAC weighed the evidence and other factors to arrive at their recommendations.

Chapter Synopses

The following sections provide high-level summaries of chapter contents.

2. Defining Research Questions and Developing the Scope Statement

Center researchers work with the Department to write key questions and detailed information about the population(s), intervention(s), comparator(s), and outcome(s) that guide the research process. Key questions typically address effectiveness and harms of a health technology through a systematic review of published clinical research. Additional key questions are addressed

through a review of clinical practice guidelines, specified Medicaid program coverage policies and private payer policies, and cost analysis studies relevant to a US context. Contextual questions may be selected to guide content for the Background section of the report and presentation. This chapter describes how scope statements are developed and what they typically include.

3. Outlining Key Milestones

The process used by Center researchers in coordination with input from the Department staff requires that topics be proposed a minimum of 6 months before the EBBRAC meeting at which the topic will be presented. This chapter gives an overview of key milestones in the research process from topic selection to report presentation.

4. Searching and Selecting Relevant Information

The Center researchers work with an information specialist to conduct searches for each topic. The searches are conducted across key online resources to identify information that addresses the scope statement's key questions. Center researchers review all of the identified information against the inclusion and exclusion criteria from the scope statement. This chapter describes the processes used to search for information, which resources are searched, and how researchers decide which information is included in the evidence report.

5. Assessing Risk of Bias

Clinical research is subject to bias, with many factors leading researchers to have more or less certainty about the findings of any individual study. Center researchers use standardized checklists to help understand what the level of risk of bias is for each included study and how this might influence the level of confidence in the study findings. This chapter describes how risk of bias is assessed and its implications for study findings.

6. Synthesizing Evidence

Center researchers summarize the relevant information and provide a synthesis of the evidence, including judgments about the overall certainty of a body of evidence, by outcome. This chapter describes the process of evidence synthesis, the different approaches that can be taken, and how Center researchers determine the overall certainty of a body of evidence.

7. Writing the Report

Center researchers draft a report for the EBBRAC comprising an Executive Summary, a full report, and appendices. This chapter describes the typical sections of EBBRAC reports and how the information for each section is presented.

8. Monitoring New Evidence and Updating Reports

Center researchers conduct targeted evidence, guideline, and policy searches on a rolling basis to assess whether new publications might affirm or change findings of prior reports.

9. Managing the Research Process

This chapter describes the Center's internal management of report documentation and content.

10. Using Evidence to Make Decisions

Clinical evidence from published studies represents only one type of information that EBBRAC members may consider when making a coverage recommendation. This chapter suggests other factors for EBBRAC members to consider while weighing the evidence, including clinical practice guidelines, policies, and other information presented in the report or at the presentation (e.g., public comment).

11. Updating the Manual

This chapter describes when Center staff updates this manual.

2. Defining Research Questions and Developing Scope Statements

After the Department staff proposes a topic for EBBRAC consideration, Center researchers draft a scope statement that describes the focus of the report by providing a brief background on the health technology, key questions to be addressed, and the structured PICO (population, intervention, comparator, and outcomes) used to guide searches, selection, and synthesis. In addition, the scope statement includes a table with detailed inclusion and exclusion criteria for the selection of relevant publications for the report and a reference list.

Center staff may consult with subject matter experts while drafting the scope statement for topics, as needed, to ensure the scope statement includes the most relevant clinical questions, the critical outcomes necessary for decision making, and that study inclusion and exclusion criteria are appropriate and justified.

The following sections describe the sections and content of a scope statement.

Background

Each scope statement includes a high-level description of the intervention in question, the clinical need and population, and other important considerations, depending on the topic. Relevant background considerations include the following:

- A brief description of the health technology and its role in care
- How the health technology was approved for use in the US (e.g., source of the data submitted with the application for approval, regulatory pathway)
- An overview of the epidemiology, prognosis, and current standard care for the population(s) or condition(s) for which the technology is being considered
- If appropriate, any known barriers to implementation of the health technology

Standard Key Questions for Health Technology Assessment Topics

KQ1. What is the clinical effectiveness of the health technology in the specified population or condition? (For some topics, this might be replaced with a question about comparative effectiveness.)

   a. Depending on the topic, does clinical effectiveness vary by patient characteristics (e.g., age, sex), disease characteristics (e.g., length of time since diagnosis), or other characteristics of interest (e.g., provider type, setting)?

KQ2. What are the harms of the health technology in the specified population or condition?

a. Depending on the topic, do harms vary by patient characteristics (e.g., age, sex), disease characteristics (e.g., length of time since diagnosis), or other characteristics of interest (e.g., provider type, setting)?

KQ3. What are the costs or cost-effectiveness of the health technology in the specified population or condition?

KQ4. What are the clinical practice guidelines for the health technology in the specified population or condition?

KQ5. What are relevant Medicaid program coverage policies and private payer policies for the health technology in the specified population or condition?

PICO: Population, Intervention, Comparator, and Outcomes

After the key questions have been determined, Center staff proposes detailed inclusion and exclusion criteria, including defining the PICO elements (population, intervention, comparator, and outcomes). The criteria are informed by the scoping work and discussions with the Department staff. The Center information specialist and researchers also use information from key sources to understand which populations are of interest, variations in the intervention and comparators, the outcomes viewed by professional societies and researchers as important, study designs used, and other important elements of published peer-reviewed studies. As the research process progresses, Center researchers may learn other important information, leading to scope clarifications or amendments to the inclusion and exclusion criteria. When this happens, Center researchers communicate proposed changes to the Department staff and document any agreed amendments in the scope statement change log. Appendix A has an example of detailed PICO inclusion and exclusion criteria.

Contextual Questions

Some topics benefit from additional context; in these cases, the Center researchers answer contextual questions in the background of the report. Typical contextual questions may address the following topics:

- Information on the current standard of care for the population or condition of interest
- Implementation considerations (e.g., shared decision making, accreditation standards, scope of practice issues, risk assessment)
- Acceptability, feasibility, and satisfaction of the health technology
- Equity issues, including how social determinants of health may affect access to the health technology

Systematic review methods are not used to answer contextual questions; however, Center researchers use other methods to identify and describe responses to the contextual questions. For example, this could include a summary of accreditation standards for a health technology of interest. However, studies cited in the contextual response are not assessed for risk of bias, and an overall judgment on the certainty of evidence (i.e., Grading of Recommendations, Assessment, Development, and Evaluation [GRADE]) is not provided.

References and Additional Sources

These sections list all sources cited in the scope statement, and they provide a list of additional sources identified during the scoping process that were not directly cited but may provide useful information to the Department staff or EBBRAC members.

Change Log

This table summarizes material changes made to the scope statement after initial approval by the Department staff, along with the date the decision was formalized.

3. Outlining Key Milestones

The following sections list important steps in the process between receipt of a potential EBBRAC report topic and finalization of the report.

Developing Key Questions and Scope Statement

- At least 6 months before the EBBRAC meeting where the topic will be presented, the Department staff shares topic of interest and any proposed PICO elements with Center staff.
- Center staff conducts preliminary searches and drafts a scope statement. Depending on the complexity of a topic, drafting a scope may require 2 to 4 weeks.

Topic Refinement and Selection

- The Department staff review the scope drafted by Center researchers, ask clarifying questions, and provide feedback that may result in edits to the scope statement.
- Center researchers take the Department staff's feedback under consideration and use the change log in the scope statement to track any changes made to the scope from this point onward.

Evidence Review and Report Writing

- Information specialist at the Center builds and executes search strategies.
- Center researchers screen and select eligible publications, as determined by the inclusion and exclusion table in the scope statement.
- Center researchers assess risk of bias for each included study to answer the key questions about clinical evidence, clinical practice guidelines, and cost.
- Center research team members abstract relevant data for critical and important outcomes.

- Center research team members write report, which is reviewed by a Center research director and the Center project leads.
- Center editor completes first full edit for consistent style and clarity of report content.
- Center staff shares full draft of report with the Department staff for review 7 weeks before the EBBRAC meeting.
- Center researchers incorporate edits to address feedback from the Department staff; a Center research director reviews the final report; and Center editor finalizes edits before sending to the Department staff.
- Final report is uploaded to Boardvantage, a platform EBBRAC committee members use to store and access meeting materials, 1 week before the EBBRAC meeting.

Report Presentation

- Public comment occurs during the EBBRAC meeting. Details for how to submit public comments can be found on the EBBRAC website.[2]
- Center staff members present report findings to EBBRAC members with support from Center research team.
- For certain topics (e.g., requiring explanation of technical procedures), the Department staff may invite a subject matter expert to respond to questions from EBBRAC members.
- The Department staff facilitates discussion of the report findings and leads a structured discussion, with the aim of achieving consensus on a recommendation for or against coverage of the health technology.

Report Finalization

- The Department staff uploads the final report and relevant EBBRAC meeting materials to the public-facing website.
- Topic is then added to a list for consideration for future surveillance and review of new relevant publications.

Documentation of Meeting Findings

- The Department staff finalizes and uploads minutes from the EBBRAC meeting to document meeting findings and recommendations from the committee.

4. Searching and Selecting Relevant Information

Center researchers use multiple sources and methods to identify clinical and economic evidence on the health technology or service of interest. Clinical evidence provides information about the efficacy and safety of the health technology or service while economic evidence provides information about the cost-effectiveness and affordability of the health technology or service.[3] Center researchers use systematic review methods, with modifications to accommodate an abbreviated timeline, to identify, critically appraise, and synthesize relevant clinical and economic evidence.[4] When conducting systematic reviews, Center researchers search multiple sources to find studies on the health technology, select studies according to predefined inclusion and exclusion criteria, and assess risk of bias for each study (see Chapter 5 for discussion of risk of bias assessment). This approach is designed to minimize bias in selecting studies to include for review and provide an accurate assessment of the body of evidence available on the health technology or service of interest. An essential component of systematic review methods is

preparing a transparent, complete, and accurate account of what was done and what was found.[5] To this end, Center researchers use the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to guide their documentation and reporting for all their systematic reviews.[5] The widely endorsed and adopted PRISMA statement consists of a detailed checklist, with accompanying explanation for each item, and study flow diagram template for reporting in systematic reviews.[5] A transparent process allows readers of the report to understand how evidence was selected, evaluated, and interpreted.[6,7]

Clinical Evidence Sources

Bibliographic Databases

For a systematic review, multiple bibliographic databases are searched to enhance overall retrieval of published studies.[4,6,8] A Center information specialist searches the following core bibliographic databases to identify published peer-reviewed studies, systematic reviews, meta-analyses, and clinical practice guidelines on the chosen topic:

- Ovid MEDLINE
- Cochrane Central Register of Controlled Trials (CENTRAL)
- Cochrane Database of Systematic Reviews (CDSR)

Depending on the topic, the Center information specialist may choose to search additional bibliographic databases. For example, if the chosen topic involves a mental health condition, then PsycINFO is also searched, or if the chosen topic involves an intervention delivered by nurses or other health professionals, then the Cumulative Index to Nursing and Allied Health Literature (CINAHL) is searched.

Other Sources

Searching bibliographic databases does not always retrieve all relevant information on a particular health intervention. There are a variety of possible reasons for this, such as a relevant study may have been published in a journal not indexed by the bibliographic databases that were searched, or a guideline produced by a professional organization has not yet been published in a journal. Information generated by government, academia, industry, and others outside traditional commercial publishing channels (also known as "gray literature") may not be indexed in a bibliographic database.[9,10] Therefore, other sources are searched to find information not retrieved in the searches of bibliographic databases.

Health Technology Assessments and Systematic Reviews

A Center information specialist searches the following sources to identify health technology assessments and systematic reviews not retrieved by searching bibliographic databases:

- Agency for Healthcare Research and Quality (AHRQ)
- Canada's Drug and Health Technology Agency (CADTH)
- Epistemonikos
- Health Quality Ontario
- Institute for Clinical and Economic Review (ICER)
- International Health Technology Assessment Database
- National Institute for Health and Care Excellence (NICE)
- Oregon Health Evidence Review Commission

- Veterans Affairs Evidence Synthesis Program
- Washington State Health Care Authority

Clinical Trial Registries

A Center information specialist searches ClinicalTrials.gov and ScanMedicine to identify ongoing and unpublished clinical trials. ClinicalTrials.gov is a website and online database of clinical research studies, maintained by the US National Library of Medicine. ScanMedicine is an online search system consolidating data from multiple clinical trial registries across the world.[11]

Center researchers use records retrieved from clinical trial registries to supplement the information reported in published studies (e.g., detailed study participant inclusion and exclusion criteria) and to monitor ongoing research on the health technology or service of interest.[12] Records from clinical trial registries are often incomplete and not regularly updated, so they may not provide enough information to assess the risk of bias for a study. Therefore, results of ongoing and unpublished studies provided as part of a clinical trial registry record are not routinely included in the evidence synthesis.

Clinical Practice Guidelines

Clinical practice guidelines are statements that include recommendations intended to optimize patient care, ideally informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.[13] Clinical practice guideline development involves both a technical process, selecting and appraising evidence, and a social process, translating evidence into recommendations.[14] During guideline development, expert opinion may be combined with empirical evidence or, in the absence of relevant research, be considered the best available evidence.[15] Ultimately, clinical practice guideline developers should consider the best available evidence, patient and physician values and preferences, and resource use when making recommendations.[15] Although several organizations and associations provide standards for creating guidelines,[13,16-18] such as using systematic review methods to select and appraise evidence and having a clear, transparent process for reaching group consensus, there is considerable variation in the quality of clinical practice guidelines.[14]

For each topic, a Center information specialist constructs a search strategy for Ovid MEDLINE to retrieve clinical practice guidelines on the condition and health technology or service of interest published within the past 5 years. As evidence-based clinical practice guidelines may take several years to develop, the 5-year time limit is imposed to ensure that the guidelines reflect the current clinical evidence and are relevant to clinical practice. In addition to Ovid MEDLINE, the following core resources are searched:

- American Medical Association
- Guidelines International Network (GIN) International Guidelines Library
- National Institute for Health and Care Excellence (NICE)
- Scottish Intercollegiate Guidelines Network (SIGN)
- Veterans Affairs/Department of Defense Clinical Practice Guidelines

GIN is a collaborative of guideline developers from around the world, consisting of organizational members, such as the American Academy of Neurology and American College of Physicians, and individual members.[17,19] The International Guidelines Library and registry contains links to

guidelines published or endorsed by GIN members, health guidelines from nonmember organizations, and guidelines in development.[19]

A Center information specialist also searches websites of professional organizations and special interest societies relevant to the chosen topic. For example, the American College of Obstetricians and Gynecologists is searched for topics related to maternal and perinatal health, and the American Academy of Pediatrics for topics related to child health. The American Psychiatric Association website is searched for topics related to mental health. For topics related to screening and prevention, the US Preventive Services Task Force website is searched. For topics related to cancer treatment, the websites of the American Cancer Society and National Comprehensive Cancer Network are searched. For topics related to heart disease, the websites of the American Academy of Family Physicians, American College of Cardiology, American Heart Association, and Heart Failure Society of America are searched.

Regulatory Bodies and Manufacturers

A Center information specialist searches regulatory sites and manufacturer websites to identify additional information on the intervention. For a topic that involves a drug, the US Food and Drug Administration (FDA) database of approved drugs (Drugs@FDA) is searched to retrieve documents relevant to the drug's approval and use (e.g., new drug application, drug label). Regulatory sites are also searched for reports of adverse events. For example, the FDA database MedWatch is searched for drug safety information and adverse event reports while the FDA Manufacturer and User Facility Device Experience (MAUDE) database is searched for safety information and adverse event reports related to medical devices. Manufacturer websites are searched for information about ongoing or unpublished studies and access to, and payment for, the intervention.

Additional Methods to Identify Clinical Evidence

Citation Chaining

Citation chaining, or snowballing, uses connections between similar research articles to find studies that may not have been retrieved in searches of bibliographic databases or other sources.[20] This technique is particularly useful for emerging or cross-disciplinary topics where terminology is not consistent. Citation chaining can refer to backward citation chaining (e.g., checking reference lists of included studies) or forward citation chaining (e.g., using Google Scholar's "cited by" feature to identify publications that cite an included study). Center researchers review reference lists of relevant systematic reviews, meta-analyses, and health technology assessments (backward citation chaining) and search for publications that have cited key included studies (forward citation chaining) to identify publications that may not have been found by searching the sources outlined above. Center researchers use various methods and tools to identify additional studies via citation chaining, such as Citationchaser,[20] an open-source tool, and Scopus, a bibliographic database, that allow users to rapidly identify references cited by and citations to a specific publication or set of publications.

Hand Searching Peer-Reviewed Journals

Bibliographic databases generally set requirements for inclusion (i.e., indexing) of articles published by peer-reviewed journals in their database. This results in many journals, and therefore their associated articles, being partially indexed (e.g., volume 3 to present), selectively

indexed (e.g., only systematic reviews), or not indexed at all. This is commonly the case for newer journals and topics that have historically been niche (e.g., transgender health, cannabis as a health care intervention). Therefore, in some instances, Center researchers may choose to hand search a small selection of journals. The decision to hand search journals is determined on a case-by-case basis but generally includes no more than 5 journals.

Search Strategy Development

Bibliographic Databases

A Center information specialist identifies key elements from the PICO framework to develop a comprehensive structured search strategy that uses keywords and controlled vocabulary terms for Ovid MEDLINE (see Box A for further detail).[3,8,9] The results of the initial MEDLINE search strategy are evaluated and terms are modified in an iterative process to achieve a precise, sensitive search. See Appendix A for an example of a structured search strategy constructed for a health technology assessment. All MEDLINE search strategies are reviewed by a second information specialist using criteria from the Peer Review of Electronic Search Strategies (PRESS) guideline.[21] Once reviewed, the MEDLINE search strategy is translated for use in other selected bibliographic databases.

Box A. Building a Structured Search Strategy

Structured searches contain controlled vocabulary terms and keywords, combined with Boolean operators.

Controlled vocabulary terms
- Are a set of standardized terms used for indexing and cataloging records
- Provide a consistent way to find information on the same concept, regardless of the terminology used in the original source
- May be unique to a particular database, such as Medical Subject Headings (MeSH) in MEDLINE

Keywords are the natural language terms used by health care practitioners, policymakers, and the public to discuss the condition and intervention of interest.

Boolean operators are simple words (AND, OR, NOT) used as conjunctions to combine or exclude terms in a structured search:
- *OR* is used to combine terms for the same concept
- *AND* is used to combine terms for different concepts
- *NOT* is used infrequently as doing so may lead to inadvertently excluding relevant results

If necessary to capture the evidence required by the key questions, searches may be amended, modified, or abbreviated to
- Incorporate additional terms and phrases (e.g., new treatment, identification of a previously unknown program)
- Incorporate other elements of interest (e.g., to retrieve studies that evaluate the economics of the condition and health technology or service)
- Accommodate database-specific limitations (e.g., no controlled vocabulary available)

Outcomes are often not included in a search strategy for systematic reviews because outcomes are not frequently referred to in the title or abstract of a published study, nor are outcomes typically indexed by a database. Therefore, including terms for outcomes in a search strategy may lead to failure to retrieve potentially relevant studies.[22,23]

Application of Search Limits

Search limits (e.g., date, language) may be applied to search strategies to reduce the number of results returned and therefore aid Center researchers in efficiently identifying studies and publications for inclusion. As such, all searches are limited to studies conducted in humans and

published in the English language. Additional limits may be applied to the search strategy judiciously, but should be justifiable.[8,24] For example:

- A search filter may be used to restrict the results by study design when the aim is to compare the effectiveness of 2 technologies or interventions (e.g., randomized controlled trials).
- A date limit may be applied to studies published within the past 5 years because the health technology or service of interest was not available before that period.

Other Sources

Searches of clinical trial registries and websites of professional organizations, special interest societies, regulatory bodies, and manufacturers are constructed according to the search capabilities of each site. For example, ClinicalTrials.gov allows users to enter keywords in predefined search boxes (e.g., condition, intervention) and apply search filters (e.g., date, ages, study phase) to create a structured search. Websites of professional organizations, special interest societies, and manufacturers often only allow for simple keyword searches of the entire site. In these instances, appropriate keywords for the condition and health technology or service are entered separately or combined if permitted.

Reference Management

An EndNote library is used to manage all identified evidence, regardless of the source. EndNote is a reference management program that allows Center researchers to maintain a searchable database of references, retrieve and store full-text documents, and insert formatted citations and a list of references into documents.

Recording and Reporting of Search Methods

Center researchers use a standardized form to document the clinical evidence search. For each source searched, details of the search (e.g., bibliographic database, date searched, number of results) are documented according to the guidelines for reporting literature searches established in an extension to the PRISMA statement.[25] This information is provided in the Methods Appendix of the report and used to construct a study flow diagram for the Findings section of the report. Appendix A has an example of a PRISMA flow diagram.

Clinical Evidence Selection

Center researchers use DistillerSR, a cloud-based systematic review platform, to manage selection of studies and clinical practice guidelines. References are exported from EndNote into DistillerSR. Duplicate references are removed in DistillerSR.

Two Center researchers independently screen titles and abstracts and review full-text articles using the inclusion and exclusion criteria defined in the scope statement. Disagreements are resolved by discussion between the 2 researchers. If consensus is not reached by discussion, then a third Center researcher reconciles the disagreement. DistillerSR tracks the number of studies excluded at each stage and reasons for exclusion during full-text review. This information is used to construct a study flow diagram for the report, in accordance with the PRISMA statement.[5] The PRISMA flow diagram shows the results of the search and selection process, tracking the number of records identified in the search to the number of studies ultimately included in the report.

Center researchers do not routinely include conference abstracts, posters, and ongoing and unpublished studies provided as part of a clinical trial registry record in the evidence synthesis.

Policy Sources

Center researchers search for federal, state, and major private payer policies related to the topic of interest, according to a list that the Department staff provided to the Center. For federal policies, Center researchers search the resources available from the Centers for Medicare & Medicaid Services for information about local and national coverage determinations. Center researchers search state and private payer websites, provider manuals, and related state statute or administrative rule websites for payer policies and related regulations on the topic. A reference for each source, including date accessed, is added to an EndNote library. Center researchers use a standardized form to document details about the policy search.

State Programs

- California Medicaid
- Florida Medicaid
- Massachusetts Medicaid
- New Jersey Medicaid
- New York Medicaid
- North Carolina Medicaid
- Oregon Medicaid and the Health Evidence Review Commission (HERC) coverage guidance (including topics under consideration)
- Pennsylvania Medicaid
- Texas Medicaid
- Washington Medicaid and the Washington Health Technology Assessment Program coverage determinations (including topics under consideration)

Private Payers

- Aetna
- Anthem Blue Cross and Blue Shield
- Highmark Blue Shield of Northeastern New York
- Capital District Physicians' Health Plan
- Cigna
- EmblemHealth
- Excellus BlueCross BlueShield
- Tufts Health Plan
- UnitedHealthcare

5. Assessing Risk of Bias

Center researchers assess threats to the internal and external validity of the evidence (defined as the risk of bias) to help understand whether results described in the report are reliable. Internal validity refers to how well a study's design, execution, analysis, reporting, and conclusions support a causal relationship between the health technology and outcomes while eliminating alternative explanations for that relationship.[26] External validity refers to whether the results of

the study can be reasonably generalized to populations or settings beyond that of the study population and setting.[27]

Center researchers use a set of standard tools, based on international, validated instruments, to assess the risk of bias for every included study in the Findings section of the report.[5,28-43] Center researchers employ a similar approach to assessing the methodological quality of clinical practice guidelines. These standard forms are filled out and responses are tracked in the DistillerSR platform. Two researchers independently assess the risk of bias of each study or guideline. In situations where the 2 researchers select different levels of risk of bias, they discuss their primary reasons for their selected level and try to resolve the conflict through discussion. If they are not successful in agreeing on a single level of risk of bias, the assigned research director serves as a third reviewer to determine the risk of bias.

Examples of considerations used to assess bias for clinical studies include the following:

- How similar baseline characteristics are between groups or clusters within the study
- Whether participants, investigators, and outcome assessors are unaware of the participant's assignment to intervention or control conditions (i.e., successful blinding)
- If outcomes are measured with valid and reliable measures
- Whether funding sources or disclosures of interest for the investigators are likely to affect study validity (i.e., conflicts of interest)

Appendix B details each of the domains and elements Center researchers consider when assessing the risk of bias for randomized studies, nonrandomized studies, and economic modeling studies. The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for a study is assessed as *high*, *moderate*, or *low* based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.

Table 1 describes example characteristics of studies that may fall into each of the 3 categories of risk of bias and is organized by study design.

Table 1. Examples of Study Characteristics for the 3 Levels of Risk of Bias by Study Design

Low Risk of Bias	Moderate Risk of Bias	High Risk of Bias
Systematic reviews		
Low-risk-of-bias systematic reviews include a clearly focused question, a literature search sufficiently rigorous to identify all relevant studies, criteria used to assess study quality and select studies for inclusion (e.g., randomized controlled trials), and assessment of similarities between studies to determine whether combining them is appropriate for evidence synthesis.	Moderate-risk-of-bias systematic reviews have incomplete information about methods that might mask important limitations or a meaningful conflict of interest.	High-risk-of-bias systematic reviews have clear flaws that could introduce significant bias.

Low Risk of Bias	Moderate Risk of Bias	High Risk of Bias
Randomized controlled trials		
Low-risk-of-bias randomized controlled trials include a clear description of the population, setting, intervention, and comparison groups; a random and concealed allocation of patients to study groups; low dropout rates; intention-to-treat analyses; and low potential for bias from conflicts of interest and funding source(s).	Moderate-risk-of-bias randomized controlled trials have incomplete information about methods that might mask important limitations or a meaningful conflict of interest.	High-risk-of-bias randomized controlled trials have clear flaws that could introduce significant bias.
Quasi-experimental studies		
Low-risk-of-bias quasi-experimental studies have a control group that is unexposed to the intervention being studied; methods are in place to prevent contamination bias; pre- and post-measures are done concurrently; and participant characteristics are balanced between groups or controlled for by propensity scores, by statistical adjustment, or both.	Moderate-risk-of-bias quasi-experimental studies have incomplete information about methods that might mask important limitations, a meaningful conflict of interest, or are at risk for contamination bias.	High-risk-of-bias quasi-experimental studies do not have a control group (i.e., before and after studies or interrupted time series) or have other clear flaws that could introduce significant bias.
Cohort studies		
Low-risk-of-bias cohort studies include a sample that is representative of the source population, have low loss to follow-up, measure and consider relevant confounding factors, and list their funding source(s) and have a low potential of bias from conflicts of interest.	Moderate-risk-of-bias cohort studies might not have measured all relevant confounding factors or adjusted for them in statistical analyses, have loss to follow-up that could bias findings, consist of a sample that is not representative of the source population, or have potential conflicts of interest that are not addressed.	High-risk-of-bias cohort studies have a clear, high risk of bias that would affect findings.
Case-control studies		
Low-risk-of-bias case-control studies include appropriate and clear consideration and selection of cases and controls, valid measures of exposures in both groups, and statistical adjustment for all major confounding variables. These studies also list their funding source(s) and have a low potential of bias from conflicts of interest.	Moderate-risk-of-bias case-control studies might not have measured all relevant confounding factors or adjusted for them in statistical analyses, might include controls that are not fully representative of cases, or might have potential conflicts of interest that are not addressed.	High-risk-of-bias case-control studies have a clear, high risk of bias that would affect findings.

Low Risk of Bias	Moderate Risk of Bias	High Risk of Bias
Cross-sectional studies		
Not applicable	Not applicable	Cross-sectional studies are hypothesis-generating studies and lack the temporal nature of a design to assess causal relationships. This study design is vulnerable to a high risk of bias. As a result, all cross-sectional studies are rated as having high risk of bias.
Case studies and series		
Not applicable	Not applicable	Case study and case series designs are descriptive, uncontrolled, and nonanalytic study designs. The methods used in these types of studies render them as having a high risk of bias, therefore, these studies are rated as having high risk of bias.
Economic modeling studies (i.e., cost and cost-effectiveness)		
Low-risk-of-bias economic evaluations include a well-described research question with economic importance and detailed methods to estimate the effectiveness and costs of the intervention. These studies provided a sensitivity analysis for all important variables, and the researchers justified the choice and values of variables. These studies also have low potential for bias from conflicts of interest and funding source(s).	Moderate-risk-of-bias economic evaluations have incomplete information about methods to estimate the effectiveness and costs of the intervention. The studies' sensitivity analyses might not consider 1 or more important variables, and the researchers did not completely justify the choice and values of variables. All of these factors might mask important study limitations.	High-risk-of-bias economic evaluations have clear flaws that could introduce significant bias. These could include significant conflict of interest, lack of sensitivity analysis, or lack of justification for the choice of values and variables.

Center researchers assess the methodological quality of the guidelines using an instrument adapted from the Appraisal of Guidelines for Research and Evaluation (AGREE) Collaboration,[28,29] and 2 researchers assign the guideline a rating of *good*, *fair*, or *poor* methodological quality based on its adherence to recommended methods and potential for biases. A good-methodological-quality guideline fulfills all or most of the criteria outlined in the instrument. A fair-methodological-quality guideline fulfills some of the criteria, and its unfulfilled criteria are not likely to alter the recommendations. A poor-methodological-quality guideline meets few or none of the criteria. Appendix B provides more details about the domains within the assessment instrument for clinical practice guidelines.

6. Synthesizing Evidence

After assessing the risk of bias of the included studies reported in the Findings section of the report, Center researchers begin the process of identifying the most relevant information from those studies to include in text and table format. A list of studies that were excluded during the full-text stage of screening is included as an appendix in the report, along with the primary reason for exclusion (i.e., there may be multiple reasons why a study is not eligible for inclusion). EBBRAC reports contain narrative summaries, evidence tables, and when possible, meta-analysis to synthesize findings across included studies. Center researchers use the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) assessment of an evidence body to make judgments about the overall certainty of a body of evidence by outcome.

Center researchers do not include written recommendations to decision makers in EBBRAC reports; instead, the Executive Summary highlights the most relevant findings and considerations for EBBRAC members to support their decision about recommendations for coverage.

Data Abstraction

Center researchers develop a form in the DistillerSR platform for abstracting data using a standard, consistent method for all included studies. Examples of types of information that Center researchers gather using the form include study location, a brief study population description, inclusion criteria, exclusion criteria, baseline characteristics of study participants, and results reported in the publication that are relevant to the outcomes selected in the PICO.

One researcher fills out the primary abstraction forms, and a second researcher performs quality assessment checks on about 10% of the abstracted studies to ensure accuracy before any abstracted data is used to create data tables, meta-analyses, or text-based synthesis of findings. Typical tables include the following:

- Tables that describe the characteristics of included studies (e.g., study design, number of participants, intervention description) and the assessed risk of bias
- Tables that present statistical findings by outcome and individual study
- Summary of findings tables with the GRADE assessment by population and outcome

Narrative Synthesis

Every EBBRAC report includes narrative synthesis. Center researchers synthesize their findings throughout the body of the report, addressing evidence, relevant cost and cost-effectiveness studies, clinical practice guidelines, and policy findings. Center researchers use the key questions to organize the narrative synthesis and prioritize content directly related to the inclusion criteria from the scope statement. Patterns of similarities and differences across included publications (e.g., study results, payer policies, guidelines) are identified and discussed throughout the report.

Meta-Analysis

Although all reports include narrative synthesis, not all reports include a meta-analysis. After the data abstraction process is complete, Center researchers assess whether there are enough commonalities in study design (e.g., comparison group types), outcomes collected (e.g., same standard validated measures), timing of collection (e.g., 12 months after the intervention), and other considerations to pool information across studies to generate a combined estimate of

effect. Center researchers follow guidelines from well-respected international sources for how to conduct meta-analyses and how to build an appropriate model. Center researchers typically use the Cochrane RevMan platform to perform the meta-analysis, but may use Stata or R platforms for meta-analysis, as appropriate.

GRADE: Grading of Recommendations, Assessment, Development, and Evaluation

Center researchers use the GRADE approach[44,45] to make judgments about the overall certainty of a body of evidence by outcome. Using a standard process, Center researchers weigh the following elements to determine how confident they are that the effect across all included studies is close to the true effect of the intervention by outcome:

- Study design limitations (i.e., risk of bias as discussed in Chapter 5)
- Inconsistency of results across studies (e.g., unexplained differences in effect sizes, confidence intervals around point estimates that do not overlap across studies, large statistical measures of heterogeneity)
- Indirectness of evidence (e.g., differences between the study population and the population of interest, use of surrogate outcomes, and indirect comparisons between groups)
- Imprecision (e.g., wide confidence intervals around the estimate of effect; uncertainty of whether the reported effect is a meaningful clinically important difference)
- Publication bias (i.e., selective publication of studies may result in overestimation or underestimation of benefits or harms related to the health technology)
- Magnitude of effect
- Dose-response gradient (e.g., the presence of a dose-response gradient may increase confidence in findings from studies without controlled designs)
- Plausibility of potential confounders

Center researchers summarize essential information in a consistent format in the GRADE summary tables throughout the report. Appendix A has an example GRADE Summary of Findings table.

GRADE System for Rating the Quality of a Body of Evidence

After Center researchers synthesize the most relevant outcome information in narrative and table formats, Center researchers assess the entire body of evidence presented in those results, weighing the elements described in the previous section. The 5 categories of GRADE rating are described in Table 2.

Table 2. GRADE System for Rating the Certainty of Evidence for Outcomes

GRADE Rating	Plain Language Description	Detailed Category Description
High	New research is very unlikely to change our understanding of the relationship between this outcome and the health technology.	Center researchers are very confident that the estimate of the effect of the intervention on the outcome lies close to the true effect. Typical sets of studies are randomized controlled trials with few or no limitations, and the estimate of effect is likely stable.
Moderate	New research may change our understanding of the relationship between this	Center researchers are moderately confident in the estimate of the effect of the intervention on the outcome. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it

GRADE Rating	Plain Language Description	Detailed Category Description
	outcome and the health technology.	is different. Typical sets of studies are randomized controlled trials with some limitations or well-performed nonrandomized studies with additional strengths that guard against potential bias and have large estimates of effects.
Low	New research is likely to change our understanding of the relationship between this outcome and the health technology.	Center researchers have little confidence in the estimate of the effect of the intervention on the outcome. The true effect may be substantially different from the estimate of the effect. Typical sets of studies are randomized controlled trials with serious limitations or nonrandomized studies without special strengths.
Very low	New research is very likely to change our understanding of the relationship between this outcome and the health technology.	Center researchers have no confidence in the estimate of the effect of the intervention on the outcome. The true effect is likely to be substantially different from the estimate of effect. Typical sets of studies are nonrandomized studies with serious limitations or inconsistent results across studies.
Not applicable	There is no research to report.	Center researchers did not identify any eligible articles.

Source. Adapted from 2 publications about GRADE.[44,45]
Abbreviation: GRADE: Grading of Recommendations, Assessment, Development, and Evaluations.

7. Writing the Report

Center researchers organize the report so that the most relevant information for EBBRAC discussion is in the Executive Summary section, with more detailed information in the body of the main report, and even greater detail in the report appendices. This chapter lists each standard section for EBBRAC reports and provides a brief description of the contents of each section. In addition to writing the report, Center researchers create a presentation based on the report's content.

Front Matter

Title Page

Table of Contents

This section lists all main headings and page numbers.

List of Tables and Figures

This is an optional section that Center staff uses when there are many tables and figures in a report that may be useful to refer to during EBBRAC discussion.

Glossary

This optional section is useful for listing key terms and abbreviations with their definitions for topics that have new, emerging, or nuanced vocabulary, or that entail multiple standard abbreviations.

Executive Summary

The Executive Summary includes the following sections, from Background through Conclusions.

Background

The Background section provides context for the key questions addressed in the report, including recent history of the topic and relevant policy context, in about 1 to 2 paragraphs.

Key Questions

The key questions in the Executive Summary are an abbreviated version of the main questions; they may not include subquestions that are detailed in the main body of the report.

Methods

The methods section in the Executive Summary orients readers to the major strategies for identifying and synthesizing the findings in the report in 1 brief paragraph.

Summary of Findings GRADE Tables

These tables synthesize the most relevant clinical evidence findings by outcome and provide an assessment of certainty of evidence and of balance of benefits and harms; they do not synthesize findings about resource use, equity, acceptability, and feasibility of implementing an intervention. Appendix A has an example GRADE summary of findings.

Key Policy Findings

The key findings present findings of interest for policymakers for their decision-making processes from select payer policies, clinical practice guidelines, and other related sources. The information is organized in bullet points by theme, is balanced and neutral, and does not provide recommendations.

Conclusions

The final section of the Executive Summary outlines key points such as main findings, shortcomings of the research, or other important considerations for implementation in 1 to 2 sentences.

Background

The Background section

- Describes the intervention of interest
- Details the clinical need and population
- Addresses recent history of the topic, including political, legal, and regulatory context
- Defines important terms used throughout the report
- Summarizes identified information relevant to contextual questions included in the scope statement

Center researchers use endnote citations throughout the main report text to cite sources, beginning in the Background section.

Key Questions

The key questions and subquestions are listed in this section and are identical to the questions developed in the scope statement. A typical report includes key questions about effectiveness, safety, cost analysis, clinical practice guidelines, and payer policies related to the intervention in question. Subquestions often ask whether the topic of the key questions varies by patient characteristics, disease characteristics, setting or provider characteristics, or other considerations related to social determinants of health. Some topics may include contextual questions that are answered in the Background section of the report, such as questions about recent developments of the topic, alternative treatments, and potential barriers to access or implementation. Appendix A has an example of key questions.

PICO

This section lists the populations, interventions, comparators, and outcomes to provide detail for identifying findings to answer the key questions. A detailed inclusion and exclusion table developed for the scope statement may be included in this section.

Methods

This section summarizes the search strategies for identifying clinical evidence and policy documentation used to identify and screen publications to answer the key questions, including which sources were searched and important limits on the searches. A more detailed description of the methods will be in Appendix A of the report.

Findings

This section includes syntheses of the findings from identified relevant publications to answer the key questions within the specifications of the PICO; it is organized in the following sections.

Clinical Evidence Review

The clinical evidence findings are organized thematically and include narrative synthesis of results from publications of studies that meet the PICO criteria. Depending on availability of similar data from included studies, this section may provide a meta-analysis of results from multiple studies. Risk of bias and other important considerations are included when interpreting results from individual studies. This section typically includes a table presenting the characteristics of each included study, tables with relevant data from included studies, and a table with an overview of relevant ongoing trials organized into the appropriate following subsections.

Effectiveness

This section synthesizes information about effectiveness outcomes. Multiple studies may contribute information to each outcome summary, and results are presented narratively and in a GRADE table format.

Safety
This section synthesizes information about safety outcomes such as serious adverse events (i.e., harms). Like the effectiveness subsection, this section presents results by outcome in narrative and GRADE table formats.

Subpopulation Considerations for Effectiveness and Safety
This section discusses any results by subpopulation characteristics of interest (e.g., age groupings of participants), and it may discuss whether the included studies reported information regarding interactions between social determinants of health and the effectiveness and safety of the intervention being studied.

Cost and Cost-Effectiveness
For topics where published cost analyses are identified, this section summarizes relevant cost information from these studies and considers relevance of the included models to a Medicaid program in a US health care context.

Clinical Practice Recommendations
This section summarizes the relevant recommendations from identified clinical practice guidelines.

Relevant Ongoing Trials
This section provides a high-level overview of ongoing trials that may be relevant to the topic of the report, whether any of the trials may add information pivotal for informing future decision making, and when results from the trials may be expected.

Payer Policies
This section summarizes relevant aspects of identified payer policies from Medicaid programs, Medicare local and national coverage determinations, and select private payers.

Discussion
This section identifies patterns across the Findings sections and offers considerations for using the findings in decision making, and it highlights potential limitations in the clinical evidence, clinical practice guidelines, payer policies, and cost analysis studies synthesized in the report.

References
The References section lists information for cited sources in order of appearance in the report, indexed by the endnote number.

Appendices

Appendix A. Search Strategies
This appendix provides additional detailed information related to searching methods (e.g., databases searched, search strategies), and may include a PRISMA diagram if it is not included in the main body of the report.

Appendix B. Detailed Inclusion and Exclusion Criteria

This appendix describes inclusion and exclusion criteria used during screening to sift through the results of the search strategies listed in Appendix A. This information is presented in a table format; it is the table from the Scope Statement.

Appendix C. Additional Tables

Some reports require additional space for more detailed tables that organize evidence, clinical practice guidelines, policy, or other information that is synthesized in the findings section of the report.

Appendix D. Included Studies

This appendix lists all included studies with their citation information.

Appendix E. Excluded Studies With Primary Reason for Exclusion

This appendix lists studies excluded at full-text review, publication information, and the primary reason for exclusion.

Appendix F. Additional Methods

This appendix describes risk of bias assessment and any additional methods considerations relevant to individual reports.

8. Monitoring New Evidence and Updating Reports

Center researchers conduct surveillance searches on a rolling basis for each completed EBBRAC report. This process typically occurs a minimum of 1 year after report finalization, and there may be gaps longer than 1 year between surveillance periods for a given report. The Department staff may request that Center staff conducts surveillance on another timeline as needed. Surveillance searches include looking for new publications of studies identified as complete or ongoing in the report (e.g., using trial identifiers), review of selected payer policies for potential updates to coverage criteria, review of included clinical practice guidelines for updated recommendations or reaffirmation of previous guidance, and a search for new clinical practice guidelines.

Center researchers provide the Department staff with a written overview of information identified in the surveillance searches and discuss with the Department staff whether an update to the report may add useful information to support the previous findings, or may change the findings of the report.

9. Managing the Research Process

Center staff uses standard templates and forms to track information gathered throughout the research process and uses a consistent set of tools, methods, and platforms to conduct research and compile the report and presentation. Center staff uses a consistent naming convention and file folder structure to organize all drafts and relevant materials. In-progress materials are stored on Oregon Health & Science University's password-protected SharePoint, and those materials are moved to a permanent location on Oregon Health & Science University's secure internal drive after report finalization.

Center staff uses Asana project management software to track report progress and key due dates; DistillerSR to sift search results, assess risk of bias, and abstract data; EndNote to organize search results and citation information for reports; Microsoft Office products for report development; and Cochrane RevMan, Stata, or R for conducting meta-analyses.

Center staff maintains notes, records, and drafts of reports prepared for EBBRAC for the duration of the Center's contract with the Department. These records are updated during each surveillance period.

10. Using Evidence to Make Decisions

Respected health technology assessment groups from around the world recommend using a structured tool to build consensus for creating a coverage recommendation based on the identified clinical evidence, while also taking into account the other sources of information described in this chapter.[3,7,46-48]

Clinical evidence represents a single type of information that EBBRAC members may consider when making a coverage recommendation. Decisions for coverage may include discussion and consideration of the following elements[3]:

- Overall clinical benefit, including effectiveness, safety, burden of illness, and need
- Patient values and preferences, including effect on patients' and caregivers' lives and ethical principles such as patient privacy and autonomy
- How the health technology may fit into current care pathways, and what other options are already available for care
- Equity and patient care, including equity of access to care and outcomes
- Cost-effectiveness
- Feasibility of adoption into the current health system, including economic and organizational feasibility

Beyond the Center staff's presentation of clinical evidence, clinical practice guidelines, and payer policies, EBBRAC members may also weigh perspectives from providers and patient preferences shared through public comments. Patients and caregivers may have a unique understanding of how the condition and treatment affect the quality of life of individuals and their families, may represent communities not represented in the current published clinical evidence literature, and may give an alternative viewpoint for how a health system is set up to manage the treatments from a patient perspective.[3]

Center staff plans to work with the Department staff to pilot a decision tool based on the GRADE Evidence to Decision framework for coverage decisions[49] during the first 12 months of the contract, with consideration for the unique context of New York state and populations served by the Medicaid program in New York. Box B below lists potential questions from this framework to facilitate discussion about recommendations for coverage.

Box B. Discussion Questions for Coverage Recommendations

Criteria of an evidence to decision framework for coverage may include the following considerations:

- How substantial are the desirable anticipated effects?
- How substantial are the undesirable anticipated effects?
- What is the overall certainty of the evidence of effects?
- Is there important uncertainty about how much people value the main outcomes?
- Does the balance between desirable effects and undesirable effects favor the option or the comparison?
- How large are the resource requirements (costs)?
- What is the certainty of the evidence of resource use?
- Does the cost-effectiveness of the option favor the option or the comparison?
- What would be the effect on health equity?
- Is the option acceptable to key stakeholders?
- Is the option feasible to implement?

Source. Adapted from the GRADE Evidence to Decision framework.[49,50]

11. Updating the Manual

Center staff and the Department staff plan to review the contents of this manual annually in January of each contract year to assess whether any updates to content are necessary. A chapter on applying these methods to assess the evidence on interventions to address social determinants of health is proposed for a future version of this manual.

Table 3. Change Log

Date	Summary of Change	Rationale

References

1. The New York State Senate. Medicaid evidence based benefit review advisory committee: section 365-D. 2015; https://www.nysenate.gov/legislation/laws/SOS/365-D. Accessed February 15, 2024.

2. New York State Department of Health. Evidence Based Benefit Review Advisory Committee (EBBRAC). 2024; https://www.health.ny.gov/health_care/medicaid/ebbrac/index.htm. Accessed February 29, 2024.

3. Ontario Health. Health technology assessments: methods and process guide. 2023; https://www.hqontario.ca/Portals/0/documents/evidence/reports/hta-methods-and-process-guide-en.pdf. Accessed October 26, 2023.

4. Higgins JPT, Thomas J, Chandler J, et al. Cochrane handbook for systematic reviews of interventions version 6.4. *Cochrane*. 2023; https://training.cochrane.org/handbook. Accessed November 2, 2023.

5. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi: 10.1136/bmj.n71.

6. Aromataris E, Munn Z. JBI manual for evidence synthesis. *JBI*. 2020; https://jbi-global-wiki.refined.site/space/MANUAL. Accessed November 2, 2023.

7. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. 2022; https://www.nice.org.uk/process/pmg36. Accessed November 2, 2023.

8. Higgins JPT, Lasserson T, Thomas J, Flemyng E, Churchill R. Methodological expectations of Cochrane intervention reviews (MECIR). *Cochrane*. 2023; https://community.cochrane.org/mecir-manual. Accessed November 3, 2023.

9. Kugley S, Wade A, Thomas J, et al. Searching for studies: a guide to information retrieval for Campbell Systematic Reviews. *The Campbell Collaboration*. 2016; http://www.campbellcollaboration.org/images/Campbell_Methods_Guides_Information_Retrieval.pdf. Accessed November 2, 2023.

10. Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med*. 2017;10(3):233-240. doi: 10.1111/jebm.12266.

11. Sadek J, Inskip A, Woltmann J, et al. ScanMedicine: an online search system for medical innovation. *Contemp Clin Trials*. 2023;125:107042. doi: 10.1016/j.cct.2022.107042.

12. Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. *BMJ*. 2012;344:d8141. doi: 10.1136/bmj.d8141.

13. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Clinical practice guidelines we can trust. *National Academies Press*. 2011; https://www.ncbi.nlm.nih.gov/books/NBK209546/. Accessed November 6, 2023.

14. De Leo A, Bloxsome D, Bayes S. Approaches to clinical guideline development in healthcare: a scoping review and document analysis. *BMC Health Serv Res*. 2023;23(1):37. doi: 10.1186/s12913-022-08975-3.

15. Lunny C, Ramasubbu C, Puil L, et al. Over half of clinical practice guidelines use non-systematic methods to inform recommendations: a methods study. *PLoS One*. 2021;16(4):e0250356. doi: 10.1371/journal.pone.0250356.

16. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. *NICE*. 2014; https://www.nice.org.uk/process/pmg20/chapter/introduction. Accessed February 14, 2024.

17. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med*. 2012;156(7):525-531. doi: 10.7326/0003-4819-156-7-201204030-00009.

18. Scottish Intercollegiate Guidelines Network. SIGN 50: a guideline developer's handbook. 2019; https://www.sign.ac.uk/our-guidelines/sign-50-a-guideline-developers-handbook/. Accessed February 15, 2024.

19. Guidelines International Network. 2024; https://g-i-n.net/. Accessed February 14, 2024.

20. Haddaway NR, Grainger MJ, Gray CT. Citationchaser: a tool for transparent and efficient forward and backward citation chasing in systematic searching. *Res Synth Methods*. 2022;13(4):533-545. doi: 10.1002/jrsm.1563.

21. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75:40-46. doi: 10.1016/j.jclinepi.2016.01.021.

22. Centre for Reviews and Dissemination. CRD's guidance for undertaking reviews in health care. *University of York*. 2009; https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Accessed November 2, 2023.

23. Frandsen TF, Nielsen MFB, Eriksen MB. Avoiding searching for outcomes called for additional search strategies: a study of Cochrane review searches. *J Clin Epidemiol.* 2022;149:83-88. doi: 10.1016/j.jclinepi.2022.05.015.

24. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. *BMC Med Res Methodol.* 2018;18(1):85. doi: 10.1186/s12874-018-0545-3.

25. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev.* 2021;10(1):39. doi: 10.1186/s13643-020-01542-z.

26. Cipriani A, Purgato M, Barbui C. Why internal and external validity of experimental studies are relevant for clinical practice? *Epidemiol Psichiatr Soc.* 2011;18(2):101-103. doi: 10.1017/s1121189x00000968.

27. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* 2005;365(9453):82-93. doi: 10.1016/S0140-6736(04)17670-8.

28. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ.* 2010;182(18):E839-E842. doi: 10.1503/cmaj.090449.

29. Brouwers MC, Kho ME, Browman GP, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ.* 2010;182(10):E472-E478. doi: 10.1503/cmaj.091716.

30. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 2017; https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf. Accessed August 28, 2018.

31. Cochrane Collaboration. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). 2021; https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2. Accessed January 13, 2022.

32. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *Int J Technol Assess Health Care.* 2005;21(2):240-245.

33. Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *BMJ.* 2013;346. doi: 10.1136/bmj.f1049.

34. National Heart, Lung, and Blood Institute. Quality assessment of case-control studies. 2014; https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed April 27, 2017.

35. National Heart, Lung, and Blood Institute. Quality assessment tool for before-after (pre-post) studies with no control group. 2018; https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed April 27, 2017.

36. National Heart, Lung, and Blood Institute. Quality assessment tool for observational cohort and cross-sectional studies. 2018; https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed May 30, 2018.

37. Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. *Arch Pathol Lab Med.* 2013;137(4):558-565. doi: 10.5858/arpa.2012-0198-RA.

38. Scottish Intercollegiate Guidelines Network. Methodology checklist 3: cohort studies. 2014; https://www.sign.ac.uk/what-we-do/methodology/checklists/. Accessed January 13, 2022.

39. Scottish Intercollegiate Guidelines Network. Methodology checklist 4: case-control studies. 2014; https://www.sign.ac.uk/what-we-do/methodology/checklists/. Accessed January 13, 2022.

40. Scottish Intercollegiate Guidelines Network. Methodology checklist 5: diagnostic studies. 2015; https://www.sign.ac.uk/what-we-do/methodology/checklists/. Accessed October 30, 2020.

41. Scottish Intercollegiate Guidelines Network. Methodology checklist 2: randomised controlled trials. 2015; https://www.sign.ac.uk/what-we-do/methodology/checklists/. Accessed October 30, 2020.

42. Scottish Intercollegiate Guidelines Network. Methodology checklist 1: systematic reviews and meta-analyses. 2015; https://www.sign.ac.uk/what-we-do/methodology/checklists/. Accessed October 30, 2020.

43. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097.

44. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336(7650):924-926. doi: 10.1136/bmj.39489.470347.AD.

45. Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE

Working Group, 2013. 2014; http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. Accessed December 15, 2015.

46. Agency for Healthcare Research and Quality (AHRQ). Methods guide for effectiveness and comparative effectiveness reviews. 2018; https://effectivehealthcare.ahrq.gov/products/collections/cer-methods-guide. Accessed November 3, 2023.

47. Institute for Quality and Efficiency in Health Care (IQWiG). General methods version 4.2. 2015; https://www.ncbi.nlm.nih.gov/books/NBK332872/. Accessed November 2, 2023.

48. Lakdawalla D, Neumann P, Wilensky G, et al. Health technology assessment in the US: a vision for the future. *USC Schaeffer Center for Health Policy & Economics and the Aspen Institute* 2021; https://healthpolicy.usc.edu/wp-content/uploads/2021/02/Health_Technology_Assessment_in_the_U.S..pdf. Accessed November 15, 2023.

49. Parmelli E, Amato L, Oxman AD, et al. GRADE evidence to decision (EtD) framework for coverage decisions. *Int J Technol Assess Health Care.* 2017;33(2):176-182. doi: 10.1017/S0266462317000447.

50. Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ.* 2016;353:i2016. doi: 10.1136/bmj.i2016.

51. Shaw B, Robalino S, King V. *Use of stereotactic body radiation therapy.* Portland, OR: Center for Evidence-based Policy; 2023.

Appendix A. Example Report Elements From a Recent HTA Report

This appendix presents examples of report elements from a recent report that Center researchers prepared for the Health Technology Clinical Committee under the Health Technology Assessment program, which is part of the Washington Health Care Authority.[51]

Example Key Questions

KQ1. What is the evidence of effectiveness for stereotactic body radiation therapy (SBRT) for patients with central nervous system cancers and inoperable stage 1 non-small cell lung cancer?

KQ2. What are the harms of SBRT in patients with included cancers?

KQ3. What is the evidence that SBRT has differential efficacy or harms in subpopulations, including:

a. Sex
b. Age
c. Site and type of cancer
d. Stage and grade of cancer
e. Setting, provider characteristics, equipment, quality assurance standards and procedures

KQ4. What is the evidence of cost and cost-effectiveness of SBRT?

Example Detailed PICO Inclusion and Exclusion Criteria

Study Component	Inclusion	Exclusion
Populations	• Adults and children with non-CNS and NSCLC (inoperable, stage 1) malignancies where treatment by radiation therapy is appropriate	• Studies in people with noncancer conditions (e.g., trigeminal neuralgia)
Interventions	• SBRT, with devices such as Gamma Knife, CyberKnife, TomoTherapy, delivered in 10 or fewer fractions	• Treatments delivered in 11 or more fractions • Interventions used for treatment planning or treatment delivery assessment only
Comparators	• Conventional (conformal) EBRT • Other forms of radiation (e.g., brachytherapy) • Chemotherapy • Surgery • No treatment	• Comparators other than those stated
Outcomes	• Effectiveness ○ Survival rate ○ Duration of symptom-free remission ○ Quality of life • Harms, including radiation exposure and complications • Cost • Cost-effectiveness	• Studies that do not report outcomes of interest • Data for treatment planning (e.g., dosing) or treatment delivery (e.g., accuracy) • Economic outcomes from studies performed in non-US countries • Economic outcomes from studies performed in the US and published more than 5 years ago
Timing	• Any point in the treatment pathway	• None stated

Study Component	Inclusion	Exclusion
Setting	• Any outpatient or inpatient clinical setting in countries categorized as very high on the UN Human Development Index	• Emergency use settings • Nonclinical settings (e.g., studies in healthy volunteers, animal models of disease) • Countries categorized other than very high on the UN Human Development Index
Study Design	• For KQ1, KQ2, and KQ3 ○ Comparative study designs (prospective, retrospective, and randomized or controlled clinical trials) • For KQ2 ○ Comparative study designs ○ Noncomparative study designs (≥ 100 participants) • For KQ4 ○ Comparative cost data and relevant economic evaluations ○ Cost-effectiveness analyses ○ Economic simulation modeling studies	• Abstracts, conference proceedings, posters, editorials, letters • Studies without a comparator (unless for harms only) • Proof-of-principle studies (e.g., technology development or technique modification) • Studies without extractable data
Sample Size	• Minimum sample size of 50 participants for comparative study designs • Minimum sample size of 100 participants for noncomparative study designs	• Studies that do not meet the minimum sample size
Publication	• Published, peer-reviewed, English-language articles	• Studies reported only as abstracts that do not allow study characteristics to be determined • Studies that cannot be located • Duplicate publications of the same study that do not report different outcomes or follow-up times, or single site reports from published multicenter studies • Studies published in languages other than English

Abbreviations. CNS: central nervous system; EBRT: external beam radiation therapy; KQ: key question; NSCLC: non-small cell lung cancer; SBRT: stereotactic body radiation therapy; UN: United Nations.

Example Search Strategy

Ovid MEDLINE(R) ALL <1946 to October 21, 2022>

Search date: October 24, 2022

1. (SBRT or SABR).ti,ab,kw.

2. (("stereotactic body" or stereotactic-body) adj1 (radiotherap* or "radio therap*" or RT or radiation or irradi* or ablati* or radioablati* or "radio ablat*")).ti,ab,kw.

3. ((stereotactic ablati* or stereotactic-ablati*) adj1 (radiotherap* or "radio therap*" or RT or radiation or irradi*)).ti,ab,kw.

4. (stereotactic radioablati* or stereotactic-radioablati*).ti,ab,kw.

5. or/1-4

6. (cyberknife* or cyber knife* or gammaknife* or gamma knife*).ti,ab,kw.

7. 6 and 5

8. ((cyberknife* or cyber knife* or gammaknife* or gamma knife*) and (SBRT or SABR)).ti,ab,kw.

9. ((cyberknife* or cyber knife* or gammaknife* or gamma knife*) adj2 (radiotherap* or "radio therap*" or RT or radiation or irradi* or ablati* or radioablati* or "radio ablat*")).ti,ab,kw.

10. or/7-9

11. 5 or 10

12. limit 11 to english language

13. (case reports or clinical conference or comment or congress or consensus development conference or consensus development conference, nih or editorial or interactive tutorial or letter or observational study, veterinary or randomized controlled trial, veterinary).pt.

14. ((phase 1* or phase i or phase ii or phase 2*) not (phase iii* or phase iv)).ti.

15. (exp Animals/ not Humans/) or (animal$1 or bovine$1 or canine$1 or cat$1 or chimpanzee$1 or cow$1 or dog$1 or feline$1 or goat$1 or hens or mice or monkey$1 or mouse or murine$1 or ovine or pig$1 or porcine or primate$1 or sheep or rabbit$1 or rat or rats or rattus or rhesus or rodent*).ti.

16. ((spine or spinal or brain or CNS or central nervous system or ventricular) not (non-spine or non-brain or non-CNS)).ti.

17. or/13-16

18. 12 not 17

19. (random* adj3 assign*).ab.

20. ("clinical trial" or clinical trial, phase iii or clinical trial, phase iv or controlled clinical trial or "multicenter study" or "randomized controlled trial").pt. or double-blind method/ or clinical trials as topic/ or clinical trials, phase iii as topic/ or clinical trials, phase iv as topic/ or controlled clinical trials as topic/ or randomized controlled trials as topic/ or early termination of clinical trials as topic/ or multicenter studies as topic/ or ((randomi?ed adj7 trial*) or (controlled adj3 trial*) or (clinical adj2 trial*) or ((single* or doubl* or tripl* or treb* or quad*) adj1 (blind* or

mask*))).ti,ab,kw. or ("2 arm" or "two arm" or "3 arm" or "three arm" or "4 arm" or "four arm" or "5 arm" or "five arm").ti,ab,kw. or quasi*.ti,ab.

21. (phase 3* or phase iii* or phase 4* or phase iv*).ti,ab.

22. (placebo* or head-to-head or (compar* adj3 (effectiveness or efficacy))).ti,ab,kw. or Comparative Effectiveness Research/

23. (active adj1 (comparator* or control$1 or treatment*)).ti,ab.

24. or/19-23

25. 18 and 24

26. cohort studies/ or longitudinal studies/ or follow-up studies/ or prospective studies/ or retrospective studies/ or cohort.ti,ab. or longitudinal.ti,ab. or prospective.ti,ab. or retrospective.ti,ab.

27. (18 and 26) not 25

28. (((comprehensive* or integrative or systematic*) adj3 (bibliographic* or review* or literature)) or (meta-analy* or metaanaly* or "research synthesis" or ((information or data) adj3 synthesis) or (data adj2 extract*))).ti,ab. or (cinahl or (cochrane adj3 trial*) or embase or medline or psyclit or (psycinfo not "psycinfo database") or pubmed or scopus or "sociological abstracts" or "web of science").ab. or ("cochrane database of systematic reviews" or evidence report technology assessment or evidence report technology assessment summary).jn. or Evidence Report: Technology Assessment*.jn. or ((review adj5 (rationale or evidence)).ti,ab. and review.pt.) or meta-analysis as topic/ or Meta-Analysis.pt.

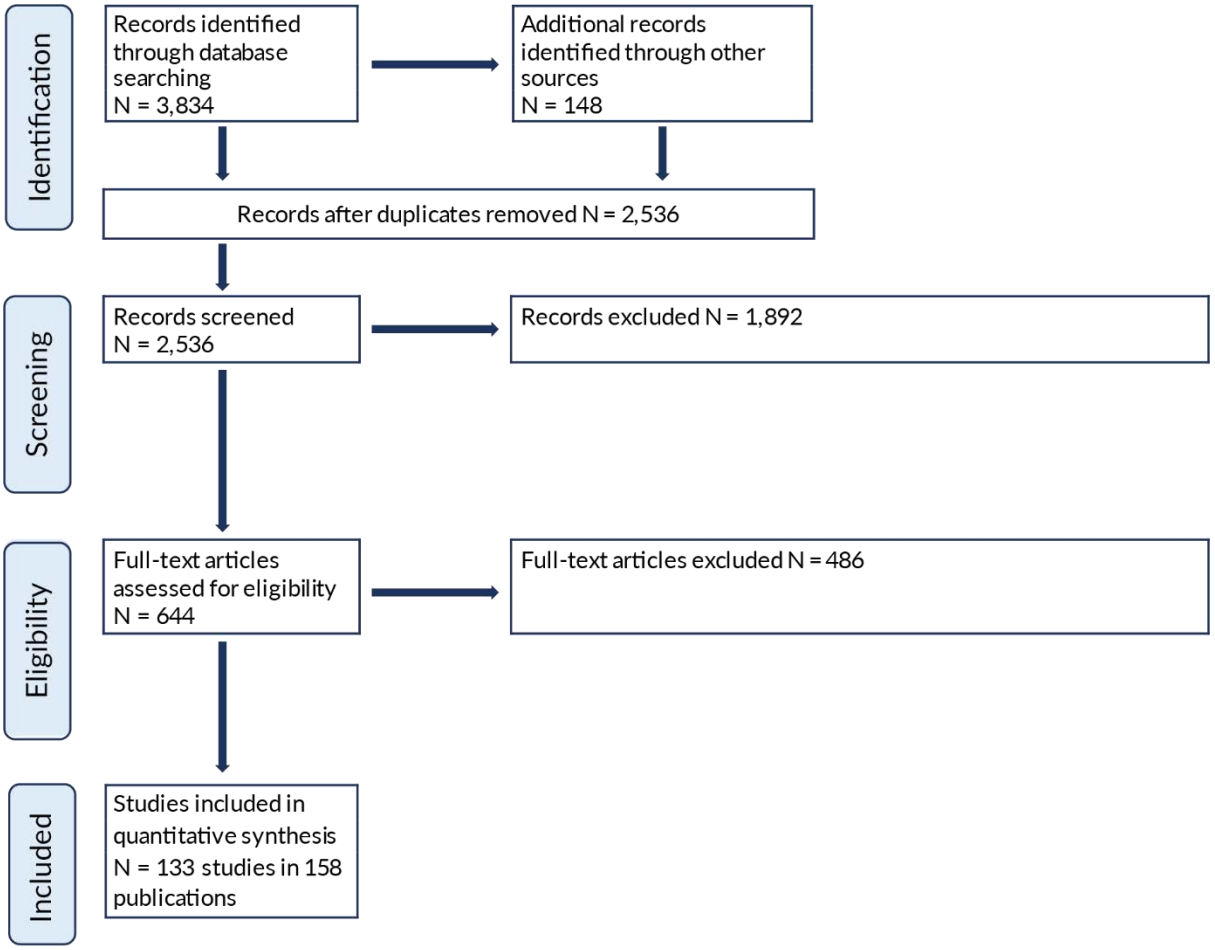
29. psychinfo.ab. or heath technology assessment.ti,ab. or ((review or umbrella or evidence) adj2 (review* or synthesis)).ti,ab.

30. or/28-29

31. (18 and 30) not (25 or 27)

32. 18 not (25 or 27 or 31)

Example Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Diagram



Example Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) Summary of Findings Table

Number of Participants (N) Number of Studies	Findings	Certainty of Evidence	Rationale
SBRT vs. surgery or no SBRT for operable early-stage NSCLC			
Overall survival			
N = 41,583 3 comparative NRSs	SBRT was associated with significantly worse outcomes than surgery for operable early-stage NSCLC; surgery was associated with around a 60% to 65% lower risk of mortality. However, 1 study did find that in patients who were medically operable, SBRT and lobectomy may be equally effective.	⨁◯◯◯ VERY LOW	Downgraded 1 level for inconsistency

Number of Participants (N) Number of Studies	Findings	Certainty of Evidence	Rationale
Progression-free survival			
N = 187 1 comparative NRS	In patients who were medically operable, SBRT and lobectomy may be equally effective (HR, 1.57; 95% CI, 0.68 to 3.64)	⨁◯◯◯ VERY LOW	Downgraded 1 level for risk of bias and 2 levels for imprecision (i.e., very wide CIs)[a]
Disease-control			
N = 60 1 RCT	In people with potentially resectable early-stage NSCLC, SBRT in combination with durvalumab was associated with significantly higher odds of having a major pathological response (OR, 16.0; 95% CI, 3.2 to 79.6) or a partial radiographic response (46.7% SBRT with durvalumab vs. 3.3% durvalumab; *P* = .001) than durvalumab alone.	⨁⨁⨁◯ MODERATE	Downgraded 1 level for risk of bias
Quality of life			
Not reported			
SBRT vs. RT for inoperable stage II NSCLC			
Overall survival			
N = 4,401 1 comparative NRS	SBRT appears to be associated with improved survival than cRT (HR, 0.79; 95% CI, 0.71 to 0.87) or hypofractionated radiotherapy (HR, 0.57; 95% CI, 0.50 to 0.66) for inoperable stage II NSCLC.	⨁⨁◯◯ LOW	Not downgraded
Progression-free survival			
Not reported			
Disease-control			
Not reported			
Quality of life			
Not reported			
SBRT vs. no SBRT for advanced NSCLC			
Overall survival			
N = 78 1 RCT	People with advanced NSCLC treated with SBRT after pembrolizumab or pembrolizumab alone had a similar overall survival (median: 15.9 months SBRT vs. 7.6 months control; HR, 0.66; 95% CI, 0.37 to 1.18) However, in subgroup analyses, men (HR, 0.42; 95%CI, 0.19 to 0.96; *P* = .04) and smokers (HR, 0.48; 95% CI, 0.25 to 0.93; *P* = .03) had significantly improved survival with SBRT compared with pembrolizumab alone.	⨁⨁⨁◯ MODERATE	Downgraded 1 level for imprecision (i.e., wide CIs)[a]
Progression-free survival			

Number of Participants (N) Number of Studies	Findings	Certainty of Evidence	Rationale
N = 78 1 RCT	People with advanced NSCLC treated with SBRT after pembrolizumab or pembrolizumab alone had a similar PFS (HR, 0.71; 95% CI, 0.42 to 1.18).	⨁⨁◯◯ LOW	Downgraded 1 level each for risk of bias and imprecision (i.e., wide CIs)[a]
Disease-control			
Not reported			
Quality of life			
Not reported			
SBRT vs. surgery or cRT for lung metastases			
Overall survival			
N = 483 4 comparative NRSs	In people with lung metastases, SBRT and surgery may be associated with similar overall survival (median survival at 2 years of around 68% to 77% in the SBRT group vs. 82% in the surgery group); however, SBRT may be associated with improved survival when compared with cRT (median survival of 26 months in the SBRT group vs. 9 months in the cRT group; $P < .001$).	⨁⨁◯◯ LOW	Not downgraded
Progression-free survival			
N = 301 3 comparative NRSs	People with lung metastases treated with SBRT had significantly worse PFS than people treated with surgery (around 3 times more likely to have progression). However, results were mixed with 1 study showing no difference between SBRT and surgery.	⨁◯◯◯ VERY LOW	Downgraded 1 level for inconsistency
Disease-control			
N = 694 4 comparative NRSs	Results were mixed with SBRT being associated with both similar and lower levels of local control than surgery for lung metastases. SBRT, however, was significantly associated with improved local control when compared with cRT. Studies reported at different times using different statistics, precluding any summary statistics (see detailed findings below).	⨁◯◯◯ VERY LOW	Downgraded 1 level for inconsistency
Quality of life			
Not reported			
SBRT vs. surgery or cRT for LCNEC of the lung			
Overall survival			
N = 3,963 2 comparative NRSs	In people with LCNEC of the lung, SBRT may be associated with improved survival when compared with cRT (HR, 0.83; 95% CI, 0.68 to 1.00)[b], but worse outcomes when compared with surgery (HR, 1.61; 95% CI, 1.36 to 1.92).	⨁⨁◯◯ LOW	Not downgraded

Number of Participants (N) Number of Studies	Findings	Certainty of Evidence	Rationale
Progression-free survival			
Not reported			
Disease-control			
Not reported			
Quality of life			
Not reported			
SBRT vs. surgery and other RT for any lung cancer			
Toxicity			
N = 138 2 RCTs	Grade 3 and higher events occurred in around 3% to 11% of SBRT group; most common were dyspnea and pneumonia, pancreatitis, and fatigue.	⨁⨁⨁◯ MODERATE	Downgraded 1 level for risk of bias
N = 221 2 comparative NRSs	Grade 3 toxicities were not common with SBRT, and included lung toxicity (including radiation pneumonitis) and chest wall pain; ranging from 3% to 14% depending on the specific toxicity.	⨁⨁◯◯ LOW	Not downgraded

Notes. [a] *Inconsistency not assessable due to only 1 study;* [b] *Inverted for consistency.*
Abbreviations. CI: confidence interval; cRT: conventional radiation therapy; HR: hazard ratio; LCNEC: large-cell neuroendocrine carcinoma; NRS: nonrandomized study; NSCLC: non-small cell lung cancer; OR: odds ratio; PFS: progression-free survival; RCT: randomized controlled trial; SBRT: stereotactic body radiation therapy;

Appendix B. Detailed Risk of Bias Considerations

Table B1. Risk-of-Bias Assessment: Randomized Controlled Trials

Domain	Domain Elements The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for a study is assessed as *high*, *moderate*, or *low* based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.
Randomization	• An appropriate method of randomization is used to allocate participants or clusters to groups, such as a computer random number generator • Baseline characteristics between groups or clusters are similar
Allocation concealment	• An adequate concealment method is used to prevent investigators and participants from influencing enrollment or intervention allocation
Intervention	• Intervention and comparator intervention applied equally to groups • Co-interventions appropriate and applied equally to groups • Control selected is an appropriate intervention
Outcomes	• Outcomes are measured using valid and reliable measures • Investigators use single outcome measures and do not rely on composite outcomes, or outcome of interest can be calculated from composite outcome • The trial has an appropriate length of follow-up and groups are assessed at same time points • Outcome reporting of entire group or subgroups is not selective
Masking (blinding) of investigators and participants	• Investigators and participants are unaware (masked or blinded) of intervention status
Masking (blinding) of outcome assessors	• Outcome assessors are unaware (masked or blinded) of intervention status
Intention-to-treat analysis	• Participants are analyzed based on random assignment (intention-to-treat analysis)
Statistical analysis	• Participants lost to follow-up unlikely to significantly bias results (i.e., complete follow-up of ≥ 80% of participants overall and nondifferential, ≤ 10% difference between groups) • The most appropriate summary estimate (e.g., risk ratio, hazard ratio) is used • Paired or conditional analysis used for crossover RCT • Clustering appropriately accounted for in a cluster-randomized trial (e.g., use of an intraclass correlation coefficient)
Other biases (as appropriate)	• List others in table footnote and describe, such as: o Sample size adequacy o Interim analysis or early stopping o Recruitment bias, including run-in period used inappropriately o Use of unsuitable crossover intervention in a crossover RCT
Interest disclosure	• Disclosures of interest are provided for authors/funders/commissioners of study • Interests are unlikely to significantly affect study validity
Funding	• There is a description of source(s) of funding • Funding source is unlikely to have a significant impact on study validity

Abbreviation. RCT: randomized controlled trial.

Table B2. Risk of Bias Assessment: Nonrandomized Studies

Domain	Domain Elements The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for study is assessed as *high*, *moderate*, or *low*, based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.
Participant selection	For cohort studies: • The 2 groups being studied are selected from source populations comparable in all respects other than factor under investigation, or statistical adjustment is used appropriately to achieve this • The study indicates how many of people asked to take part did so in each of the groups being studied • The likelihood some eligible participants might have outcome at time of enrollment is assessed and considered in analysis • Fewer than 20% of individuals or clusters in each arm of study dropped out before study was completed For case-control studies: • Cases and controls are clearly specified and defined, with inclusion and exclusion criteria applied appropriately • Cases may be selected by meeting inclusion criteria, controls may be selected by meeting inclusion criteria and then being matched to cases • Sampling selection (ratio of cases to control) is justified • Cases and controls selected from same population and same timeframe; when not all cases and controls are selected from same population, these are randomly selected • Among cases, investigators confirm that exposure occurred before development of disease being studied and/or likelihood that some eligible participants might have outcome at time of enrollment is assessed and considered in analysis
Intervention	• The assessment of exposure to intervention is reliable • Exposure level or prognostic factors are assessed at multiple times across length of study, if appropriate • For case-control studies, assessors of (intervention) exposure status are unaware (masked or blinded) to case or control status of participants, and there is a method to limit effects of recall bias on assessment of exposure to intervention
Control	• Control condition represents an appropriate comparator
Outcome	• There is a precise definition of outcomes used • Outcomes are measured using valid and reliable measures, evidence from other sources is used to demonstrate method of outcome assessment is valid and reliable • Investigators use single outcome measures and do not rely on composite outcomes, or outcome of interest can be calculated from composite outcome • The study has an appropriate length of follow-up for outcome reported and groups are assessed at same time points • Outcome reporting of entire group or subgroups is not selective • When patient-reported outcomes are used, there is a method for validating measure
Masked outcome assessment	• The assessment of outcome(s) is made blind to exposure status. Where outcome assessment blinding was not possible, there is recognition that knowledge of exposure status could have influenced assessment of outcome.

Domain	Domain Elements The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for study is assessed as *high*, *moderate*, or *low*, based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.
	• For case-control study: assessors of exposure status are unaware (masked or blinded) of case or control status of participant
Confounding	• The main potential confounders are identified and considered in design and analysis of study
Statistical analysis	• Comparison is made between full participants and those who dropped out or were lost to follow-up, by exposure status • If groups were not followed for an equal length of time, analysis was adjusted for differences in length of follow-up • All major confounders are adjusted for using multiple variable logistic regression or other appropriate statistical methods • Confidence intervals (or information used to calculate them) are provided • For case-control studies that use matching, conditional analysis is conducted or matching factors are adjusted for in analysis
Other biases (as appropriate)	• List others in table footnote and describe • Sample size adequacy
Interest disclosure	• Disclosures of interest are provided for authors/funders/commissioners of study • Interests are unlikely to significantly affect study validity
Funding source	• There is a description of source(s) of funding • Funding source is unlikely to have a significant impact on study validity

Table B3. Risk of Bias Assessment: Economic Modeling Studies

Domain	Domain Elements The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for study is assessed as *high*, *moderate*, or *low* based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.
Target population	• Target population and care setting described • Describe and justify basis for any target population stratification, identify any previously identifiable subgroups • If no subgroup analyses were performed, justify why these were not required
Perspective	• State and justify analytic perspective (e.g., societal, payer, etc.)
Time horizon	• Describe and justify time horizon(s) used in analysis
Discount rate	• State and justify discount rate used for costs and outcomes
Comparators	• Describe and justify selected comparators • Competing alternatives appropriate and clearly described
Modelling	• Model structure (e.g., scope, assumptions made) is described and justified • Model diagram provided, if appropriate • Model validation is described (may involve validation of different aspects such as structure, data, assumptions, and coding and different validation models such as comparison with other models) • Data sources listed and assumptions for use justified • Statistical analyses are described
Effectiveness	• Estimates of efficacy/effectiveness of interventions are described and justified

Domain	Domain Elements The elements included in each domain are assessed and rated as *yes*, *no*, *unclear*, or *not applicable* based on performance and documentation of individual elements in each domain. The overall risk of bias for study is assessed as *high*, *moderate*, or *low* based on assessment of how well overall study methods and processes were performed to limit bias and ensure validity.
	• The factors likely to have an impact on effectiveness (e.g., adherence, diagnostic accuracy, values, and preferences) are described and an explanation of how these were factored into analysis is included • The quality of evidence for relationship between intervention and outcomes, and any necessary links, is described
Outcomes	• All relevant outcomes are identified, measured, and valued appropriately (including harms/adverse events) for each intervention, and justification for information/assumptions is given • Any quality of life measures used in modelling are described and use justified • Any other outcomes that were considered but rejected are described with rationale for rejection • Ethical and equity-related outcomes are considered and included when appropriate
Resource use/costs	• All resources used are identified, valued appropriately, and included in analyses • Methods for costing are reporting (e.g., patient level) • Resource quantities and unit costs are both reported • Methods for costing time (e.g., lost time, productivity losses) are appropriate and a justification is provided if time costs are not considered
Uncertainty	• Sources of uncertainty in analyses are identified and justification for probability distributions used in probabilistic analyses are given • For scenario analyses, values and assumptions tested are provided and justified
Results	• All results are presented in a disaggregated fashion, by component, in addition to an aggregated manner • All results are presented with undiscounted totals before discounting and aggregation • Natural units are presented along with alternative units (e.g., QALYs) • The components of incremental cost-effectiveness ratio (ICER) are shown (e.g., mean costs of each intervention in numerator and mean outcomes of each intervention in denominator) • Results of scenario analyses, including variability in factors such as practice patterns and costs, are reported and described in relation to reference (base) case
Interest disclosure	• Disclosures of interest are provided for authors/funders/commissioners of study • Interests are unlikely to significantly affect study validity
Funding source	• There is a description of source(s) of funding • Funding source is unlikely to have a significant impact on study validity

Abbreviations. ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year.

Table B4. Methodological Quality Assessment: Clinical Practice Guidelines

Domain	Domain Elements Assessment indicates how well guideline methodology and development process were performed to limit bias and ensure validity for elements in domain (each domain rated as *good*, *fair*, or *poor* overall based on performance and documentation of elements)
Rigor of development: evidence	• Systematic literature search meets quality standards for a systematic review (i.e., comprehensive search strategy with, at a minimum, 2 or more electronic databases) • The criteria used to select evidence for inclusion is clear and appropriate • The strengths and limitations of individual evidence sources is assessed and overall quality of body of evidence assessed
Rigor of development: recommendations	• Methods for developing recommendations clearly described and appropriate • There is an explicit link between recommendations and supporting evidence • The balance of benefits and harms is considered in formulating recommendations • The guideline has been reviewed by external expert peer reviewers • The updating procedure for guideline is specified in guideline or related materials (e.g., specialty society website)
Editorial independence	• There is a description of source(s) of funding and views of funder(s) are unlikely to have influenced content or validity of guideline • Disclosures of interests for guideline panel members are provided and are unlikely to have a significant impact on overall validity of guideline (e.g., a process for members to recuse themselves from participating on recommendations for which a significant conflict is provided)
Scope and purpose	• Objectives specifically described • Health question(s) specifically described • Target population(s) for guideline recommendations is specified (e.g., patients in primary care) and target users for guideline (e.g., primary care clinicians)
Stakeholder involvement	• Relevant professional groups represented • Views and preferences of target population(s) sought (e.g., clinicians and patients)
Clarity and presentation	• Recommendations are specific and unambiguous • Different management options are clearly presented • Key recommendations are easily identifiable
Applicability	• Provides advice and/or tools on how recommendation(s) can be put into practice • Description of facilitators and barriers to its application • Potential resource implications considered • Criteria for implementation monitoring, audit, and/or performance measures based on guideline are presented