

# Evidence-based Practice Center Systematic Review Protocol

## Project Title: Closing the Quality Gap Series: Revisiting the State of the Science

### The Effects of Bundled Payment Strategies on Health Care Spending and Quality of Care

#### I. Background and Objectives for the Systematic Review

This review is part of the Closing the Quality Gap series, which aims to provide critical analysis of the existing literature on quality improvement strategies for a selection of diseases and practices. The review will focus on “bundled payment,” a strategy for health care quality improvement and cost containment. Other reviews in the series will address a range of quality improvement topics arising from Agency for Healthcare Research and Quality (AHRQ) portfolios.

We define “bundled payment” as a health care provider payment method in which the payment is related to the predetermined expected costs of an episode of care. The definition of “bundled payment” used in this review will include several related concepts that have been referred to as “bundling,” “packaging,” “episode-based payment,” and “warranties.” These concepts refer to different ways to aggregate services into a single unit of payment. Specific payment models may include some or multiples of these aggregation methods. We differentiate between the following different types of aggregation:

- Aggregation of services *longitudinally in time* for an episode of care. The episode is defined to encompass services related to a health care treatment or condition within a defined time window. For example, a single payment could include a surgical procedure and followup care. Distinctions are also sometimes made between “packaging” of services provided during a single patient encounter and “bundling” of services during multiple visits.
- Aggregation of services *across providers* who may be practicing in different care settings. For example, a single payment could be made for inpatient hospital facility services and physician professional services during an inpatient stay.
- *Warranties* refer to payment arrangements where payment for services related to treatment complications is aggregated into the unit of payment. Providers assume financial risk for the cost-of-care defects above a predetermined amount.

We differentiate between the above types of payment methods, in which payment is related to an episode of related services, and payment methods such as global payment or capitation, where payment is made for management of a defined patient population.

To some extent, the notion of bundling is inherent in many current provider payment methods. For example, current Medicare payments for coronary artery bypass graft surgery include a single payment to the hospital for the facility portion of all services during the inpatient stay and a single payment to the surgeon covering services in a 90-day “global period” including the surgical procedure and routine preprocedural and postprocedural services. Newer bundled payment methods are distinguished by the inclusion of multiple providers in disparate care settings that were previously paid separately.<sup>1</sup> For example, a bundled payment for coronary artery bypass graft surgery could include presurgical services, facility and physician fees for the inpatient surgical procedure, and followup care, including monitoring and cardiac rehabilitation and treatment of complications. In addition, the bundles in some newer payment programs span longer time periods than previously used payment methods. Bundled payment program designs are expected to vary widely in terms of what types of procedures or diagnoses are used as the “anchor” of the episodes of care, what other types of services are included in the episode, how the time period of the episode is delimited, and what types of providers are included.

While all bundled payment programs share the common element of payment related to the expected cost of an episode of care, specific payment mechanisms are expected to vary. One possible mechanism is to make a single, prospective payment for the episode of care, similar to Medicare prospective payments for inpatient care admissions. Other mechanisms could include a blend of payment methods. A “shared savings” approach would blend retrospective fee-for-service reimbursement with periodic bonus payments equal to a share of the difference between actual and expected payments for episodes, given that actual costs are lower than a threshold level set below expected costs. Some bundled payment programs include an element of pay-for-performance related to quality measures.<sup>2</sup> Bundled payments could also be used in conjunction with (or in addition to) other new payment and delivery models, such as shared savings for accountable care organizations or medical homes.<sup>2</sup> The review will not be limited by the types of financial incentives employed; all of the above methods will be included.

The intent of bundled payment systems is to decrease health care spending while improving the quality of care.<sup>3</sup> Bundled payment would create a financial incentive for providers to reduce the number and cost of services contained in the bundle.<sup>4</sup> Providers would have discretion over how to allocate their resources in order to treat the patient’s episode most effectively. In particular, Miller has postulated that bundled payment could motivate and enable providers to eliminate services that are low value (from the perspective of health outcomes), duplicative, or unnecessary.<sup>1</sup> Another potential effect would be to encourage coordination of care by holding multiple providers in multiple settings jointly accountable, through shared payment, for the total cost of care for a given treatment or condition.<sup>4</sup>

Several types of undesired effects of bundled payment have also been postulated. Providers could potentially increase the number of episodes provided.<sup>1</sup> Instead of eliminating low-value services, bundled payment could lead to underuse of appropriate services, with potential adverse effects on patient outcomes.<sup>4</sup> In the absence of robust risk adjustment of bundled payments, providers may select low-risk patients and avoid those with higher risks (and costs).<sup>4</sup> Concerns have also been raised about the

administrative feasibility of bundled payment programs, particularly in establishing accountability and a mechanism for distributing payment among otherwise independent providers who participate in an episode.<sup>5,6</sup>

Given the uncertainties about the effects of bundled payment on spending and quality, a review of the evidence on its effects is needed. This review should help readers to 1) understand what the evidence shows about the effects of bundled payment on health care spending and quality of care and 2) understand key design and contextual features of bundled payment programs and their association with effectiveness.

## II. The Key Questions

### Question 1

What does the evidence show on the effects of bundled payment versus usual (predominantly fee-for-service) payment on health care spending and quality?

- **Population(s).** Individuals receiving medical services reimbursed through bundled payment and comparison groups receiving the same services reimbursed through conventional payment.
- **Interventions.** “Bundled payment”: payment related to the expected cost of an episode of care defined around a particular treatment or condition. As described below, bundled payment systems vary in important ways, including the definition of the unit of payment (“bundle”), the payment mechanism, and payment adjustments for patient risk and quality of care.
- **Comparators.** The comparators will be “usual” payment methods that could include a range of methods, including fee-for-service, per-diem, per-discharge, and capitation payments. We will not limit our study to any specific comparators. We expect that in most studies, the “usual” payment comparator will include a mix of payment methods with heterogeneity both within providers (providers reimbursed via different methods by different payers) and between providers (e.g., geographic variation in the prevalence of capitation).
- **Study Outcomes.** The main types of study outcomes of interest are health care spending/resource utilization and quality of care. Of secondary interest is evidence that bundled payment led to specific types of care redesign intended to affect spending and quality. We will also abstract information on effects on the risk profiles of treated patient populations, since risk selection is a potential adverse effect of bundled payment. Study outcomes include:
  - Health care spending (allowed charges) per episode.
  - Health care spending per capita.
  - Utilization rates for specific services (e.g., readmission rate).

- Utilization rates for episodes of care.
  - Provider cost/resource use to deliver episodes (e.g., cost per implanted device, average length of inpatient stay).
  - Provider financial risk.
  - Administrative cost of payment method.
  - Quality of care, considered in the following categories used by the National Quality Measures Clearinghouse<sup>7</sup>:
    - Structure: a feature of a health care organization or clinician relevant to its capacity to provide health care.
    - Process: a health care service provided to, on behalf of, or by a patient appropriately based on scientific evidence of efficacy or effectiveness.
    - Outcome: a health state of a patient resulting from health care.
    - Access: a patient's or enrollee's attainment of timely and appropriate health care.
    - Patient experience of care: a patient's or enrollee's report concerning observations of and participation in health care.
  - Care redesign by providers (descriptive information on responses to bundled payment, such as use of practice-embedded care coordinators, changes in referral practices, and changes in implant purchasing methods).
  - “Unbundling”—that is, behavior of providers that results in separate payment for bundled service, such as moving the date of service outside of the time window of the unit of payment.
  - Average risk/severity of patients treated. Studies may report on this as a measure of a potential adverse effect of bundled payment on risk selection or access to care.
- **Timing.** Minimum duration of followup equal to length of episode.
  - **Settings.** Health care providers participating in bundled payment programs and comparison providers.

## Question 2

Does the evidence show differences in the effects of bundled payment systems by key design features?

- **Population(s).** Same as for Key Question (KQ) 1.
- **Interventions.** Subsets of the bundled payment systems in KQ 1, characterized by the following key design features:
  - Definition of the “bundle.”
    - “Anchor” of the bundle: acute condition, chronic condition, major procedure, and minor procedure.
    - Types of services included (inpatient, ambulatory, postacute, etc.).

- Payment methodology used (e.g., prospective payment, shared savings, pay-for-performance bonuses).
- Risk-adjustment methods.
- Use of quality measurement for adjustment of payment or eligibility thresholds.
- Method of distribution of bundled payment among participating providers.
- **Comparators.** Same as for KQ 1.
- **Study Outcomes.** Same as for KQ 1.
- **Timing.** Same as for KQ 1.
- **Settings.** Same as for KQ 1.

### Question 3

Does the evidence show differences in the effects of bundled payment systems by key contextual factors?

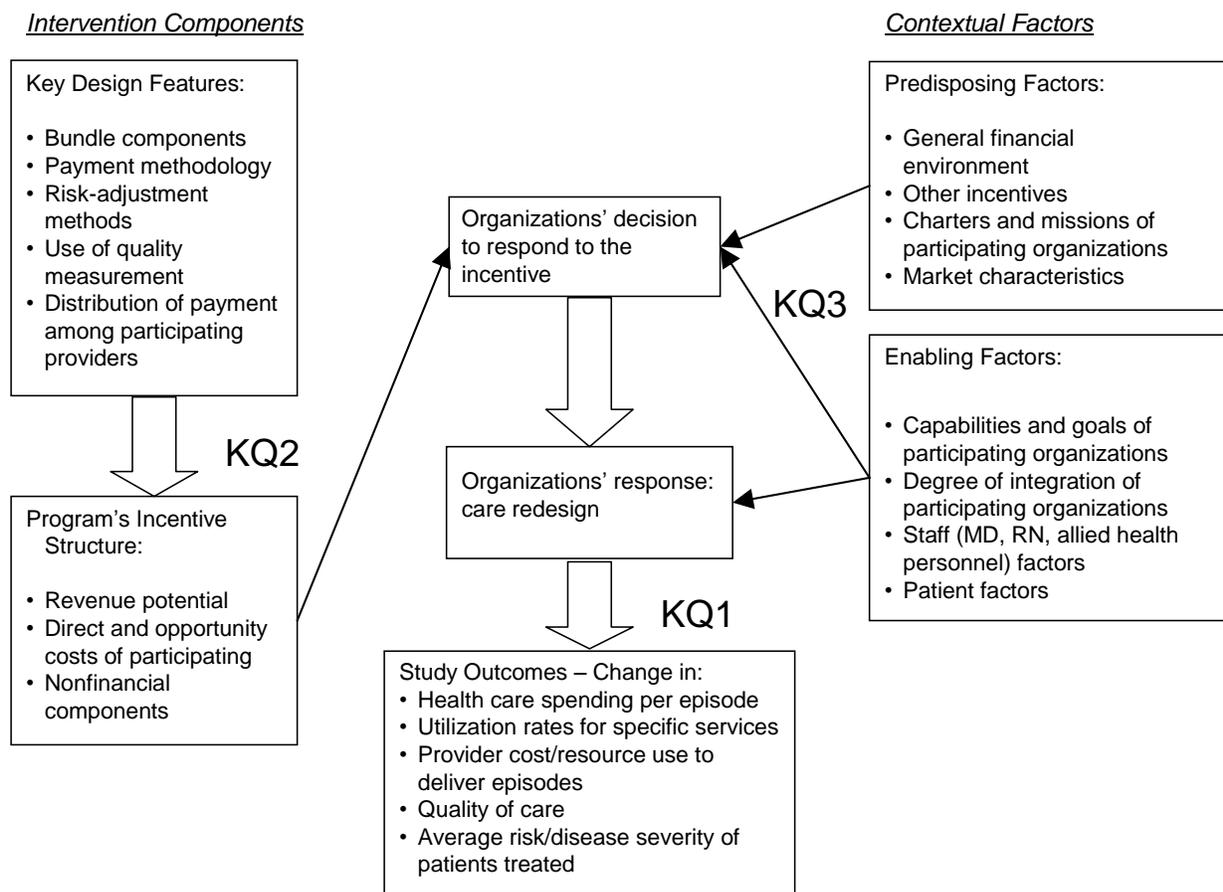
- **Population(s).** Same as for KQ 1.
- **Interventions.** Subsets of the bundled payment systems in KQ 1, characterized by the following key contextual factors, as well as others noted in the conceptual framework:
  - Types of health care–delivery organizations included.
  - Degree of integration of health care–delivery organizations.
  - Number of payers involved, market share characteristics, and relationship with participating providers.
  - Competitiveness of market for payers and health care–delivery organizations.
- **Comparators.** Same as for KQ 1.
- **Study Outcomes.** Same as for KQ 1.
- **Timing.** Same as for KQ 1.
- **Settings.** Same as for KQ 1.

### III. Analytic Framework

We propose the conceptual model in Figure 1 to understand the response to the implementation of a bundled payment model among organizations participating in the delivery of an episode of care. This model is based on one developed by Dudley et al.<sup>8</sup> to describe the response of organizations to payment incentives in general. It draws from the health services research literature and incorporates more general economic concepts, such as opportunity costs, that often are not addressed in research about specific incentives. The original model was grounded in Andersen’s Behavioral Model of Health Services Use.<sup>9</sup> The Andersen model was modified to apply to organizations rather than to individuals seeking access to care.

In the figure below, we propose several key design features that will define a particular set of incentives (and disincentives) associated with any specific bundled payment strategy. The impact of these design features are addressed by KQ 2. The financial and nonfinancial characteristics of these incentives are primary determinants of the “need” an organization has to change its practice in response to the modified payment policy. This response, however, may be mediated by key contextual factors, including both predisposing and enabling factors. Predisposing factors include the general financial environment, other incentives outside of the bundled payment program, market variables, and characteristics of participating provider organizations such as charter and mission. Enabling factors include the capabilities and goals of participating organizations, the degree to which these organizations are integrated, and staff and patient-level characteristics. The impact of these contextual factors is addressed by KQ 3. The center of the model reflects how organizations respond to the incentives created by bundled payment through care redesign. KQ 1 addresses how different potential responses affect study outcomes, including health care spending, health care quality, and the other outcomes listed above in the description of KQ 1.

**Figure 1.** Analytic framework for review of the effects of bundled payment strategies on health care spending and quality of care



Adapted from:

Source: [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov)

Published Online: June 23, 2011

Dudley RA, Frolich A, Robinowitz DL, et al. Strategies To Support Quality-based Purchasing: A Review of the Evidence. Technical Review No. 10. (Prepared by Stanford–University of California San Francisco Evidence-based Practice Center under Contract No. 290-02-0017). Rockville, MD: Agency for Healthcare Research and Quality; July 2004. AHRQ Publication No. 04-0057.

Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav* 1995 Mar;36(1):1-10. PMID: 7738325.

## IV. Methods

### A. Criteria for Inclusion/Exclusion of Studies in the Review

Studies will be included that address the populations, interventions, comparators, and study outcomes described above. All study designs will be included, including experimental, observational, and descriptive studies. Relevant grey literature, including government reports and other material identified from sources listed in Section B, will be included. The publication date will be limited to January 1, 1985, and later, because health care financing has changed over time, limiting the generalizability of earlier findings to the current health care system. If considered necessary, study authors will be contacted for additional data.

The following studies will be excluded: 1) studies that did not report any of the outcomes of interest; 2) studies that did not report on a bundled payment intervention as defined above; and 3) background articles.

Studies of interventions implemented in countries other than the United States will be included only if they meet broad criteria for generalizability to the United States. These criteria include:

- The country's delivery system provides similar types of services to the U.S. system (i.e., not low-income countries that provide a much different mix of services).
- The comparison payment method is predominantly fee-for-service as in the United States (e.g., not salary).
- The delivery context in which the intervention was implemented is similar to one that exists somewhere in the United States.
- The bundled payment intervention meets other inclusion criteria and the study addresses the key study outcomes of interest.

The search strategy will not use language restrictions; studies in other languages that fit all other inclusion criteria will be included if the necessary translation expertise is available. The final report will note how many studies were excluded due to language constraints and whether that is likely to affect the conclusions.

## **B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies To Answer the Key Questions**

The objective of the search strategy is to identify all published bundled payment evaluations.

A librarian will perform the initial literature search. Two trained reviewers will scan the titles/abstracts of the list run by the librarian and select studies for full-text screen. For each of the selected studies, reviewers will perform further reference mining by scanning titles listed in the reference section to identify additional articles to be included. Reviewers will reconcile their selections and make joint decisions, following all the inclusion/exclusion criteria listed in previous sections.

We propose to use the following search terms:

bundl\*[tiab] OR episode[tiab] OR “prospective payment”[tiab] OR warranty[tiab] OR warranti\*[tiab]

AND

payment[tiab] OR finance\*[tiab] OR reimburse\*[tiab] OR incentive\*[tiab]

AND

trial[tiab] OR compare\*[tiab] OR effect\*[tiab] OR impact[tiab] OR outcome\*[tiab]

We propose to search the following sources:

- Databases
  - Cochrane Library of systematic reviews
  - PubMed (National Library of Medicine, includes MEDLINE)
- Other sources
  - Reports from government agencies including GAO, CMS (also contractors)
  - [Clinicaltrials.gov](http://Clinicaltrials.gov)
  - References of included studies
  - References of relevant reviews
  - Citation tracking of included studies using Web of Science
  - Personal files from related topic projects
  - References provided by Technical Expert Panel members

## **C. Data Abstraction and Data Management**

Data will be independently abstracted by two researchers trained in the critical assessment of evidence. The following data will be abstracted from included studies:

- Trial name.
- Setting and context (including but not limited to number of payers involved, market share, payer relationship with participating providers).
- Provider population characteristics (including but not limited to provider organization type[s], provider organization staffing, and profit status);
- Patient population characteristics (including but not limited to sex, age, ethnicity, diagnosis and/or disease severity, and baseline health care utilization).
- Eligibility and exclusion criteria.
- Interventions (including “anchor” procedure or diagnosis, services included in the bundle, payment methodology used, risk-adjustment methods, and use of quality measurement).
- Any cointerventions.
- Comparisons.
- Results for each outcome.
- Funding source.

#### **D. Assessment of Methodological Quality of Individual Studies**

We will assess the methodological quality of individual studies following methodology outlined in the AHRQ *Methods Guide for Effectiveness and Comparative Effectiveness Research* (hereafter, *Methods Guide*).<sup>10</sup> Each individual study will be given a summary rating of good (low risk of bias), fair, or poor (high risk of bias). Studies rated “poor” will also be given a brief explanation of the basis for the rating. The rating will be based on the following list of criteria.

- Several core elements apply to trials, as well as to observational studies:
  - Similarity of groups at baseline in terms of baseline characteristics and prognostic factors.
  - Extent to which valid primary outcomes were described.
  - Blinding of subjects and providers.
  - Blinded assessment of the outcome.
  - Intention-to-treat analysis.
  - Differential loss to followup between the compared groups or overall high loss to followup.
  - Conflict of interest.
- For trials, two additional elements are important:
  - Methods used for randomization.
  - Allocation concealment.

- For observational studies (which are expected to represent most or all of the reviewed studies), still other additional elements will be considered:
  - Sample size.
  - Methods for selecting participants (inception cohort, methods to avoid selection bias).
  - Methods for measuring exposure variables.
  - Methods for dealing with any design-specific issues, such as recall bias, interviewer bias, et cetera.
  - Analytical methods to control confounding.

## E. Data Synthesis

Our a priori analytic plan is to summarize the evidence for effectiveness of bundled payment in comparison with usual payment methods. The evidence of risks (e.g., patient selection) will also be summarized. We do not plan to conduct any quantitative synthesis of results, because we expect low similarity between studies. Heterogeneity will be assessed, based on analysis of the data abstracted on intervention design and contextual factors.

We will perform stratified analyses by study type (e.g., cohort, cross-sectional) and possibly other dimensions, such as key design features of the bundled payment programs (e.g., bundled payment for acute vs. chronic care episodes). We will perform a narrative synthesis. Major findings of the studies will be further presented in tables to compare different interventions.

## F. Grading the Evidence for Each Key Question

We will assess the overall strength of evidence for intervention effectiveness by using guidance outlined in the *Methods Guide*.<sup>10</sup> This method is based loosely on one developed by the Grade Working Group and classifies the grade of evidence according to the following criteria:

**High** = High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.

**Moderate** = Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.

**Low** = Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.

**Insufficient** = Evidence is either unavailable or does not permit estimation of an effect.

The evidence grade is based on four primary domains (required) and two optional domains. The required domains are risk of bias, consistency, directness, and precision; the additional domains are coherence, residual confounding, and strength of association. One other optional domain, dose-response association, will not be used in this review. Publication bias will also be considered; while not considered a separate domain of strength of evidence, it is related to strength of evidence, particularly consistency and precision. A brief description of these domains is displayed in Appendix A. For this review, we will use both this explicit scoring scheme and the global implicit judgment about “confidence” in the result. Where the two disagree, we will choose the lower classification.

### **G. Applicability**

The purpose of reviewing the literature is to consider what strategies can be generalized to other practices. Studies that are conducted among highly selected samples of patients, or have limited complexities of treatment, or are pilot projects in geographic areas and delivery systems with relatively advanced predisposing and enabling characteristics will likely have lower generalizability and external validity for implementation in other environments. We will consider interventions that have consistent findings in multiple environments as having a greater likelihood of greater generalizability.

We will assess the applicability of the studies reviewed, providing a summary of the most important characteristics of the body of reviewed studies that affect applicability and a description of their expected effects on applicability. In addition to the population and setting factors discussed above, we will examine other elements of the PICOTS (population, intervention, comparator, outcomes, timing, and setting) framework to characterize threats to applicability. Information collected on the interventions and context for KQs 2 and 3 will inform this analysis.

## V. References

1. Miller HD. From volume to value: better ways to pay for health care. *Health Aff (Millwood)* 2009 Sep-Oct;28(5):1418-28. PMID: 19738259.
2. Schneider EC, Hussey PS, Schnyer C. *Payment reform: analysis of models and performance measurement implications*. Santa Monica, CA: RAND; 2011.
3. Rosenthal MB. Beyond pay for performance—emerging models of provider-payment reform. *N Engl J Med* 2008 Sep 18;359(12):1197-200. PMID: 18799554.
4. Medicare Payment Advisory Commission. *Report to the Congress: reforming the delivery system*. Washington, DC: MedPAC; June 2008.
5. Goldsmith J. Analyzing shifts in economic risks to providers in proposed payment and delivery system reforms. *Health Aff (Millwood)* 2010 Jul;29(7):1299-304. PMID: 20606177.
6. Corrigan J, McNeill D. Building organizational capacity: a cornerstone of health system reform. *Health Aff (Millwood)* 2009 Mar-Apr;28(2):w205-w15. PMID: 19174381.
7. National Quality Measures Clearinghouse (NQMC). 2010. (Accessed July 9, 2010, at <http://www.qualitymeasures.ahrq.gov/selecting-and-using/using.aspx>)
8. Dudley R, Frolich A, Robinowitz D, et al. *Strategies To Support Quality-based Purchasing: A Review of the Evidence*. Technical Review No, 10 (Prepared by Stanford–Ujiversity of California San Francisco Evidence-base Practice Center under Contract No. 290-02-0017). Rockville, MD: Agency for Healthcare Research and Quality; July 2004. AHRQ Publication No. 04-0057.
9. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav* 1995 Mar;36(1):1-10. PMID: 7738325.
10. *Methods Guide for Effectiveness and Comparative Effectiveness Review*. Rockville, MD: Agency for Healthcare Research and Quality; March 2011. AHRQ Publication No. 10(11)-EHC063-EF. Chapters available at: [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov).

## VI. Definition of Terms

1. **Bundled payment:** a health care provider payment method in which the payment is related to the predetermined expected costs of an episode of care.
2. **Usual payment:** payment methods currently in use including fee-for-service, per-diem, capitation, and per-discharge payment.
3. **Episode of care:** encompasses services related to a medical treatment or condition, within a defined time window, and typically spanning multiple providers and care settings.

## VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

## VIII. Review of Key Questions

For all EPC reviews, key questions were reviewed and refined as needed by the EPC with input from the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.

## IX. Technical Experts

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

## X. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than \$10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

## Appendix A: Grading the strength of a body of evidence: required domains and their definitions

Domain	Definition and Elements	Score and Application
Risk of Bias	<p>Risk of bias is the degree to which the included studies for a given outcome or comparison have a high likelihood of adequate protection against bias (i.e., good internal validity), assessed through two main elements:</p> <ul style="list-style-type: none"> <li>• Study design (e.g., RCTs or observational studies).</li> <li>• Aggregate quality of the studies under consideration.</li> </ul> <p>Information for this determination comes from the rating of quality (good/fair/poor) done for individual studies.</p>	<p>Use one of three levels of aggregate risk of bias:</p> <ul style="list-style-type: none"> <li>• Low risk of bias</li> <li>• Medium risk of bias</li> <li>• High risk of bias</li> </ul>
Consistency	<p>The principal definition of consistency is the degree to which reported effect sizes from included studies appear to have the same direction of effect. This can be assessed through two main elements:</p> <ul style="list-style-type: none"> <li>• Effect sizes have the same sign (i.e., are on the same side of “no effect”).</li> <li>• The range of effect sizes is narrow.</li> </ul>	<p>Use one of three levels of consistency:</p> <ul style="list-style-type: none"> <li>• Consistent (i.e., no inconsistency)</li> <li>• Inconsistent</li> <li>• Unknown or not applicable (e.g., single study)</li> </ul> <p>As noted in the text, single-study evidence bases (even mega-trials) cannot be judged with respect to consistency. In that instance, use: “Consistency unknown (single study).”</p>
Directness	<p>The rating of directness relates to whether the evidence links the interventions directly to health outcomes. For a comparison of two treatments, directness implies that head-to-head trials measure the most important health or ultimate outcomes.</p> <p>Two types of directness, which can coexist, may be of concern. Evidence is indirect if:</p> <ul style="list-style-type: none"> <li>• It uses intermediate or surrogate outcomes instead of health outcomes. In this case, one body of evidence links the intervention to intermediate outcomes and another body of evidence links the intermediate to the most important (health or ultimate) outcomes.</li> <li>• It uses two or more bodies of evidence to compare interventions A and B—e.g., studies of A vs. placebo and B vs. placebo, or studies of A vs. C and B vs. C but not A vs. B.</li> </ul> <p>Indirectness always implies that more than one body of evidence is required to link interventions to the most important health outcomes.</p> <p>Directness may be contingent on the outcomes of interest. EPC authors are expected to make clear the outcomes involved when assessing this domain.</p>	<p>Score dichotomously as one of two levels of directness:</p> <ul style="list-style-type: none"> <li>• Direct</li> <li>• Indirect</li> </ul> <p>If indirect, specify which of the two types of indirectness account for the rating (or both, if that is the case)—namely, use of intermediate/ surrogate outcomes rather than health outcomes and use of indirect comparisons. Comment on the potential weaknesses caused by, or inherent in, the indirect analysis. The EPC should note if both direct and indirect evidence was available, particularly when indirect evidence supports a small body of direct evidence.</p>
Precision	<p>Precision is the degree of certainty surrounding an effect estimate with respect to a given outcome (i.e., for each</p>	<p>Score dichotomously as one of two levels of precision:</p>

Domain	Definition and Elements	Score and Application
	<p>outcome separately).</p> <p>If a meta-analysis was performed, this will be the confidence interval around the summary effect size.</p>	<ul style="list-style-type: none"> <li>• Precise</li> <li>• Imprecise</li> </ul> <p>A precise estimate is an estimate that would allow a clinically useful conclusion. An imprecise estimate is one for which the confidence interval is wide enough to include clinically distinct conclusions. For example, results may be statistically compatible with both clinically important superiority and inferiority (i.e., the direction of effect is unknown), a circumstance that will preclude a valid conclusion.</p>
Coherence	<p>Coherence is the degree of plausibility of results in relation to epidemiology or, in some cases, biology and pathophysiology.</p>	<p>This additional domain does not need to be described or noted unless something “implausible” has emerged, in which case EPC authors should comment on it.</p>
Residual confounding	<p>Occasionally, in an observational study, residual confounders would work in the direction opposite that of the observed effect. A case in point is when a study is biased against finding an effect and yet it finds an effect. Thus, had these confounders not been present, the observed effect would have been even larger than the one observed.</p>	<p>Score as three levels:</p> <ul style="list-style-type: none"> <li>• Confounding unlikely to explain observed effect: Plausible residual confounders are more likely to have decreased the observed effect than to have increased the observed effect.</li> <li>• Confounding may explain observed effect: Plausible residual confounders are unlikely to have decreased the observed effect and could be responsible for observed effect.</li> <li>• Cannot assess.</li> </ul>
Strength of association (magnitude of effect)	<p>Strength of association refers to the likelihood that the observed effect is large enough that it cannot have occurred solely as a result of bias from potential confounding factors.</p>	<p>Score as two levels:</p> <ul style="list-style-type: none"> <li>• Strong: large effect size that is unlikely to have occurred in the absence of a true effect of the intervention (e.g., relative risk &gt;5).</li> <li>• Weak: small enough effect size that it could have occurred solely as a result of bias from confounding factors (e.g., relative risk &lt;5).</li> </ul>

Abbreviations: EPC = Evidence-based Practice Center; RCT = randomized controlled trial.