Chapter 22: The Process of a Systematic Review and Meta-analysis

M. Hassan Murad; Roman Jaeschke; PJ Devereaux; Kameshwar Prasad; Alonso Carrasco-Labra; Thomas Agoritsas; Deborah J. Cook; Gordon Guyatt

Clinical Scenario

CLINICAL SCENARIO

Should Patients Undergoing Noncardiac Surgery Receive β-Blockers?

You receive a request for consultation from a general surgeon regarding the perioperative management of a 66-year-old man undergoing hip replacement surgery in 2 days. The patient has a history of type 2 diabetes and hypertension and is a smoker. He has no history of heart disease. The patient's blood pressure is 135/80 mm Hg. Because the patient has multiple *risk factors* for heart disease, you are considering whether he should be treated perioperatively with β-blockers to reduce the risk of death, nonfatal myocardial infarction, and other vascular complications.

Finding the Evidence

Being aware that a large amount of literature exists on this controversial topic, you decide to conduct a search that will provide you with an accurate and rapid overview of current best *evidence*. Because the question is about therapy, you are particularly interested in finding a recent *systematic review* and *meta-analysis* of *randomized clinical trials* (RCTs) that deal with this topic. Using the free *federated search engine* ACCESSSS (http://plus.mcmaster.ca.ezproxy.library.dal.ca/accessss; see Chapter 5, Finding Current Best Evidence), you enter these search terms: beta blockers, perioperative, and mortality.

Starting with the summaries at the top of your search output, you locate 2 relevant preappraised summaries on the "management of cardiac risk for noncardiac surgery." Both summaries cite the results of a large systematic review and meta-analysis published in 2008,[1] along with references to current US and European *clinical practice guidelines*. However, you notice that the last updates of these chapters date back 4 to 6 months ago. You therefore look further down in your search output to check preappraised research (see Chapter 5, Finding Current Best Evidence) and rapidly identify a more recently published systematic review and meta-analysis addressing your question and that was highly rated for relevance and newsworthiness by clinicians from 4 specialties.[2] You download the full text of the article reporting this meta-analysis.

Systematic Reviews and Meta-Analysis: An Introduction

Definitions

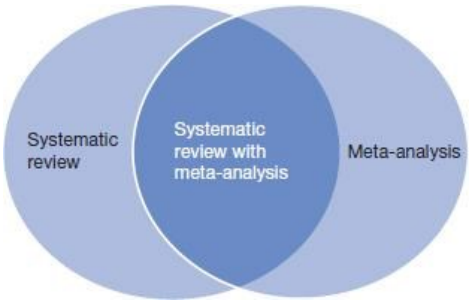
A systematic review is a summary of research that addresses a focused clinical question in a systematic, reproducible manner. Systematic reviews can provide estimates of therapeutic efficacy, *prognosis*, and diagnostic test accuracy and can summarize the evidence for questions of "how" and "why" addressed by *qualitative research* studies. Although we will refer to other sorts of questions, this chapter focuses on systematic reviews that address the effect of therapeutic interventions or harmful *exposures* on *patient-important outcomes*.

A systematic review is often accompanied by a meta-analysis (a statistical pooling or aggregation of results from different studies) to provide a single best estimate of effect. The pooling of studies increases precision (ie, narrows the *confidence intervals* [CIs]), and the single best effect estimate generated facilitates clinical decision making. Therefore, you may see a published systematic review in which the authors chose not to do a meta-

analysis, and you may see a meta-analysis conducted without a systematic review (ie, studies were combined statistically but were not selected following a comprehensive, explicit, and reproducible approach) (Figure 22-1. Most useful clinically will be a well-performed systematic review—the methods for which we describe in this chapter—with an accompanying meta-analysis.

FIGURE 22-1

The Overlap of Study Designs: Systematic Review and Meta-analysis



In contrast to systematic reviews, traditional *narrative reviews* typically address multiple aspects of the disease (eg, etiology, diagnosis, prognosis, or management), have no ex-plicit criteria for selecting the included studies, do not include systematic assessments of the *risk of bias* associated with *primary studies,* and do not provide quantitative best estimates or rate the confidence in these estimates. The traditional narrative review articles are useful for obtaining a broad overview of a clinical condition but may not provide a reliable and unbiased answer to a focused clinical question.

Why Seek Systematic Reviews?

When searching for evidence to answer a clinical question, it is preferable to seek a systematic review, especially one that includes a meta-analysis, rather than looking for the best individual study or studies. The reasons include the following:

1. Single studies are liable to be unrepresentative of the total body of evidence, and their results may therefore be misleading.
2. Collecting and appraising a number of studies take time you probably do not have.
3. A systematic review is often accompanied by a meta-analysis to provide the best estimate of effect that increases precision and facilitates clinical decision making.
4. If the systematic review is performed well, it will likely provide all of the relevant evidence with an assessment of the best estimates of effect and the confidence they warrant.
5. Systematic reviews include a greater range of patients than any single study, potentially enhancing your confidence in applying the results to the patient before you.

A Synopsis of the Process of a Systematic Review and Meta-analysis

In applying the *Users' Guides*, you will find it useful to have a clear understanding of the process of conducting a systematic review and meta-analysis. Figure 22-2 shows how the process begins with the definition of the question, which is synonymous with specifying eligibility criteria for deciding which studies to include in a review. These criteria define the population, the exposures or interventions, and the outcomes of interest. A systematic review also may restrict studies to those that minimize the risk of bias. For example, systematic reviews that address a question of therapy often will include only RCTs.

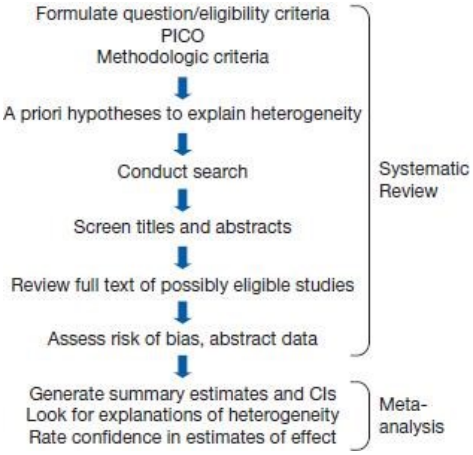
FIGURE 22-2

The Process of Conducting a Systematic Review and Meta-analysis

In a systematic review without meta-analysis, the step of generating summary estimates and confidence intervals is not applicable. If the systematic review includes a meta-analysis and presents estimates of effect from individual studies, seeking explanation for heterogeneity and rating confidence

in estimates is possible.

Abbreviations: CI, confidence interval; PICO, Patient, Intervention, Comparison, Outcome.



Having specified their selection criteria, reviewers will conduct a comprehensive search of the literature in all relevant medical databases, which typically yields a large number of potentially relevant titles and abstracts. They then apply the selection criteria to the titles and abstracts, arriving at a smaller number of articles that they retrieve. Once again, the reviewers apply the selection criteria, this time to the complete reports.

Having completed the culling process, the reviewers assess the risk of bias of the individual studies and abstract data from each study. Finally, they summarize the results, including, if appropriate, a quantitative synthesis or meta-analysis. The meta-analysis provides *pooled estimates* (ie, combined estimates) of the effect on each of the outcomes of interest, along with the associated CIs. Meta-analyses frequently include an examination of the differences in effect estimates across included studies in an attempt to explain differences in results (exploring *heterogeneity*). If based on previously specified hypotheses about possible differences in patients, interventions, or outcomes that may explain differences in results, such explorations become more credible (see Chapter 25.2, How to Use a Subgroup Analysis).

Judging the Credibility of the Effect Estimates

When applying the results of a systematic review to patient care, you can look for estimates of effect. A systematic review without a meta-analysis typically presents results from individual studies; the meta-analysis adds a single pooled (combined) estimate of effect, with an associated CI, for each relevant outcome. Pooled estimates could be for therapy outcomes (eg, death, myocardial infarction, quality of life, late catastrophic adverse effects), estimates of the properties of diagnostic tests (eg, *likelihood ratios*), or estimates of patients' likely outcomes (eg, prognosis). Clinicians need to know the extent to which they can trust these estimates.

Two fundamental problems can undermine this trust. One is the extent to which systematic review authors have applied rigorous methods in conducting their review. We refer to this as the *credibility* of the review.[3] By credibility, we mean the extent to which the design and conduct of the review are likely to have protected against misleading results.[4] As you will see, credibility may be undermined by eligibility criteria that are inappropriate or not specified, the conduct of an inadequate search, and the omission of risk of bias assessments of individual studies (see Box 22-1 for issues to be considered in the credibility of the review process; these issues are applicable to any systematic review, with or without a meta-analysis).

BOX 22-1

Users' Guides for Credibility of the Systematic Review Process

Did the Review Explicitly Address a Sensible Clinical Question?
Was the Search for Relevant Studies Detailed and Exhaustive?
Was the Risk of Bias of the Primary Studies Assessed?
Did the Review Address Possible Explanations of Between-Study Differences in Results?
Did the Review Present Results That Are Ready for Clinical Application?
Were Selection and Assessments of Studies Reproducible?
Did the Review Address Confidence in Effect Estimates?

A highly credible review—one that has adhered to methodologic stand-ards—may nevertheless leave us with only very low confidence in estimates of effect. Common reasons for this include the following: the individual studies may be plagued by high risk of bias and inconsistent results, even the pooled (combined) sample sizes may be small and the results may be imprecise, and the patients enrolled in the studies may differ in important ways from those in whom we are interested. This chapter deals with credibility assessment of the review process; the next chapter (Chapter 23, Understanding and Applying the Results of a Systematic Review and Meta-analysis) will guide you in deciding how much confidence we can place on estimates of effect in the presence of a credible review process.

Was the Process Credible?

Did the Review Explicitly Address a Sensible Clinical Question?

A systematic review has, relative to a traditional narrative review, a narrow focus and addresses a specific question that—for questions of therapy or harm—is defined by particular patients, interventions, comparisons, and outcomes. When review authors conduct a meta-analysis, the issue of how narrow or wide is the scope of the question becomes particularly important. Let us look at these hypothetical examples of 4 meta-analyses with varying scope:

1. A meta-analysis that pooled results from all modalities of cancer therapy for all types of cancer to generate a single estimate of the effect on mortality.

2. A meta-analysis that pooled the results of the effect of all doses of all antiplatelet agents (including aspirin, sulfinpyrazone, dipyridamole, ticlopidine, and clopidogrel) on major thrombotic events (including myocardial infarctions, strokes, and acute arterial insufficiency in the lower extremities).

3. A meta-analysis that pooled the results of the effect of all doses of all antiplatelet agents on mortality in patients with clinically manifest atherosclerosis (whether in the heart, brain, or lower extremities).

4. A meta-analysis that pooled the results of the effect of a wide range of aspirin doses to *prevent* thrombotic stroke in patients presenting with a transient ischemic attack (TIA) due to carotid artery disease.

Clinicians will clearly be uncomfortable with the first meta-analysis, which addresses all treatments for all cancers. Clinicians are unlikely to find the second and third meta-analyses on antiplatelet agents in major thrombotic events and mortality useful because they remain too broad. In contrast, most clinicians may be comfortable with the fourth, more focused meta-analysis of aspirin and thrombotic stroke, although they may express concerns about pooling across a wide range of aspirin doses.

What makes a meta-analysis too broad or too narrow? When deciding whether the question posed in the meta-analysis is sensible, clinicians need to ask themselves whether the underlying biology is such that they would anticipate more or less the same *treatment effect* across the range of patients included (Box 22-2). They should ask a parallel question about the other components of the study question: Is the underlying biology such that, across the range of interventions and outcomes studied, they expect more or less the same treatment effect? Clinicians also can construct a similar set of

questions for other areas of clinical inquiry. For example, across the range of patients, ways of testing, and *reference* or *gold standard* for diagnosis, does one expect more or less the same likelihood ratios associated with studies that examine a diagnostic test[5] (see Chapter 18, Diagnostic Tests)?

BOX 22-2

Were Eligibility Criteria for Inclusion in the Systematic Review Appropriate?

Are results likely to be similar across the range of included patients (eg, older and younger, sicker and less sick)?
Are results likely to be similar across the range of studied interventions or exposures (eg, for therapy, higher dose or lower dose; for diagnosis, test results interpreted by experts or nonexperts)?
Are results likely to be similar across the range of ways the outcome was measured (eg, shorter or longer follow-up)?

Clinicians reject a meta-analysis that pools data across all modes of cancer therapy for all types of cancer because they know that some cancer treatments are effective in certain cancers, whereas others are not effective. Combining the results of these studies would yield an estimate of effect that would make little sense or be misleading for most of the interventions. Clinicians who reject the meta-analysis on all antiplatelet agents and mortality in patients with atherosclerosis would argue that the biologic variation in antiplatelet agents is likely to lead to important differences in treatment effect. Furthermore, they may contend that there are important differences in the biology of atherosclerosis in the vessels of the heart, brain and neck, and legs. Those who would endorse this meta-analysis would argue for the similar underlying biology of antiplatelet agents—and atherosclerosis in different parts of the body—and thus anticipate a similar magnitude of treatment effects.

For the last, more focused review, most clinicians would accept that the biology of aspirin action is likely to be similar in patients whose TIA reflected right-sided or left-sided brain ischemia, in patients older than 75 years and in younger patients, in men and women, across different aspirin doses, during periods of *follow-up* ranging from 1 to 5 years, and in patients with stroke who have been identified by the attending physician and those identified by a team of experts. The similar biology is likely to result in a similar magnitude of treatment effect, which explains the comfort of the meta-analysis authors with combining studies of aspirin in patients who have had a TIA.

The clinician's task is to decide whether, across the range of patients, interventions or exposures, and outcomes, it is plausible that the intervention will have a similar effect. This judgment is possible only if the review authors have provided a precise statement of what range of patients, exposures, and outcomes they decided to include; in other words, the explicit eligibility criteria for their review.

In addition, systematic review authors must specify the criteria for study inclusion related to the risk of bias. Generally, these should be similar to the most important criteria used to evaluate the risk of bias in primary studies[6] (Box 22-3). Explicit eligibility criteria not only facilitate the decision regarding whether the question was sensible but also make it less likely that the authors will preferentially include or exclude studies that support their own previous conclusions or beliefs.

BOX 22-3

Guides for Selecting Articles That Are Most Likely to Provide Results at Lower Risk of Bias

Therapy	Were patients randomized?
	Was follow-up complete?
Diagnosis	Was the patient sample representative of those with the disorder?
	Was the diagnosis verified using credible criteria that were independent of the items of medical history, physical examination, laboratory tests, or imaging procedures under study?
Harm	Did the investigators find similarity in all known determinants of outcome or adjust for differences in the analysis?
	Was follow-up sufficiently complete?
Prognosis	Was there a representative sample of patients?
	Was follow-up sufficiently complete?

Clinicians may legitimately ask, even within a relatively narrowly defined question, whether they can be confident that results will be similar across patients, interventions, and outcome measurement. Referring to the question of aspirin use by patients with a TIA, the effect could conceivably differ in those with more or less severe underlying atherosclerosis, across aspirin doses, or during short-term and long-term follow-up. Thus, at the time of examining the results, we need to ask whether the assumption with which we started proved accurate: was the effect the same across patients, interventions, and outcomes? We return to this issue in the next chapter (see Chapter 23, Understanding and Applying the Results of a Systematic Review and Meta-analysis).

Was the Search for Relevant Studies Detailed and Exhaustive?

Systematic reviews are at risk of presenting misleading results if they fail to secure a complete, or at least representative, sample of the available eligible studies. To achieve this objective, reviewers search bibliographic databases. For most clinical questions, searching a single database is insufficient and can lead to missing important studies. Therefore, searching MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials is recommended for most clinical questions.[7] Searching other databases may be required, depending on the nature of the review question. The systematic review authors check the reference lists of the articles they retrieve and seek personal contact with experts in the area. It also may be important to examine recently published abstracts presented at scientific meetings and to look at less frequently used databases, including those that summarize doctoral theses and databases of ongoing trials held by pharmaceutical companies or databases of ongoing registered trials.

Another important source of unpublished studies is the US Food and Drug Admin-istration (FDA) reviews of new drug applications. A study that evaluated the risk of dyspepsia associated with the use of nonsteroidal anti-inflammatory drugs found that searching FDA records yielded 11 trials, of which only 1 was published.[8] Another study of FDA reports found that they included numerous unpublished studies, and the findings of these studies can appreciably alter the estimates of effect.[9] Unless the authors of systematic reviews tell us what they did to locate the studies, it is difficult to know how likely it is that relevant studies were missed.

Reporting bias occurs in a number of forms, the most familiar of which is the failure to report or publish studies with negative results. This *publication bias* may result in misleading results of systematic reviews that fail to include unpublished studies.[10,11]

If authors include unpublished studies in a review, they should try to obtain full reports, and they should use the same criteria to appraise the risk of bias of both published and unpublished studies. There is a variety of techniques available to explore the possibility of publication bias, but none of them are fully satisfactory. Systematic reviews based on a small number of studies with limited total sample sizes are particularly susceptible to

publication bias, especially if most or all of the studies have been sponsored by a commercial entity with a vested interest in the results.

Another increasingly recognized form of reporting bias occurs when investigators measure a number of outcomes but report only those that favor the *experimental intervention* or those that favor the intervention most strongly (*selective outcome reporting bias*). If reviewers report that they have successfully contacted authors of primary studies and were assured of the full disclosure of results, concern about reporting bias decreases.

Reviewers may go even farther than simply contacting the authors of primary studies. They may recruit these investigators as collaborators in their review, and in the process, they may obtain individual patient records. Such *individual patient data meta-analysis* can facilitate powerful analyses (addressing issues such as true *intention-to-treat* analyses and informed *subgroup analyses*), which may strengthen the inferences from a systematic review.

Was the Risk of Bias of the Primary Studies Assessed?

Even if a systematic review includes only RCTs, knowing the extent to which each individual trial used safeguards against bias is important. Differences in study methods might explain important differences among the results.[12] For example, less rigorous studies sometimes overestimate the effectiveness of therapeutic and preventive interventions.[13] Even if the results of different studies are consistent, determining their risk of bias is still important. Consistent results are less compelling if they come from studies with a high risk of bias than if they come from studies with a low risk of bias.

Consistent results from *observational studies* putatively addressing treatment issues also should raise concern. Clinicians may systematically select patients with a good prognosis to receive therapy, and this pattern of practice may be consistent over time and geographic setting. There are many examples of observational studies that found misleading results subsequently contradicted by large RCTs. For example, considerable preclinical and epidemiologic evidence suggested that antioxidant vitamins reduced the risk of prostate cancer. However, a trial of 35 533 healthy men found that dietary supplementation with vitamin E significantly increased the risk of prostate cancer.[14] Similarly, laboratory experiments suggested that antioxidants may slow or prevent atherosclerotic plaque formation, but a trial of 14 641 male physicians found that neither vitamin E nor vitamin C supplementation reduced the risk of major cardiovascular events.[15] Many other examples and a discussion of misleading results of observational studies and RCTs can be found in Chapter 11.2, Surprising Results of Randomized Trials.

There is no one correct way to assess the risk of bias.[16] Some reviewers use long checklists to evaluate risk of bias, whereas others focus on 3 or 4 key aspects of the study. When considering whether to trust the results of a review, check to see whether the authors examined criteria similar to those we have presented in other chapters of this book (see Chapter 7, Therapy [Randomized Trials]; Chapter 14, Harm [Observational Studies]; Chapter 18, Diagnostic Tests; and Chapter 20, Prognosis). Reviewers should apply these criteria with a relatively low threshold (such as restricting eligibility to RCTs) in selecting studies (Box 22-3) and more comprehensively (such as considering *concealment*, *blinding*, and *stopping early* for benefit) in assessing the risk of bias of the included studies. The authors of systematic reviews should explicitly report the extent of the risk of bias of each included study in their review.

Did the Review Address Possible Explanations of Between-Study Differences in Results?

Studies included in a systematic review are unlikely to show identical results. Whether or not their review includes a meta-analysis, systematic review authors should attempt to explain the reasons for variability in results. When the studies are combined in a meta-analysis, the difference in results becomes easily quantifiable. Chance always represents a possible explanation. Alternatively, differences in the characteristics of the patients enrolled, in the way the intervention was administered, in the way the outcome was assessed, or in the risk of bias may be responsible. For example, the intervention may be more effective in older patients than in younger patients or in those with diabetes than in those without diabetes. We often refer to inconsistency in results among studies as heterogeneity.

Systematic review authors should hypothesize possible explanations for heterogeneity (a priori, when they plan the review) and test their hypotheses in a subgroup analysis. Subgroup analyses may provide important insights, but they also may be misleading (see Chapter 25.2, How to Use a Subgroup Analysis). In Chapter 23, Understanding and Applying the Results of a Systematic Review and Meta-analysis, we discuss how to evaluate heterogeneity and how it affects the confidence in estimates.

Did the Review Present Results That Are Ready for Clinical Application?

If you and your patients are told that treatment lowers the risk of myocardial infarction by 50%, it sounds impressive, but that could mean a reduction from 1% to 0.5% or from 40% to 20%. In the former situation, when the *risk difference* (also referred to as *absolute risk reduction*) is 0.5%, your patient may decide to decline a treatment with appreciable adverse effect, *burden*, or cost. In the latter situation, that is much less likely to be the case. Therefore, you and your patients need to know the absolute effect of the intervention. The absolute benefit (or harm) that patients will achieve with therapy depends on their *baseline risk* (the likelihood of the outcome when receiving no or standard therapy).

For example, statins reduce fatal and nonfatal cardiovascular events[17] by approximately 25% (*relative risk* [RR], 0.75); the absolute benefit, however, may be greater for a patient with an elevated Framingham risk score (or other risk stratification method) than for a patient with a low score (Box 22-4).

BOX 22-4

The Impact of Baseline Risk on the Magnitude of Absolute Risk Reduction

Patient 1	Patient 2
65-year-old male smoker with cholesterol level of 250 mg/dL, high-density lipoprotein (HDL) of 30 mg/dL, and systolic blood pressure of 140 mm Hg	50-year-old female smoker with cholesterol of 170 mg/dL, HDL of 55 mg/dL, and systolic blood pressure of 130 mm Hg
Absolute risk of having a cardiac event during the next 10 years: 28%	Absolute risk of having a cardiac event during the next 10 years: 2%
Risk after treatment with statin: 28% × 0.75 = 21%	Risk after treatment with statin: 2% × 0.75 = 1.5%
Absolute risk reduction: 28% − 21% = 7%	Absolute risk reduction: 2% − 1.5% = 0.5%

Although we are primarily interested in absolute effects, relative effects tend to be much more consistent across studies (see Chapter 9, Does Treatment Lower Risk? Understanding the Results). That is the reason that meta-analyses of binary outcomes usually should and do combine and present relative effects, such as the relative risk, *odds ratio*, or occasionally *hazard ratio*. So how, then, do we determine the absolute effects in which we are really interested? The best way is to obtain an estimate of the patients' baseline risk (ideally from an observational study of a representative population, from a risk-stratification instrument, or, if neither is available, from the randomized trials in the meta-analysis)[18] and then use the relative risk[19] to estimate that patient's risk difference.

Review authors also can present outcomes that are *continuous variables* in ways that are more or less useful and applicable. For instance, the weighted mean difference and standardized mean difference represent common statistical approaches for pooling across studies. Clinicians, however, may have difficulty grasping the significance of the effect of a respiratory rehabilitation program presented as a weighted mean difference of 0.71 units on the Chronic Respiratory Questionnaire (CRQ) scale. They may have less difficulty if told that the *minimal important difference* on the CRQ is 0.5 units. Clinicians are likely to have at least equal difficulty if told that the treatment effect on disease-specific *health-related quality of life* is a standardized mean difference of 0.71. Again, they may have less difficulty if told that 0.2, 0.5, and 0.8 may represent small, moderate, and large effects. Clinicians are likely to have the least amount of difficulty if told that 30% of patients have an important improvement in function as a result of the program (a *number needed to treat* of approximately 3).[20]

Were Selection and Assessments of Studies Reproducible?

As we have seen, authors of systematic reviews must decide which studies to include, the extent of risk of bias, and what data to abstract. These decisions always require judgment by the reviewers and are subject to both mistakes (ie, *random errors*) and bias (ie, *systematic errors*). Having 2 or more people participate in each decision guards against errors, and if there is good agreement beyond chance among the reviewers, the clinician can have more confidence in the results of the systematic review. Systematic reviewers often report a measure of agreement (eg, a measure of *chance-corrected agreement* such as the *? statistic*) (see Chapter 19.3, Measuring Agreement Beyond Chance) to quantify their level of agreement on study selection and appraisal of the risk of bias.

Did the Review Address Confidence in Effect Estimates?

As we have pointed out, a review can follow optimal systematic review and meta-analytic methods, and the evidence may still warrant low confidence in estimates of effect. Ideally, systematic review authors will explicitly address the risk of bias that can diminish confidence in estimates as well as imprecision (ie, wide CIs) and inconsistency (ie, large variability in results from study to study). If systematic review authors do not make explicit assessments themselves, they should at least provide the information you need to make your own assessment. The next chapter (Chapter 23, Understanding and Applying the Results of a Systematic Review and Meta-analysis) describes in detail how the systematic review authors—or you, in the absence of the authors doing so explicitly—can address these issues to make an appropriate rating of confidence in estimates of effect.

CLINICAL SCENARIO RESOLUTION

Returning to our opening scenario, the systematic review and meta-analysis you located included 11 trials that enrolled more than 10 000 patients who were having noncardiac surgery and were randomized to either β-blockers or a *control group*.[2] The trials addressed the main outcomes of interest (death, nonfatal myocardial infarction, and nonfatal stroke). The β-blocker, dose, timing, and duration of administration all varied across the trials.

The systematic review authors had searched MEDLINE, EMBASE, CINAHL, the Cochrane Library Central Register of Randomised Controlled Trials, and other trial databases and registries. They also checked the reference lists of identified articles and previous systematic reviews for additional references. They did not restrict the search to a particular language or location. They had 2 independent reviewers assess trial eligibility and select studies, and disagreements were resolved by a third review author. They did not quantitatively report the agreement level among reviewers, a feature you would have preferred to know.

The systematic review authors used the *Cochrane Collaboration* risk of bias assessment methods. They explicitly described the risk of bias of each trial by reporting on the adequacy of generation of the allocation sequence, allocation concealment, and blinding of participants, personnel, and outcome assessors. As part of the meta-analysis, the authors conducted a separate *sensitivity analysis* that excluded the studies with a higher risk of bias. They tested for publication bias. They did not report that they had contacted authors of primary studies, which you would have preferred they did.

Overall, you conclude that the credibility of the methods of this systematic review and me-ta-analysis is moderate to high, and you decide to examine the estimates of effect and the associated confidence in these estimates.

References

1. Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet*. 2008;372(9654):1962-–1976. [PubMed: 19012955]

2. Bouri S, Shun-Shin MJ, Cole GD, Mayet J, Francis DP. Meta-analysis of secure randomised controlled trials of β-blockade to prevent perioperative death in non-cardiac surgery. *Heart*. 2014;100(6):456-–464. [PubMed: 23904357]

3. Alkin M. *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage Publications Inc; 2004.

4. Oxman AD. Checklists for review articles. *BMJ*. 1994;309(6955):648-–651. [PubMed: 8086990]

5. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120(8):667-–676. [PubMed: 8135452]

6. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci*. 1993;703:125-–134. [PubMed: 8192290]

7. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. http://handbook.cochrane.org.ezproxy.library.dal.ca/. Accessed July 26, 2014.

8. MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol*. 2003;56(1):44-–51. [PubMed: 12589869]

9. McDonagh MS, Peterson K, Balshem H, Helfand M. US Food and Drug Administration documents can provide unpublished evidence relevant to systematic reviews. *J Clin Epidemiol*. 2013;66(10):1071-–1081. [PubMed: 23856190]

10. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*. 1997;315(7109):640-–645. [PubMed: 9310565]

11. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA*. 1998;279(4):281-–286. [PubMed: 9450711]

12. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609-–613. [PubMed: 9746022]

13. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. 2011;(4):MR000012.

14. Klein EA, Thompson IM Sr, Tangen CM, et al. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA*. 2011;306(14):1549-–1556. [PubMed: 21990298]

15. Sesso HD, Buring JE, Christen WG, et al. Vitamins E and C in the prevention of cardiovascular disease in men: the Physicians' Health Study II randomized controlled trial. *JAMA*. 2008;300(18):2123-–2133. [PubMed: 18997197]

16. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282(11):1054-–1060. [PubMed: 10493204]

17. Taylor F, Huffman MD, Macedo AF, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013;1:CD004816.

18. Guyatt GH, Eikelboom JW, Gould MK, et al; American College of Chest Physicians. Approach to outcome measurement in the prevention of thrombosis in surgical and medical patients: *Antithrombotic Therapy and Prevention of Thrombosis*, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2)(suppl):e185S-–e194S.

19. Murad MH, Montori VM, Walter SD, Guyatt GH. Estimating risk difference from relative association measures in meta-analysis can infrequently pose interpretational challenges. *J Clin Epidemiol*. 2009;62(8):865-–867. [PubMed: 19230610]

20. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. *Res Synth Methods*. 2012;2(3):188-–203.