

Systematic reviews and meta-analyses in regional anesthesia and pain medicine (Part I): guidelines for preparing the review protocol

Michael J Barrington,¹ Ryan S D'Souza,² Edward J Mascha,³ Samer Narouze,⁴ George A Kelley⁵

¹Department of Anesthesia and Perioperative Pain Medicine, Oregon Health & Sciences University, Portland, Oregon, USA

²Department of Anesthesiology and Perioperative Medicine, Mayo Clinic Hospital, Rochester, Minnesota

³Departments of Quantitative Health Sciences and Outcomes Research, Cleveland Clinic, Cleveland, Ohio, USA

⁴Center for Pain Medicine, Western Reserve Hospital, Cuyahoga Falls, Ohio, USA

⁵Department of Epidemiology and Biostatistics, West Virginia University, Morgantown, West Virginia, USA

Correspondence to

Michael J Barrington, Department of Anesthesia and Perioperative Pain Medicine, Oregon Health & Sciences University, Portland, OR 97239, USA; barringm@ohsu.edu

This article has been copublished with permission in *Regional Anesthesia & Pain Medicine* and *Anesthesia & Analgesia*. All rights reserved in respect of *Regional Anesthesia & Pain Medicine*, © American Society of Regional Anesthesia & Pain Medicine 2023 and in respect of *Anesthesia & Analgesia*, © International Anesthesia Research Society. The articles are identical except for minor stylistic and spelling differences in keeping with each journal's style. Either citation can be used when citing this article.

Received 26 June 2023

Accepted 26 June 2023

Published Online First

9 November 2023



► <http://dx.doi.org/10.1136/rapm-2023-104802>



© American Society of Regional Anesthesia & Pain Medicine 2024. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Barrington MJ, D'Souza RS, Mascha EJ, et al. *Reg Anesth Pain Med* 2024;**49**:391–402.

ABSTRACT

Comprehensive resources exist on how to plan a systematic review and meta-analysis. The objective of this article is to provide guidance to authors preparing their systematic review protocol in the fields of regional anesthesia and pain medicine. The focus is on systematic reviews of healthcare interventions, with or without an aggregate data meta-analysis. We describe and discuss elements of the systematic review methodology that review authors should prespecify, plan, and document in their protocol before commencing the review. Importantly, authors should explain their rationale for planning their systematic review and describe the PICO framework—participants (P), interventions (I), comparators (C), outcomes (O)—and related elements central to constructing their clinical question, framing an informative review title, determining the scope of the review, designing the search strategy, specifying the eligibility criteria, and identifying potential sources of heterogeneity. We highlight the importance of authors defining and prioritizing the primary outcome, defining eligibility criteria for selecting studies, and documenting sources of information and search strategies. The review protocol should also document methods used to evaluate risk of bias, quality (certainty) of the evidence, and heterogeneity of results. Furthermore, the authors should describe their plans for managing key data elements, the statistical construct used to estimate the intervention effect, methods of evidence synthesis and meta-analysis, and conditions when meta-analysis may not be possible, including the provision of practical solutions. Authors should provide enough detail in their protocol so that the readers could conduct the study themselves.

Systematic reviews and meta-analyses of randomized clinical trials are regarded as the pinnacle of evidence-based medicine.¹ The reputed worth of this approach is deserved when the systematic review collates all evidence relevant to a prespecified, important clinical question using methods that are explicit and systematically constructed to minimize bias in the selection of studies and evidence synthesis.^{2,3} Meta-analysis refers to statistical techniques used to combine the results of multiple trials, generating a composite larger sample size and a single numerical estimate of effect contrasting the outcomes of 2 groups receiving different therapies. In this article, we present guidelines for both the systematic review combined with meta-analysis of effect estimates (the usual scenario) and the

WHAT IS ALREADY KNOWN ON THIS TOPIC?

⇒ Comprehensive resources exist on how to prepare a protocol in advance of performing a systematic review and meta-analysis. Systematic review protocols in regional anesthesia and pain medicine often lack foundational content within those resources.

WHAT DOES THIS GUIDELINE ADD?

⇒ This guideline identifies key content within existing resources to support authors preparing their systematic review protocols.

HOW MIGHT THIS GUIDELINE AFFECT RESEARCH, PRACTICE, OR POLICY?

⇒ This guideline may assist protocol developers as well as clinician peer reviewers tasked with reviewing and improving the quality of systematic review manuscripts submitted to the journals of *Regional Anesthesia and Pain Medicine* and the Regional Anesthesia and Acute Pain Medicine section of *Anesthesia & Analgesia*.

systematic review performed with no subsequent meta-analysis. Throughout, the primary term used will be systematic review.

Comprehensive resources on this topic include the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) initiative,^{2,4} and the *Cochrane Handbook for Systematic Reviews of Interventions*.⁵ Despite clearly formulated, promulgated, and accepted guidelines, and the commonly included author statement of compliance with PRISMA, content foundational to preparing and planning a systematic review is frequently absent in review protocols and the methods section of manuscripts submitted to *Regional Anesthesia and Pain Medicine* and the Regional Anesthesia and Acute Pain Medicine section of *Anesthesia & Analgesia*. Poor compliance with PRISMA is also evident in other fields.⁶ Moreover, the content within systematic review manuscripts frequently deviates substantially from what the authors documented in their protocol. An analysis of 17 systematic reviews investigating interventions for chronic postsurgical pain reinforces these concerns.⁷

Authors are submitting systematic reviews and meta-analyses with increasing frequency to medical

journals; therefore, addressing omissions in scientific content and format considered fundamental to a systematic review is needed. Our target readers are¹ the content expert clinician who is considering conducting a systematic review in the fields of regional anesthesia and pain medicine and² clinician peer reviewers tasked with reviewing systematic review and meta-analysis manuscripts. One intent is to clarify expectations required of systematic review authors. Importantly, medical journals require authors to follow the 2015 PRISMA Protocols (PRISMA-P) Explanation and Elaboration paper (or subsequent versions) when drafting their review protocol. Uploading the PRISMA-P checklist on submission and before the manuscript is sent for peer review is a journal requirement.²

In this first article, Part I, we describe and discuss elements of the systematic review methodology that should be prespecified, planned, and documented in the protocol before commencing the formal review process. These elements include (1) developing the correct framework; (2) explaining the rationale for the review and framing the healthcare question; (3) defining and prioritizing primary and other outcomes; (4) specifying eligibility criteria; (5) documenting information sources and developing search strategies; (6) describing plans to manage data; (7) risk of bias assessment; (8) planned methods of evidence synthesis; and (9) methods used to evaluate heterogeneity, small study effects (nonreporting bias), and quality (certainty) of the evidence. For Part I, we draw heavily on the PRISMA initiative^{2,4} and the *Cochrane Handbook*.⁵ In addition, authors should refer to a critical appraisal instrument for assessing systematic reviews, for example, A Measurement Tool to Assess Systematic Reviews (AMSTAR 2).⁸ In the companion article, Part II, we describe the formal review process. While the focus of these guidelines is on a traditional aggregate data systematic review with meta-analysis of randomized clinical trials, much of the presented information also applies to network⁹ and individual participant data¹⁰ meta-analyses, for both of which specific PRISMA extension statements exist.

THE PICO FRAMEWORK IS CENTRAL TO PLANNING A SYSTEMATIC REVIEW AND META-ANALYSIS

PICO is a framework for operationalizing a clinical research question. The PICO framework comprises participants (P), interventions (I), comparators (C), outcomes (O), and related elements (study design, practice settings, timeframe, and length of follow-up).^{2,11} These related elements can also be incorporated into the acronym. For example, the T in PICOT refers to time, and the S in PICOS to study design. PICO and related elements (henceforth referred to as PICO or PICO items) inform the reader of key content. For example, a review title that includes PICO elements informs the reader. We recommend that authors use PICO to frame their important clinical question (PRISMA-P Item 7)² and write this in the final paragraph of the introduction to make the reader aware of this without having to read further. PICO also provides a framework for determining the scope of the review, designing and executing the search strategy, specifying the eligibility criteria (PRISMA-P Item 8), and identifying potential sources of heterogeneity (PRISMA-P Item 15 a).² The PICO framework also informs data extraction (PRISMA-P Item 12).²

PICO items have variants documented a priori by content experts in their review protocol. For example, the population of patients having breast surgery for cancer (P) comprises multiple surgical subtypes that will likely modify the treatment effect of regional anesthesia techniques (I). This diversity in the patient population may influence the review authors' decision to proceed

with meta-analysis (PRISMA-P Item 15 a).² Authors may decide to limit the scope of their review by excluding subpopulations or maintain breadth by addressing multiple subpopulations.^{11,12} Clinical experts will also be aware of regional anesthesia intervention (I) characteristics (eg, versions and evolution) that potentially modify its effectiveness. Authors will next want to consider the comparator interventions (C), including inactive control interventions (eg, no intervention, placebo, sham procedure, usual care), active control interventions (eg, a different intervention, a variant of the same intervention), and cointerventions.¹² In their protocol, authors should plan how specified intervention groups will be used in their planned evidence synthesis and reporting. They may decide to build contingencies by specifying both specific and broader intervention groups.¹² This process mitigates the risk of review authors making ad hoc decisions after execution of the search strategy and selecting studies.

Overall summary: PICO should relate directly to the question asked. In the protocol, authors should document variants of the PICO items: specifically, population characteristics that may influence the intervention treatment effect, and characteristics of the intervention that may modify its effect.

IMPORTANCE OF PREDEFINED PROTOCOL AND REGISTRATION

While some have raised opposition,¹³ guidelines mandate that systematic review authors document their methods within the protocol in advance of conducting the systematic review.² Performing a systematic review and meta-analysis requires multiple decisions and judgments. As authors write their review protocol, they establish the context for and scope of the review, develop key priorities, and develop a systematic approach that aims to minimize bias in the review findings. The protocol is contractual in nature and describes what the authors will do during the review process. Clinical content experts will clearly need to engage stakeholders with the breadth of expertise and perspective required for the design and documentation of the protocol for their planned systematic review.² These would include systematic review methodologists, information specialists, and end consumers.

Medical journals expect that authors will use the 2015 PRISMA-P Explanation and Elaboration paper (or subsequent versions) when drafting their review protocol.² Such journals mandate that authors place their protocol in a publicly available registry. This promotes accountability and protects against arbitrary methodological changes and selective outcome reporting. Describing and explaining amendments is a 2020 PRISMA reporting requirement.⁴ Put simply, authors document what they intended to do vs what they did, including the reason for any change. The review protocol with enclosed methods is itself an important, stand-alone document. In the following sections, we describe methodological content (in addition to the PICO framework) to be specified in the review protocol in advance of performing the systematic review.

Overall summary: The protocol for the systematic review and meta-analysis contains the planned review methods, outcomes, and analyses. The protocol indicates the existence of a plan for the review process. The review methods should be structured, transparent, and reproducible.

DESCRIBE THE RATIONALE FOR THE REVIEW AND FRAME THE IMPORTANT HEALTHCARE QUESTION

The systematic review begins with authors describing their rationale and objectives for performing the review (PRISMA-P Item 6).² Listing review objectives allows authors to state

their question more broadly. In their review protocol, authors describe how an intervention in a specified population will likely produce the expected outcome. Framing the important healthcare question that authors seek to answer is the critical item that helps authors maintain focus and determine the scope of their review. If there is no important clinical question, there is no basis for the systematic review. The clinical question guides many aspects of the review process, including but not limited to eligibility criteria, search strategy, study selection, and data extraction. The review question drives the development of the methods. As an example, many systematic reviews compare two interventions, and hence a pairwise meta-analysis is used. However, when multiple interventions for one condition are evaluated, then a network meta-analysis is appropriate.⁹

Extensive methodological expertise and resources are required to complete a systematic review and meta-analysis. Therefore, prospective authors should ask themselves the following: What makes this systematic review important and interesting? Will the results of our review move the field forward? Will this review address answerable questions and bridge important gaps in our knowledge?¹¹ Authors search the existing literature and registers of systematic reviews to avoid repeating a similar review. However, an existing systematic review may need updating as new evidence emerges. As an example, authors justified the need to update a meta-analysis to address the role of duloxetine in ameliorating knee stiffness (a previous review addressed knee osteoarthritis pain).¹⁴ Updating a meta-analysis is also appropriate if previous reviews did not have sufficient evidence to definitively answer the clinical question. Authors considering a systematic review may want to use the decision tree algorithm from the Panel for Updating Guidance for Systematic Reviews (PUGs) for help in deciding whether a new or updated review is needed.¹⁵

We recommend again that authors frame their important healthcare question (broad or narrow) using PICO items and related elements (PRISMA-P Item 7).^{2,11} Consider three systematic reviews, where the population was total knee arthroplasty; however, they varied in scope based on the clinical question and PICO: (1) to assess if any regional anesthesia blockade vs none improved a broad range of clinical outcomes¹⁶; (2) to assess if femoral nerve blockade alone or in combination with other blocks vs no femoral nerve blockade improved pain and adverse events¹⁷; and (3) to assess if femoral nerve blockade with sciatic nerve blockade vs femoral nerve blockade alone improved analgesic outcomes.¹⁸ These three systematic reviews have questions with a narrowing focus achieved through PICO selection. Authors should decide the scope of their review by deciding if generalizing across PICO elements is appropriate and whether extracted information would provide clinically relevant information.

Overall summary: Reviewers and readers should expect to see the clinical question framed in the introduction using PICO and related elements (eg, timeframe, study type).

LIST, DEFINE, AND PRIORITIZE PRIMARY AND OTHER OUTCOMES

The primary outcome extends from the rationale for and objectives of the review. The primary outcome is chosen so that the clinical question can be answered. We recommend authors specify one or two primary outcomes, together with a statistical analysis plan for these outcomes. Secondary outcomes generally relate to the primary outcome and are consistent with the systematic review objectives.¹⁹ Authors in their systematic

review protocol should list, define, and prioritize outcomes as primary or secondary¹¹ and provide a rationale for their choices (PRISMA-P Item 13).² This process mitigates the risk of selecting and reporting outcomes once results of the review are known.

While the *Cochrane Handbook* recommends evaluating harms, systematic reviews in our field typically comprise efficacy trials, whose authors do not routinely explain in their methods how they captured adverse outcomes. Nevertheless, data for this outcome should be extracted, and in the absence of adequate data on harms, quantitative or otherwise, a recommendation is made in the discussion section of the manuscript that future trials include more detailed information on harms. Providing direction for future research is an important but often overlooked aspect of systematic reviews.

Authors should describe how they will manage multiplicity of outcomes and analyses.¹² This issue is mitigated by limiting the number of reported outcomes.¹⁹ Fully anticipating the characteristics (eg, diversity of reported outcomes) of eligible studies in advance is not possible. However, if the review objective is to evaluate quality of analgesia, then authors should anticipate reporting of pain from the primary studies at different levels of stimulation and at multiple time points (*Cochrane Handbook* section 6.2.4, Repeated observations on participants).²⁰ Repeated observations (eg, measurement of pain) on the same participants at different time points produces data that are considered statistically dependent and increases the probability of reporting false positive results. Managing this multiplicity is challenging.²¹ Review authors should consider prespecifying a hierarchy of measures as described in *Cochrane Handbook* section 3.2.4.¹² For example, if the Brief Pain Inventory was reported, then a hierarchy could be prespecified: the pain subscale from the Brief Pain Inventory, then pain scores with movement, then one time point selected that would be considered by stakeholders (patients, caregivers) important.¹² In the protocol development phase, authors may decide to draft a ‘concept’ relating to the research question and refine this when they conduct the review.^{11,12} Plans to refine outcome definitions (eg, timeframe or time point) should be described in the review protocol² and reported with transparency in the final manuscript.⁴ For example, the decision to select the time point most frequently reported vs time elapsed should be transparent. Authors should state in the protocol the minimal clinically important difference for their primary outcome. For example, in a meta-analysis of analgesic benefits of erector spinae plane block after breast surgery, Hussain *et al*²² a priori defined a reduction of 30 mg oral morphine in the first 24 hours postoperatively to be clinically important.

Overall summary: Authors should select their primary outcome with a relevant and important time point, if appropriate, motivated by the rationale, objectives, and clinical question they are asking. Review authors should anticipate challenges associated with outcome multiplicity and analyses.

DETAIL THE ELIGIBILITY CRITERIA FOR SELECTING STUDIES, INTENDED SOURCES OF INFORMATION, AND SEARCH STRATEGIES

The quality of the database search strategy comprises a core element of the systematic review search plan. We recommend authors adhere to the Peer Review of Electronic Search Strategies (PRESS) standard, including having the search strategy peer-reviewed by a second information specialist unaffiliated with the systematic review.²³ The most important aspect of a search strategy is its sensitivity, otherwise known as recall.

The goal is to capture as many studies as possible that meet the eligibility criteria, unrestricted by language and publication status. The goal is to reduce bias in identifying and selecting source reports, thus improving reproducibility. The review question and objectives drive the eligibility criteria. In their review protocol, authors should explicitly define their eligibility criteria (inclusion and exclusion criteria) using study characteristics (PICO items).

Authors should document their decisions on report characteristics (years covered, language, publication status) (PRISMA-P Items 8–9).^{2 12} For non-English-language articles, free online translators exist; however, professional language translation services may be required.²⁴ It is recommended that information be sought from sources that are unpublished, noting that these data results may not have been peer-reviewed. These include industry/manufacture data, trial registries, clinical study reports, regulatory data, the reference lists of source reports, conference proceedings, and the gray literature. For nonreporting biases, refer to Chapter 7, section 7.2.3,²⁵ and Chapter 13 of the *Cochrane Handbook*.²⁶

Explicitly defining the eligibility criteria (1) reduces the risk of bias in identifying and selecting studies and (2) drives the search strategy terminology. The search strategy follows from the research question, eligibility criteria, and PICO items. During protocol development, clinical content experts will need to collaborate with an information specialist to develop robust searching methods. This specialist would likely be a health sciences librarian with expertise in systematic review searching.²³ The information specialist or librarian will need to know the following information: registered protocol, anticipated review title, whether this is an update of an existing review, PICO elements, search terms prespecified by authors with clinical expertise, and examples of citations identified from preliminary searches that should be captured in the search strategy. At least three electronic databases should be searched for potentially eligible studies—for example, PubMed, which MEDLINE is nested in, Embase, and the Cochrane Central Register of Controlled Clinical Trials. Despite Google Scholar being recommended as a primary database,²⁷ it is important to understand that it does not allow for highly sensitive (refined) searching, thereby leading to many false positives and potentially wasted effort. In their protocol, authors should include a draft of the search strategy querying one database, with a description of the planned approach for other databases (PRISMA 20 Item 7)² that is mature enough so that authors considering reviewing the same topic would avoid duplication of effort. An example of a draft search strategy is shown in table 1.

Overall summary: The most important aspect of a search strategy is its sensitivity. The goal is to capture as many studies as possible that meet the eligibility criteria, unrestricted by language and publication status.

DESCRIBE AND PLAN HOW KEY DATA ELEMENTS WILL BE MANAGED, SELECTED, AND EXTRACTED

In their review protocol, authors need to describe how they will manage records and data throughout the systematic review (PRISMA-P Item 11a).² Authors should describe their process for selecting studies through each phase of the review (eligibility, screening, and inclusion) (PRISMA-P Item 11b), and how they will extract key data elements from reports (PRISMA-P Item 11c). Free online programs are available for title and abstract study screening.²⁸ Key data elements to be extracted, such as

Table 1 Details of search strategy required in the review protocol

No. of steps	Query*
1	Radiculopathy/
2	(radiculopath\$ or radiculitis or radiculitides or (nerve root adj1 (disorder\$ or inflammation\$ or avulsion\$ or compress\$)) or radicular-pain).ab,kf,ti.
3	1 or 2
4	Blood platelets/
5	(platelet\$ or thrombocyte\$).ab,kf,ti.
6	4 or 5
7	Injections, epidural/
8	((epidural or extradural or peridural or transforaminal) adj inject\$).ab,kf,ti.
9	7 or 8
10	3 and 6 and 9

*Database: MEDLINE[A1]; Filter: Ovid MEDLINE(R) 1946 to Present and Epub Ahead of Print, In-Process & Other NonIndexed Citations and Ovid MEDLINE(R) Daily. This search strategy is applicable to the MEDLINE database, a similar search strategy with the same search terms and Boolean exclusion operators will be used to query EMBASE, Web of Science, Scopus, etc, from database inception to study start date. Example of a draft search strategy querying the MEDLINE database is displayed. The topic of the systematic review is pain intensity in patients with radiculopathy who receive epidural administration of platelet rich plasma versus placebo. This is the level of detail of a draft search strategy required in the protocol.

PICO items, including prespecified time points, need to be defined (PRISMA-P Item 12).² Chapter 5 of the *Cochrane Handbook* contains a relevant checklist of items (table 5.3.a, section 5.5.11).²⁹ The authors of Chapter 5 recommend that the review team construct an outline of the tables and figures to be included in the review.²⁹ Develop these outlines so that the required data are collected to populate the table of characteristics of included studies and to facilitate risk of bias assessment. Review authors should also consider developing outlines of the evidence profile³⁰ and summary of findings tables.^{30 31} The majority of information required for the latter can be collected before the formal review process.

An example of the detail required is reported in a meta-analysis evaluating the efficacy of adductor canal block for knee arthroplasty.³² Tables 1–3 describe the characteristics of included studies (participants, outcomes, and intervention/cointervention respectively). Tables 4 and 5 report the evidence profile and summary of findings.³² Authors should plan for data assumptions and simplification (PRISMA-P Item 12).² As an example, a meta-analysis appraising evidence for intraoperative methadone defined postoperative oral morphine equivalents as their primary outcome.³³ For studies that reported opioid dosage, the authors manually converted opioid dosage to morphine equivalents.

Overall summary: Authors should describe their process for selecting studies. They should identify key data elements they plan to extract. To facilitate this, we recommend author's construct (during development of the protocol) an outline of the tables and figures they plan to include in the review.

DESCRIBE HOW RISK OF BIAS FOR OUTCOME VARIABLES WILL BE ASSESSED IN INDIVIDUAL STUDIES AND ACROSS STUDIES

Inherent to assessing the internal validity of a study is the risk of bias that it systematically overestimates or underestimates the true intervention effect on a specified outcome variable compared with the result derived from a 'perfect' trial. This is distinct from quality-compromising aspects such as inadequate

Table 2 Worked example of how model choice for a meta-analysis can affect the overall pooled results and subsequent conclusions

Model	Difference in means (95% CI)	P value
FE	-0.55 (-0.70 to -0.40)	<0.001
RE _{D&L}	-0.90 (-1.36 to -0.44)	<0.001
RE _{HKSJ}	-0.90 (-1.45 to -0.35)	0.003
IVhet	-0.55 (-1.27 to 0.17)	0.13
QE	-0.38 (-1.12 to 0.36)	0.31

Data are from the work of Park *et al*⁶⁰ on preemptive epidural analgesia for acute and chronic postthoracotomy pain in adults. Negative values indicate improvement, that is, reductions, in pain intensity in the preemptive vs control group. p, α value; Q (p), Cochran Q statistic and α value for Q; QE, quality effects model with studies weighted by Cochran Risk of Bias Results reported in the original meta-analysis; RE_{D&L}, traditional random effects model of DerSimonian and Laird; RE_{HKSJ}, random effects model with Hartung-Knapp-Sidik-Jonkman adjustment. Results are similar for Q (p) [(116.3 (<0.001)), I^2 (95% CI) [85.4% (78.3–90.2)]] and given that the same approach is historically used for analysis across all models. FE, fixed effects model; IVhet, inverse variance heterogeneity model.

sample size. At least 393 assessment tools are available for assessing the methodological quality and risk of bias for various study designs.³⁴ However, medical journals require that authors of systematic reviews use the most recent Cochrane Risk of Bias assessment instrument, currently RoB2,³⁵ when assessing risk of bias for parallel group randomized controlled trials in which participants are assigned at the participant level. Using signaling questions, the RoB2 instrument assesses bias as ‘low risk,’ ‘high risk,’ or ‘some concerns’ across 5 domains: (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcome, and (5) selection of the reported result. Based on the results of these domain assessments, an overall risk of bias of either ‘low risk,’ ‘high risk,’ or ‘some concerns’ is generated.³⁵ While a decision tree helps to guide both domain and overall bias assessments, these may be overridden by the investigators based on their own judgments. Specific Cochrane tools are also available for investigators who plan to include randomized crossover trials, cluster-randomized, or other trial designs.

Medical journals require that authors, at the protocol stage, describe methods used to assess risk of bias and whether this is at the outcome level, the study level, or both. Both levels are most common and preferred—bias is assessed for a particular outcome variable at the study level and then results aggregated across studies (see PRISMA-P Item 14).² Specific methodological tasks include the following:

1. Describe the process for extracting outcome data elements. Medical journals require that data elements used to evaluate the risk of bias should be extracted using two people working independently and a process defined on how to resolve disagreement.
2. Define the criteria used for ROB 2 categories for each bias domain. For example, if the methods used to generate the randomization sequence or concealment allocation are not described in a source study, then consider documenting a priori what risk this represents. For example, in their meta-analysis, Hussain *et al* decided a priori that studies that did not use a sham block were assigned a high risk of detection bias.²²
3. Describe information that will be provided to readers to support the risk of bias judgments and if all outcomes included in the summary of findings tables^{30,31} will be assessed for risk of bias (considered mandatory for Cochrane reviews, Handbook Boxes 7.3.b).^{25,36}

4. Describe how the risk of bias assessments will be used in data synthesis (sensitivity analyses).² We recommend authors plan to test how sensitive the estimate of treatment effect is to evidence of risk of bias.³⁷

DESCRIBE THE STATISTIC USED TO ESTIMATE THE INTERVENTION EFFECT, METHODS OF DATA SYNTHESIS, AND META-ANALYSIS

The effect measure is the statistic that compares the outcome data between two study groups.²⁰ For dichotomous (binary) data, effect measures include risk ratio, relative risk, OR, risk difference, absolute risk reduction, attributable risk, and number needed to treat. For continuous data, effect measures such as mean difference or standardized mean difference are often used.^{20,37,38} Numerical rating pain scales are ordinal scales but are typically treated as continuous data. For time-to-event data, hazard ratios are commonly used. Refer to [table 2](#) in the accompanying Part II paper for more information on effect measures. Effect measures are either ratio (eg, OR) or difference (eg, mean difference) measures, which by comparing outcome data between two groups describe the magnitude of the intervention effect. The true effect of an intervention is not known and can only be estimated, hence the use of the terms ‘treatment effect estimate,’ ‘intervention effect estimate,’ or ‘effect estimate.’ The term ‘effect size’ may be used interchangeably with ‘effect estimate’; however, of note, the former term is also used to describe versions of standard mean difference with a denominator that comprises variance. Effect estimates are reported with both a point estimate (eg, mean difference or OR) and the SE or CI.

A meta-analysis of effect estimates is possible when the aggregated effect estimates and their variances from individual studies are known or calculated. The effect estimates are then combined, generating a summary numerical estimate of the effectiveness of the therapy. However, it is recommended that a meta-analysis only occurs when there is an acceptable level of homogeneity between two or more included trials.³⁸ When the number of studies pooled is small, the ability to generalize beyond the included studies is limited; however, this could provide direction regarding the need for future research on the topic of interest.

Authors should prespecify in the protocol the effect measures (for dichotomous and continuous data) they plan to use for their analyses (PRISMA-P Item 15b)² including how these metrics will be calculated. For example, several methods exist for calculating the standardized mean difference (Hedges’ g and d, Cohen’s d, etc), and some studies do not include the exact data needed to calculate the effect size. While authors of reviews can contact the corresponding authors to retrieve this information, they may have limited success.³⁹ Therefore, using well-described methods, the authors should attempt to calculate their effect estimate from the data provided in the article vs simply discarding the study.^{20,40}

After describing how the effect estimate is calculated from individual studies, the a priori method of pooling studies needs to be described in detail. Broadly, there are two approaches or models used to combine the evidence from multiple studies: fixed-effect and random-effects. Fixed-effect models are limited to estimating within-study error and assume that there is a single true treatment effect in the population of interest. Differences in treatment effects are assumed to be sampling error only. This model is appropriate when inference is limited to the specific population (fixed-effect) or specific studies (fixed-effects) in the analysis.⁴¹ However, most investigators are interested in inferring beyond the studies included in the meta-analysis to a larger set of similar populations. To address this, random-effects

Table 3 Examples of clinical and methodological heterogeneity

Type of heterogeneity	Summary title of cited study	Key parameters highlighting heterogeneity
<i>Clinical heterogeneity</i>		
Participant characteristics (sex, age, baseline disease severity, ethnicity, comorbidities)	Meta-analysis comparing high concentration of plain local anesthetic vs low concentration of local anesthetic with opioids for labor epidural analgesia ⁵⁷	Nulliparous only vs mixed patients; geographical diversity (seven different countries) with likely variation in practice patterns
	Systematic review that pools participants having videoassisted, robotic and open surgical techniques (unpublished)	Different surgical types will have different recovery profiles
Intervention characteristics (type, dose, frequency of dose, route of delivery)	Meta-analysis comparing analgesic efficacy of liposomal bupivacaine vs bupivacaine hydrochloride for prevention of postoperative pain ⁵⁸	Drug dose differences between study arms
	Systematic review evaluating change in pain intensity in chronic noncancer pain after neuromodulation interventions (unpublished)	Type of interventional modality each with distinct mechanism (intrathecal drug delivery systems and spinal cord stimulation)
Outcome measurements (timing, length of follow-up, repeated measurements)	Systematic review evaluating efficacy and safety of magnesium for management of chronic pain ⁵⁹	Length of follow-up
<i>Methodological heterogeneity</i>		
Individual study design	Meta-analysis comparing efficacy of topical nonsteroidal anti-inflammatory agents vs placebo for osteoarthritis ⁶⁰	Pooling of both randomized controlled trials and observational studies
Methodology/quality of trials (differences in allocation concealment, differences in participant blinding)	Systematic review of truncal regional nerve blocks comparing analgesia between intermittent local anesthetic bolus vs continuous infusion techniques ⁶¹	Variable blinding of participants and/or study personnel; variable allocation concealment

models attempt to estimate between-study variance so that one can (hypothetically) generalize to similar populations beyond those included in the meta-analysis. Contrary to a fixed-effect(s) model, a random-effects model assumes that there is a distribution of treatment effects in the population of interest. This distribution is estimated/characterized from the effects reported from

individual studies.³⁷ Several random-effects models exist,^{41–45} all of which use different methods to estimate between-study variance.

In 1986, DerSimonian and Laird observed that systematic reviews lacked a consistent assessment of homogeneity of treatment effect before pooling.⁴² The DerSimonian and Laird

Table 4 Exploring heterogeneity

Terminology	Explanation	
Clinical heterogeneity	Due to variability in participants, interventions, and outcomes	
Methodological heterogeneity	Due to variability in study design, methods of assessment, risk of bias	
Statistical heterogeneity	Variation in effect estimates more than would be expected due to chance, a consequence of clinical and methodological diversity	
Statistical test, analysis	Explanation	Limitations/explanation
Cochran's Q test	Tests the null hypothesis of homogeneity. Q is the (chi-square statistic) test reported at bottom of the forest plot. It assesses if the observed differences are likely due to chance. A low P value, a large chi-square relative to df provides statistical evidence of heterogeneity.	Low power (failure to reject null hypothesis) when the number of studies is small. P value of .10 often used as threshold of statistical significance. May detect unimportant heterogeneity with large number of studies.
Tau ² (τ ²)	Absolute measure of between-study heterogeneity.	Higher values suggest greater between-study heterogeneity. Imprecise when the number of studies is low.
I ² statistic	Derived from Cochran's Q. Relative measure of heterogeneity, proportion of variability in effect size estimates that is real (variation beyond chance) compared with that occurring by chance.	Imprecise when the number of studies is low. Both the point estimate and a measure of its precision, the 95% CI, should be reported. When all the 95% CIs of the effect size of all the studies in a meta-analysis overlap, I ² will be 0.
Subgroup/moderator analysis	Examples: Will subgroups of studies with different attributes (participant or intervention, risks of bias) have different effect estimates?	Often used to identify source of heterogeneity. Subgroups are best identified a priori.
Metaregression	Describes association between effect size and study level covariates. Ask this question: Are there factors that modify effectiveness of a treatment?	Describes variation between studies. Results are observational in nature. Limited by the number of studies compared with number of covariates. Multiple metaregression often challenging because of missing data for different covariates from different studies.
Sensitivity analysis	Examples: test how sensitive the effect estimate is to (1) deleting each study from the model once, (2) deleting studies that are outliers (results, size), and (3) deleting studies based on quality aspects such as risk of bias.	Often used to identify source of heterogeneity.
CI of summary mean treatment effect estimate	Describes the uncertainty of the summary treatment effect estimate (single numerical estimate) generated from meta-analysis.	Is a measure of precision of the mean treatment effect estimate generated from multiple studies. Does not describe degree of heterogeneity among studies.
Prediction interval	Describes the spread of underlying effects in the studies included in a random-effects meta-analysis. That is, the prediction interval describes the degree of heterogeneity in meta-analysis. Describes the width of the distribution.	Based on assumption of normality of underlying effects estimates. Is a measure of dispersion of individual results and provides a wider range of expected treatment effects compared with CI of the mean effect estimate.

Table 5 Summary

Before commencing their systematic review, authors will be aware of and adhere to the following	
Use PICO and related elements	To frame their clinical question and determine the scope of the review To design their search strategy, specify eligibility criteria, and inform planned data extraction To prespecify potential sources of clinical and methodological heterogeneity
Predefine their protocol and place in a publicly available registry	The protocol will contain their planned review methods, outcomes, and analyses, which will be structured, transparent, and reproducible The protocol will document what authors intended to do and what they actually did
Explain the rationale for the review and articulate their clinical question	Framing the important healthcare question that authors seek to answer is a critical item The clinical question will subsequently determine the scope of the review, choice of outcome and drive development of the methods The clinical question will guide many aspects of the review process including eligibility criteria, search strategy, study selection, data extraction, and evidence synthesis
List, define, and prioritize outcomes as primary and secondary	The primary outcome extends from the rationale for the review, its objectives and is chosen so that the clinical question can be answered Authors should limit the number of outcomes reported and/or acknowledge that secondary outcomes that do not directly address the review objectives should be considered exploratory Define a priori clinical importance for the primary outcome
List eligibility criteria for selecting studies and document intended sources of information and search strategies	Explicitly define eligibility criteria (inclusion and exclusion criteria) using study characteristics (PICO) and document decisions on report characteristics (such as years covered, language, publication status) At least three electronic databases should be searched for potentially eligible studies Include a draft of the search strategy querying one database with a description of the planned approach for other databases
Describe and plan how data elements will be managed, selected, and extracted	Plan to collect data that will populate tables of study characteristics, summary of findings, evidence profile, risk of bias, and that required to perform meta-analysis Describe their process for selecting studies (eligibility, screening, and inclusion)
Describe how risk of bias will be assessed	Using RoB2, the most recent Cochrane risk of bias assessment instrument for randomized controlled trials Describe information that will be provided to readers to support their risk of bias judgments
Describe effect measure, methods of data synthesis and meta-analysis	Prespecify the effect measure used to estimate the treatment effect and how it will be calculated Describe in detail methods of evidence synthesis including rationale and choice of statistical model
Describe how risk of bias due to results missing in the evidence synthesis will be evaluated	Plan for both a qualitative and quantitative assessment of small-study effects (nonreporting biases)
Describe planned methods to identify and evaluate sources of heterogeneity	Plan to identify clinical and methodological heterogeneity using preidentified clinical covariates Be explicit about methods to test for and quantify statistical heterogeneity Consider a limited number of predefined sensitivity, subgroup, moderator, and/or meta regression analyses Acknowledge that a meta-analysis with a small number of studies may give vulnerable results and/or provide limited information on sources of heterogeneity
Describe conditions when meta-analysis may not be possible and provide practical solutions	For example, <10 studies with primary outcome data, different effect measures, bias, substantial unresolved heterogeneity Consider a method of evidence synthesis that does not involve meta-analysis of effect estimates
Describe how quality (certainty) of the evidence will be evaluated	Using the GRADE process

GRADE, Grading of Recommendations Assessment, Development and Evaluation; PICO, participants (P), interventions (I), comparators (C), outcomes (O), and related elements; RoB2, Risk of Bias 2.

random-effects model is the most used and is available in most statistical packages. Random-effects models incorporate statistical heterogeneity into the overall pooled estimate but do not explain it. It is important to note that the decision of a fixed vs random-effects model should never be because of a statistical test of heterogeneity.^{38 41} In general, we suggest that a random-effects model be chosen over a fixed-effect(s) model and that the original citation for the specific random-effects model used be provided.

No statistical model is perfect, and alternatives to traditional fixed-effects and random-effects models exist. Two such models are the inverse-variance heterogeneity (IVhet) model of Doi *et al*^{46 47} and the quality effects (QE) model.^{48 49} The QE model incorporates weights based on overall risk of bias/study quality scores into the pooled result. [table 2](#) illustrates an example of how choice of model choice can affect the pooled results and conclusion. The data are based on a meta-analysis by Park *et al*⁵⁰ on preemptive epidural analgesia for acute and chronic postthoracotomy pain in adults. The primary outcome was the difference in means (preemptive minus control) pain intensity, measured

using a numerical rating scale, 4 hours after surgery. Eighteen studies representing 1003 participants (500 intervention, 503 control) were included. In the original meta-analysis, the authors used the random-effects model of DerSimonian and Laird.⁴² As can be seen by the pooled estimates as well as the 95% CIs and α values for the pooled estimates, 3 models yielded results considered to be statistically significant ($p < 0.05$), while two did not. Thus, the first three models may lead one to the conclusion that preemptive epidural analgesia reduces pain intensity, while the other two would lead to the opposite conclusion.

While all models (except the fixed-effect model) incorporate between-study heterogeneity into the overall pooled result, none of models explain the source(s) of the heterogeneity. Note that the models listed are intended to be illustrative only and not an exhaustive list of models available for a meta-analysis of effect estimates. Review authors should consider the IVhet and QE models due to their better performance compared with the traditional random-effects models. Authors should provide their rationale for the choice of model.

Authors need to report whether their statistical tests are based on the Z- or t-distribution; are 1- or 2-tailed; and the α value and subsequent CI level chosen (eg, 95%). It is recommended that the Hartung-Knapp-Sidik-Jonkman adjustment be applied to a random-effects meta-analysis.⁵¹ This adjustment modifies the SE of the point estimate as well as multiplying the SE based on the t- vs Z-distribution—the end result being a wider and more accurate CI. The rationale for this adjustment is that CIs based on the Z-distribution tend to be inappropriately too narrow. This adjustment is available in various meta-analytic statistical routines.

It is recommended that a 95% prediction interval be calculated if a random-effects, IVhet, or QE model is chosen.⁵² A traditional 95% CI gives lower and upper limits on where the mean (average) treatment effect across all studies is expected to lie. CIs reflect the precision with which the mean effect size is estimated. A 95% prediction interval estimates the range for 95% of the individual treatment effects of the included studies—in addition to where the treatment effect of a new study from a comparable population would lie with 95% probability. A prediction interval is a measure of dispersion of individual results. This will provide a wider range of expected treatment effects compared with 95% CI.

For example, a recent meta-analysis found that in the absence of local infiltration analgesia, adding local anesthetic infiltration between the popliteal artery and capsule of the knee after total knee arthroplasty reduced resting pain scores at 6 hours by a weighted difference in means of -1.33 (95% CI, -1.57 to -1.09).⁵³ However, the 95% prediction interval (ie, expected range) for what one might expect if someone conducted their own randomized controlled trial in a similar population resulted in a wider interval (-2.04 to -0.62). Note that the expected effect size (reduced pain scores) now varies from clinically important to trivial. Potentially, this therapy is effective in some populations or patients but not in others. Using again the meta-analysis by Park *et al*,⁵⁰ the 95% CI based on the frequentist approach and random-effects model was -1.36 to -0.44 . This suggests that we are 95% confident (because the 95% CI will miss the true effect 5% of the time) that the mean reductions in pain lie between -1.36 and -0.44 on a numerical rating scale. However, the 95% prediction interval suggests that an individual study from a comparable population would reduce pain scores between -2.78 and 0.98 . As with the above example, the therapy may be more effective in some populations compared with others. Formulas and examples for calculating 95% prediction intervals can be found in the article by IntHout *et al*.⁵²

Overall summary: The effect measure (ratio or difference) is the statistical construct that compares outcome data between two groups estimating the magnitude of the intervention effect. Effect estimates from individual studies are combined using statistical models to generate a single parameter to estimate the intervention effect. This parameter, for example, the OR or standardized mean difference, is reported with its CI as a measure of precision. The prediction interval is a measure of dispersion of individual effect sizes, and we recommend when there are a sufficient number of studies that it be presented alongside effect estimates for random-effects meta-analyses.

DESCRIBE HOW THE RISK OF BIAS DUE TO MISSING RESULTS IN THE EVIDENCE SYNTHESIS WILL BE ASSESSED

Publication bias, or nonreporting bias (the term preferred by authors of Chapter 13 of the *Cochrane Handbook*²⁶), refers to the selective reporting of research results/manuscripts for publication, or accepted for publication, which most often occurs with smaller studies. Entire completed studies may remain unpublished, or specific results from published studies may be

missing or reported in a format that precludes meta-analysis. Making efforts to obtain unpublished reports will reduce the risk of missing data from entire studies.²⁶ Red flags for nonreporting bias include substantial methodological changes (eg, prioritization of published outcomes inconsistent with intended primary outcomes described in the trial registry), or the reporting of outcomes with *P* values alone and omitting summary statistics.

There are several approaches to assessing the risk of bias from results missing from the evidence synthesis. A funnel plot is a scatter plot of study size or precision (SE or inverse of SE) on the vertical axis against the intervention effect estimates on the horizontal axis. Smaller studies tend to show different and larger estimated treatment effects than larger ones.⁵⁴ Therefore, smaller studies generate estimates that scatter widely at the bottom part of the plot. In contrast, the precision of the estimated treatment effects of the larger studies is increased (less scatter) relative to the smaller studies. In the absence of bias and between-study heterogeneity, the scatter will be due to sampling variation alone and typically generates an inverted ‘funnel shape’ or a funnel plot that is symmetrical.⁵⁴

Two of the most common approaches for assessing small-study effects (ie, nonreporting biases) are to examine the funnel plot for asymmetry and use Egger’s regression-intercept test. However, the Doi plot should be considered for all meta-analyses because it may be more intuitive than the funnel plot and the Luis Furuya-Kanamori (LFK) index more robust than Egger’s regression-intercept test.⁵⁵ Funnel plot asymmetry has multiple possible causes, including (1) nonreporting biases (eg, publication bias, selective outcome reporting); (2) poor methodological quality leading to spuriously inflated effects, often in smaller studies; (3) true heterogeneity (ie, magnitude of treatment effect differs, or is shifted, according to study size); and (4) chance (because the collected studies are subject to sampling variability).⁵⁴ In the end, though, if there is evidence of asymmetry in the funnel plot, authors may not be confident about the cause(s). Furthermore, lack of asymmetry should not be a reason to claim there was no reporting bias or other issues. In general, asymmetry in a funnel plot is described as evidence of ‘small study effects.’ However, this should not be equated in absolute terms with publication or reporting bias. Small-study effect results (publication bias, etc) are also incorporated into the assessment of the quality (certainty) of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) instrument described later in this paper.⁵⁶

Overall summary: Both a qualitative (figure) and quantitative assessment of small-study effects, that is, nonreporting biases, should be described in the protocol.

DESCRIBE PLANNED METHODS TO EVALUATE AND IDENTIFY SOURCES OF HETEROGENEITY OF RESULTS ACROSS STUDIES

A meta-analysis assumes some degree of clinical and methodological homogeneity between trials. Despite pooling studies that appear similar, there will likely be heterogeneity because of diversity in participants, interventions, outcomes, practice environment, design, and methods. Therefore, the individual studies may have very different results. In this setting, heterogeneity confounds the interpretation of the single numerical estimate the meta-analysis generates. Clinicians may lack confidence in the result. However, the presence of heterogeneity may present opportunities to identify best practice. For example, authors may have preidentified study-level clinical covariates that subsequently explain or resolve heterogeneity. Therefore,

importantly, in the protocol, authors should use their content expertise to identify and document a priori potential sources of clinical and/or methodological heterogeneity before screening studies and pooling the results.

Evaluating heterogeneity begins with identifying a priori clinical and methodological diversity driving variability in study results. Heterogeneity is 1 of 5 domains used to evaluate quality (certainty) of the evidence (see GRADE description that follows).⁵⁶ The presence of significant heterogeneity may be a principal factor when deciding or not to proceed with meta-analysis of effect estimates (PRISMA-P Item 15 a).² Note that heterogeneity is driven by the PICO framework, which determines the scope of the review and the diversity of included studies. Table 3 provides examples of clinical and methodological heterogeneity.

Statistical heterogeneity refers to inconsistency in the magnitude of the treatment effect estimate between studies that is more than would be expected or explained by sampling variability, measurement error, or chance. In their review protocol, authors should be explicit about how they are going to measure and address statistical heterogeneity (PRISMA-P Item 15b).² Commonly reported methods to test for and quantify statistical heterogeneity include Cochran's Q test, Tau² (τ^2), and the I² statistic.^{38 62} These statistics have limitations: one example is a meta-analysis comprised of a small number of eligible studies, with small sample sizes and heterogeneous effect estimates. In this scenario, there are limitations with Cochran's Q test (low power to reject null hypothesis of homogeneity)^{38 62} and the I² statistic (lack of precision and bias in point estimate).⁶³ As an example, I² values <25% suggest that heterogeneity may be very low; however, a value of 0 does not exclude heterogeneity (failure to reject the null hypothesis may imply lack of statistical power). The CI of I² is often wide, and it may cross thresholds of I² used to categorize heterogeneity: very low (<25%), low (25% to <50%), moderate (50% to <75%), and large (\geq 75%).⁶² However, these limitations are related to the number of studies pooled in the meta-analysis. The uncertainty surrounding Tau² and the point estimate of I² can be substantial with a small number of studies. When important statistical heterogeneity is unexplained by clinical or methodological factors (the usual scenario),³⁸ then homogeneity is questioned, and the validity of the meta-analysis results are potentially compromised. In the setting of substantial unresolved heterogeneity, the results of a meta-analysis may be misleading or at least difficult to interpret, and authors may need to consider a method of evidence synthesis that does not involve meta-analysis of effect estimates.^{38 64 65} It is reasonable for authors to state this conditional option a priori in their review protocol (PRISMA-P Items 15a, 15d).

Review authors should consider a limited number of predefined sensitivity, subgroup, moderator, and/or meta regression analyses (PRISMA-P Item 15 c).² Note that the absence of statistical heterogeneity does not negate such preplanned analyses because, as with a single randomized trial, it is relevant to consider how consistent the treatment effect is across levels of key patient baseline or study-specific covariates. When developing a protocol that includes effect modification analysis, authors can consider the recently developed 10-item Credibility of Effect Modification Analyses (ICEMAN) checklist⁶⁶ for meta-analysis.

Sensitivity analysis refers to analyses conducted to assess the robustness of results across various utilized methods or assumptions (*Cochrane Handbook* section 10.14).³⁸ Most meta-analyses do not include a large number of studies. Therefore, it is recommended that influence analysis, a form

of sensitivity analysis, be planned to examine the overall results, including statistical heterogeneity, where each study is deleted from the model once. In addition, and regardless of the number of pooled studies, outlier analysis, another form of sensitivity analysis, should be preplanned. One approach to address outlier analysis is to examine results by deleting effect sizes from studies in which their 95% CIs fall completely outside the overall pooled 95% CI. Finally, some meta-analyses may include one or more large studies that comprise the majority of the weight, for example, 50% or more, when pooled. However, rather than avoiding a meta-analysis, the result being a loss of potentially important information, it is more appropriate to plan and to conduct a sensitivity analysis with these studies deleted from the model to see how it affects the overall pooled results.

Subgroup and moderator analyses refer to analyses based on factors such as population characteristics (eg, sex) and/or variants of the intervention or components of the outcome (eg, length of follow-up).³⁸ Subgroup analyses do not include the calculation of between-study variance, whereas moderator analyses do. When planning such analyses, there should be a clear rationale (biologic, clinical, methodologic) and/or existing research suggesting potential subgroup differences.³⁸ A key aspect of such analyses is not simply to assess the meta-analysis treatment effect within levels of a characteristic, but to test the treatment-by-covariate interaction.²¹

Guidelines encourage and show preference for predefined analyses over post hoc analyses. Therefore, authors should clarify when their selected analyses were established. For example, in a meta-analysis that compared the effect of regional anesthesia vs general anesthesia on cancer recurrence, the authors were transparent regarding predefined subgroup analyses and post hoc subgroup analyses.⁶⁷

Metaregression merges meta-analytic and regression principles to explore heterogeneity.^{38 68} Metaregression determines if and how much a study-level covariate contributes to heterogeneity of the treatment effects between studies. It is a method to assess treatment effect heterogeneity at the study level³⁸ (*Cochrane Handbook* section 10.11.4).³⁸ Covariates may be study-specific (eg, drug dosage) or ecological (ie, requiring individual subject data to assess patient-level factors).

Metaregression allows the author to estimate the treatment effect while controlling for differences across studies, as well as assess which covariates account for most of the heterogeneity. Furthermore, the metaregression approach is weighted and reduces the probability of false positive findings compared with subgroup analysis.⁶⁸ Since the study is the unit of analysis, the ability to conduct a metaregression is often obviated by the need for many studies to assess covariate effects.⁶⁸ The covariates included in a metaregression should be few and prespecified in the review protocol (PRISMA-P Item 15 c).²

Acknowledging these limitations, we recommend including no more than one metaregression covariate for every 10 studies/effect sizes for a continuous variable and four studies/effects sizes per group for a categorical covariate.^{68 69} Therefore, given the small number of studies included in most meta-analyses, subgroup, moderator, and metaregression analyses may not be feasible. In addition, it is important to understand that analyses such as metaregression within the context of an aggregate data meta-analysis do not support causal inferences because covariates are not randomly assigned in studies. In fact, metaregression analyses are targeting moderators of the treatment effect, not new interventions. Therefore, any observed findings would need to be tested in original randomized controlled trials.

Finally, any preplanned analyses should include the choice of model(s) and software, including version, used to conduct the analyses. For example, moderator analyses can occur where both study-level and categorical analyses occur using a fixed, random, or mixed (fixed and random) approach. Refer to [table 4](#) for more detail on methods for exploring heterogeneity, including their limitations.

Overall summary: Authors should document potential sources of heterogeneity and limited sensitivity/subgroup/moderator/metaregression analyses a priori in the protocol. Predefining potential effect modifiers and sources of heterogeneity is important but does not mitigate the risk of multiple comparisons—adjustments are still warranted. A meta-analysis with a small number of studies may give vulnerable results and provide limited information on sources of heterogeneity, and subgroup, moderator, and metaregression analysis may not be feasible. No statistical methods can overcome the potential limitations created by meta-analyzing a small number of studies.

DESCRIBE CONDITIONS WHEN META-ANALYSIS MAY NOT BE POSSIBLE AND PRACTICAL SOLUTIONS

Meta-analysis of effect estimates may not be feasible or needed. Reasons include paucity of studies, or paucity of studies with required outcomes, different effect measures, bias (missing studies, missing data), and heterogeneity. Even if there is little literature on a topic, the results of a systematic review may be unique (no previous review published) but insufficient to perform meta-analysis. In this scenario, the systematic review without meta-analysis may be of value even if to indicate the lack of data and direction for future research. If there is substantial variation in results, especially if this includes the direction of effect, it may be disingenuous to report a single numerical estimate of the treatment effect. For example, the authors who conducted a systematic review assessing the efficacy and safety of magnesium for treatment of chronic pain initially planned for meta-analysis.⁵⁹ However, the authors stated that the presence of significant heterogeneity among the included studies precluded any meta-analyses, and they therefore conducted a systematic review without meta-analysis.

Authors planning a meta-analysis should be mindful that a systematic review alone can provide a robust review of the evidence and that there are guidelines to assist them preparing their review in this format.⁶⁴ Per recommendations from the *Cochrane Handbook* (Chapter 12.1 a),⁶⁴ it is valid to build contingencies into the protocol analysis plan if a meta-analysis is not possible. PRISMA recommends describing the type of summary planned when quantitative synthesis is not possible (PRISMA-P Items 15a, 15d).² Furthermore, contingencies that generalize the scope of a predefined PICO synthesis that initially addressed a very narrow question may also be included. This strategy enables capture and synthesis of a larger number of studies in situations where studies are lacking on a narrowed, specific topic.

Examples of building contingencies into the protocol analysis plan in one or more groups of the PICO elements at a broader level are as follows (note differences in italicized wording): (a) ‘the effect of *any lower extremity regional anesthetic block* on...’ instead of ‘the effect of *only femoral nerve blockade* on...’; (b) ‘the effect of multimodal analgesia on postoperative pain score at *any time-point up to 24 hours*’ instead of ‘the effect of multimodal analgesia on postoperative pain score at *2 hours only*’; and ‘the effect of intranasal fentanyl in *children* and adolescents on...’ instead of ‘the effect of intranasal fentanyl *in children* on...’ Despite the lack of specificity in these broader questions, they

may still address an important question relevant to clinical practice, identify specific areas where evidence is lacking, and thus provide an avenue for future research efforts. In [table 3](#) in our accompanying Part II paper, we provide examples of presenting and reporting systematic reviews without meta-analysis.

Overall summary: Heterogeneity that remains strong despite accounting for a priori sources of heterogeneity is an important factor when deciding to proceed with meta-analysis of effect estimates (PRISMA-P Item 15 a).²

DESCRIBE METHODS USED TO ASSESS THE QUALITY (CERTAINTY) AND STRENGTH OF THE EVIDENCE (GRADE)

Medical journals require authors use the GRADE process to evaluate the quality (certainty) of the evidence and strength of recommendations from the body of evidence for all outcomes that they report.⁷⁰ This should be conducted after completing risk of bias assessment and all statistical analyses, assuming a meta-analysis is included for the latter. The GRADE method assesses the overall quality (certainty) of evidence based on the following five domains: risk of bias,⁷¹ heterogeneity (inconsistency),⁷² indirectness,⁷³ precision,⁷⁴ and publication bias⁷⁵ (PRISMA-P Item 17).² How this should be reported is included in the accompanying Part II paper.

Overall summary: Review authors are required to use the GRADE process to report the quality (certainty) of evidence in summary of finding and evidence profile tables.

CONCLUSIONS

Performing a systematic review and performing a meta-analysis are substantial, complex, resource-intensive processes. The systematic review team should include meta-analysis and information specialists. Guidelines warrant transparency in the design, conduct, and reporting of a systematic review so that its results can be correctly interpreted. Transparency begins with authors describing their rationale for a systematic review, and this principle should be evident throughout all stages of the review process. The meta-analysis process is likely to involve some change partly related to the uncertainty of the characteristics of eligible studies when the protocol is designed. Publication of this article simultaneously in the two journals *Regional Anesthesia and Pain Medicine* and *Anesthesia & Analgesia* highlights the need for systematic review and meta-analysis authors to read and to consider its content. However, we acknowledge that we have reinforced important foundational concepts and content currently existing in the important resources of PRISMA, Cochrane, and AMSTAR. Authors should also refer to these resources as they develop, implement, and report the systematic review protocol.

Contributors MJB conceived and led the development of this project (Parts I and II); drafted the original and subsequent versions of this manuscript including its format, structure, and important intellectual content; and prepared the original manuscript for submission and subsequent revisions for resubmission. RSD contributed important intellectual content to this manuscript, drafted substantial content for versions of this manuscript, developed its format and structure, and approved the final version. EJM contributed important intellectual content to the manuscript, developed its format and structure, and approved the final version. SN contributed important intellectual content to the manuscript and approved the final version. GAK contributed important intellectual content to this manuscript, provided mentorship and content expertise, drafted substantial content for versions of this manuscript, developed its format and structure, and approved the final version. This manuscript was handled by: Thomas R. Vetter, MD, MPH, MFA.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests MJB is a Regional Anesthesia & Acute Pain Medicine Executive Section Editor of Anesthesia & Analgesia. RSD: None. EJM is Statistical Editor for Anesthesia & Analgesia. SN is a President of American Society of Regional Anesthesia and Pain Medicine. GAK is a Statistical Consultant for Regional Anesthesia and Pain Medicine.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- Chloros GD, Prodomidis AD, Giannoudis PV. Has anything changed in evidence-based medicine? *Injury* 2022.
- Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647:g7647.
- Chandler J, Cumpston M, Thomas J, et al. Chapter 1: introduction. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.
- Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Innocenti T, Feller D, Giaggio S, et al. Adherence to the PRISMA statement and its association with risk of bias in systematic reviews published in rehabilitation journals: a meta-research study. *Braz J Phys Ther* 2022;26:100450.
- McGregor RH, Warner FM, Linde LD, et al. Quality of meta-analyses of non-opioid, pharmacological, perioperative interventions for chronic postsurgical pain: a systematic review. *Reg Anesth Pain Med* 2022;47:263–9.
- Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of Healthcare interventions, or both. *BMJ* 2017;358:j4008.
- Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162:777–84.
- Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;313:1657–65.
- Thomas J, Kneale D, McKenzie JE, et al. Chapter 2: determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- McKenzie JE, Brennan SE, Ryan RE, et al. Chapter 3: defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Waugh N. Misgivings about PROSPERO in solving the problem of duplicate systematic reviews. *BMJ* 2013;347:bmj.f5496.
- Chen B, Duan J, Wen S, et al. An updated systematic review and meta-analysis of Duloxetine for knee osteoarthritis pain. *Clin J Pain* 2021;37:852–62.
- Garner P, Hopewell S, Chandler J, et al. Panel for updating guidance for systematic reviews (PUGs). Panel for updating guidance for systematic reviews (PUGs). When and how to update systematic reviews: consensus and checklist. *BMJ* 2016;354:i3507.
- Macfarlane AJR, Prasad GA, Chan VWS, et al. Does regional anesthesia improve outcome after total knee arthroplasty? *Clin Orthop Relat Res* 2009;467:2379–402.
- Chan E-Y, Fransen M, Parker DA, et al. Femoral nerve blocks for acute postoperative pain after knee replacement surgery. *Cochrane Database Syst Rev* 2014;2014. Issue 5. Art. No.: CD009941:CD009941.
- Zorrilla-Vaca A, Li J. The role of Sciatic nerve block to complement femoral nerve block in total knee Arthroplasty: a meta-analysis of randomized controlled trials. *J Anesth* 2018;32:341–50.
- Vetter TR, Mascha EJ. Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. *Anesth Analg* 2017;125:678–81.
- Higgins JPT, Li T, Deeks JJ. Chapter 6: choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: www.training.cochrane.org/handbook
- Bender R, Bunce C, Clarke M, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857–65.
- Hussain N, Brull R, Noble J, et al. Statistically significant but clinically unimportant: a systematic review and meta-analysis of the analgesic benefits of erector spinae plane block following breast cancer surgery. *Reg Anesth Pain Med* 2021;46:3–12.
- McGowan J, Sampson M, Salzwedel DM, et al. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;75:40–6.
- Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol* 2019;111:127–34.
- Boutron I, Page MJ, Higgins JPT, et al. Chapter 7: considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Page MJ, Higgins JPT, Sterne JAC. Chapter 13: assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: www.training.cochrane.org/handbook
- Bramer WM, Rethlefsen ML, Kleijnen J, et al. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev* 2017;6:245.
- Harrison H, Griffin SJ, Kuhn I, et al. Software tools to support title and abstract screening for systematic reviews in Healthcare: an evaluation. *BMC Med Res Methodol* 2020;20:7.
- Li T, Higgins JPT, Deeks JJ. Chapter 5: collecting data. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. introduction- GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
- Schünemann HJ, Higgins JPT, Vist GE, et al. Chapter 14: completing "summary of findings" tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Hussain N, Brull R, Zhou S, et al. Analgesic benefits of single-shot versus continuous adductor canal block for total knee arthroplasty: a systemic review and meta-analysis of randomized trials. *Reg Anesth Pain Med* 2023;48:49–60.
- D'Souza RS, Gurrieri C, Johnson RL, et al. Intraoperative methadone administration and postoperative pain control: a systematic review and meta-analysis. *Pain* 2020;161:237–43.
- Stone JC, Glass K, Clark J, et al. The methodological standards for epidemiological research (MASTER) scale demonstrated a unified framework for bias assessment. *J Clin Epidemiol* 2021;134:52–64.
- Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898.
- Higgins JPT, Savović J, Page MJ, et al. Chapter 8: assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: www.training.cochrane.org/handbook
- Pace NL. Research methods for meta-analyses. *Best Pract Res Clin Anaesthesiol* 2011;25:523–33.
- Deeks JJ, Higgins JPT, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- Gibson CA, Bailey BW, Carper MJ, et al. Author contacts for retrieval of data for a meta-analysis on exercise and diet restriction. *Int J Technol Assess Health Care* 2006;22:267–70.
- Follmann D, Elliott P, Suh I, et al. Variance imputation for overviews of clinical trials with continuous response. *J Clin Epidemiol* 1992;45:769–73.
- Borenstein M, Hedges LV, Higgins JPT, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;1:97–111.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
- DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007;28:105–14.
- Biggestaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat Med* 1997;16:753–68.
- Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med* 2002;21:3153–9.
- Doi SAR, Barendregt JJ, Khan S, et al. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. *Contemp Clin Trials* 2015;45:130–8.
- Doi SAR, Furuya-Kanamori L, Thalib L, et al. Meta-analysis in evidence-based Healthcare: a paradigm shift away from random effects is overdue. *Int J Evid Based Healthc* 2017;15:152–60.
- Doi SAR, Furuya-Kanamori L. Selecting the best meta-analytic Estimator for evidence-based practice: a simulation study. *Int J Evid Based Healthc* 2020;18:86–94.

- 49 Doi SAR, Barendregt JJ, Khan S, *et al.* Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. *Contemp Clin Trials* 2015;45:123–9.
- 50 Park SK, Yoon S, Kim BR, *et al.* Pre-Emptive epidural analgesia for acute and chronic post-Thoracotomy pain in adults: a systematic review and meta-analysis. *Reg Anesth Pain Med* 2020;45:1006–16.
- 51 IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard Dersimonian-Laird method. *BMC Med Res Methodol* 2014;14:25. doi: 10.1186/1471-2288-14-25:25.
- 52 IntHout J, Ioannidis JPA, Rovers MM, *et al.* Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 2016;6:e010247.
- 53 Hussain N, Brull R, Sheehy B, *et al.* Does the addition of iPACK to adductor canal block in the presence or absence of periarticular local anesthetic infiltration improve analgesic and functional outcomes following total knee arthroplasty? A systematic review and meta-analysis. *Reg Anesth Pain Med* 2021;46:713–21.
- 54 Sterne JAC, Sutton AJ, Ioannidis JPA, *et al.* Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:bmj.d4002.
- 55 Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. *Int J Evid Based Healthc* 2018;16:195–203.
- 56 Balshem H, Helfand M, Schünemann HJ, *et al.* GRADE guidelines: 3. rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
- 57 Zhang L, Hu Y, Wu X, *et al.* A systematic review and meta-analysis of randomized controlled trials of labor epidural analgesia using moderately high concentrations of plain local anesthetics versus low concentrations of local anesthetics with opioids. *JPR* 2021;Volume 14:1303–13.
- 58 Dinges HC, Wiesmann T, Otremba B, *et al.* The analgesic efficacy of liposomal bupivacaine compared with bupivacaine hydrochloride for the prevention of postoperative pain: a systematic review and meta-analysis with trial sequential analysis. *Reg Anesth Pain Med* 2021;46:490–8.
- 59 Park R, Ho AM-H, Pickering G, *et al.* Efficacy and safety of magnesium for the management of chronic pain in adults: a systematic review. *Anesth Analg* 2020;131:764–75.
- 60 Zeng C, Wei J, Persson MSM, *et al.* Relative efficacy and safety of topical non-Steroidal anti-inflammatory drugs for osteoarthritis: a systematic review and network meta-analysis of randomised controlled trials and observational studies. *Br J Sports Med* 2018;52:642–50.
- 61 Jagannathan R, Niesen AD, D'Souza RS, *et al.* Intermittent bolus versus continuous infusion techniques for local anesthetic delivery in peripheral and Truncal nerve analgesia: the current state of evidence. *Reg Anesth Pain Med* 2019;44:447–51.
- 62 Higgins JPT, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
- 63 von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med Res Methodol* 2015;15:35.
- 64 McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane, 2022. Available: <https://www.training.cochrane.org/handbook>
- 65 Campbell M, McKenzie JE, Sowden A, *et al.* Synthesis without meta-analysis (swim) in systematic reviews: reporting guideline. *BMJ* 2020;368:l6890. doi: 10.1136/bmj.l6890:l6890.
- 66 Schandelmaier S, Briel M, Varadhan R, *et al.* Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ* 2020;192:E901–6.
- 67 Zhang YL, Pei LJ, Sun C, *et al.* Regional anesthesia and cancer recurrence in patients with late-stage cancer: a systematic review and meta-analysis. *Chinese Medical Journal* 2021;134:2403–11.
- 68 Baker WL, White CM, Cappelleri JC, *et al.* Understanding heterogeneity in meta-analysis: the role of meta-regression. *Int J Clin Pract* 2009;63:1426–34.
- 69 Fu R, Gartlehner G, Grant M, *et al.* Conducting quantitative synthesis when comparing medical interventions: AHRQ and the effective health care program. October 25, 2010. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality (US), 2008. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21433407>
- 70 Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 2. framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
- 71 Guyatt GH, Oxman AD, Vist G, *et al.* GRADE guidelines: 4. rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- 72 Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 7. rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
- 73 Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 8. rating the quality of evidence—Indirectness. *J Clin Epidemiol* 2011;64:1303–10.
- 74 Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines 6. rating the quality of evidence—Imprecision. *J Clin Epidemiol* 2011;64:1283–93.
- 75 Guyatt GH, Oxman AD, Montori V, *et al.* GRADE guidelines: 5. rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64:1277–82.