Users' guide to the surgical literature: how to use a systematic literature review and meta-analysis

Mohit Bhandari, MD, MSc; P.J. Devereaux, MD; Victor Montori, MD, MSc; Claudio Cinà, MD, MSc; Ved Tandan, MD, MSc; Gordon H. Guyatt, MD, MSc; for the Evidence-Based Surgery Working Group

An evidence-based approach to surgery incorporates patients' circumstances or predicaments, identifies knowledge gaps and frames questions to fill those gaps, conducts efficient literature searches, critically appraises the research evidence and applies that evidence to patient care. The practice of evidence-based medicine, therefore, is a process of lifelong self-directed learning in which caring for patients creates a need for clinically important information about diagnoses, prognoses, treatments and other health care issues.[1,2]

Readers are able to utilize several types of summarized information from expert opinion and textbook reviews to systematic reviews. Traditional, or *narrative*, reviews, by definition, do not use a systematic approach to identifying information on a particular topic. Moreover, narrative reviews often pose background-type questions and provide a general overview of a topic such as those found in book chapters and instructional course lectures. A background question is, for example, "What is the epidemiology, clinical presentation, treatment options and prognosis following femoral shaft fractures in adults?" We use the term *systematic review* for any summary of the medical literature that attempts to address a focused clinical question with explicit strategies for the identification and appraisal of the available literature (Table 1 and Table 2); *meta-analysis* is a term used for systematic reviews that use quantitative methods (i.e., statistical techniques) to summarize the results. Systematic reviews typically pose a foreground-type question. Foreground questions are more specific and provide insight into a particular aspect of management. For instance, investigators may provide a systematic review of plating versus nailing of humeral shaft fractures on nonunion rates (foreground question) rather than a general review of how bone heals after all treatments of humeral shaft fractures (background question).

Whereas systematic reviews (and meta-analyses) have become popular in surgery, they are not without limitations. The quality of the systematic review is influenced by the quality of the primary studies being reviewed. However, in the absence of large, definitive clinical trials, meta-analyses can provide important information to guide patient care as well as future clinical research.

In applying the suggested guidelines (Table 1) you will gain a clear understanding of the process of conducting a systematic review (Table 2).

The conduct and interpretation of systematic reviews in surgery is often

From the Surgical Outcomes Research Centre, Department of Surgery, St. Joseph's Healthcare Hamilton and McMaster University, and the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.

Accepted for publication June 13, 2003.

Correspondence to: *Dr. Mohit Bhandari, Department of Clinical Epidemiology and Biostatistics, Rm. 2C12, McMaster University Health Sciences Centre, 1200 Main St. W, Hamilton ON L8N 3Z5; fax 905 524-3841; bhandari@sympatico.ca*

The Evidence-Based Surgery Working Group members include: Stuart Archibald, MD;[†‡] Mohit Bhandari, MD;[†] Charles H. Goldsmith, PhD;[‡§] Dennis Hong, MD;[†] John D. Miller, MD;[*†‡] Marko Simunovic, MD, MPH;[†‡§¶] Ved Tandan, MD, MSc;[*†‡§] Achilleas Thoma, MD;[†‡] John Urschel, MD;[†‡] Sylvie Cornacchi, MSc[†‡]*

**Department of Surgery, St. Joseph's Hospital, †Department of Surgery, ‡Surgical Outcomes Research Centre and §Department of Clinical Epidemiology and Biostatistics, McMaster University, and ¶Hamilton Health Sciences, Hamilton, Ont.*

This manuscript is based, in part, on: Guyatt GH, Rennie D, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. *Chicago: AMA Press; 2002 and Bhandari M, Guyatt GH, Montori V, Devereaux PJ, Swiontkowski MF. User's guide to the orthopaedic literature: how to use a systematic literature review.* J Bone Joint Surg Am *2002;84:1672-82.*

 ' 2004 Canadian Medical Association

challenging given the paucity of clinical trails available on any given topic. However, if investigators adhere to proper methodology, they can provide conclusions drawn from a comprehensive study with limited bias.

Clinical scenario

You are an orthopedic surgeon who has recently joined a group of orthopedic surgeons practising in an academic centre. You have an interest in injuries about the foot and ankle and have noticed that the treatment of ruptures of the Achilles tendon differs from that of your recent experience acquired during your fellowship training. Your colleagues prefer nonoperative treatment of Achilles tendon ruptures because they believe that outcomes are good with this technique. Having trained with an orthopedic surgeon who preferred operative repair for Achilles tendon ruptures for the same reasons of improved outcome, you begin to wonder whether your new colleagues know something your fellowship supervisor did not.

You decide to challenge another colleague who uses nonoperative treatment to provide a study to support her choice. She replies, "There's one randomized trial from Europe, but I'm sure there is lots of information on this topic in the literature. Why don't you present a summary of the information on this topic at next week's grand rounds?"

Intrigued by this opportunity, you gladly accept your colleague's challenge and begin to look for relevant information.

The search

You quickly determine from talking with your colleagues and your fellowship supervisor that there have been a number of randomized trials comparing operative and nonoperative treatment of acute ruptures of the Achilles tendon. Realizing that your 1-week deadline will not be sufficient to summarize all of the articles, you decide to focus your literature search to identify any recent reviews of this topic. Being relatively proficient on the Internet, you select your favourite search site, National Library of Medicine's PubMed at www.pubmed.gov. You select the "Clinical Queries" section and choose a search for systematic reviews. You type in the words "Achilles tendon." This identifies 12 documents. You review the titles of the 12 potentially relevant studies and are happy to find a systematic review and meta-analysis of operative versus nonoperative treatment for acute ruptures of the Achilles tendon.[3] You retrieve this article for further review.

Are the results of this review valid?

Did the review explicitly address a sensible clinical question?

Consider a systematic overview that pooled results from all fracture therapies (both surgical and medical) for all types of fractures to generate a

Table 1

Users' Guides for How to Use Review Articles
Are the results valid? • Did the review explicitly address a sensible clinical question?
• Was the search for relevant studies detailed and exhaustive?
• Were the primary studies of high methodologic quality?
• Were assessments of studies reproducible?
What are the results? • Were the results similar from study to study?
• What are the overall results of the review?
• How precise were the results?
How can I apply the results to patient care? • How can I best interpret the results to apply them to the care of my patients?
• Were all clinically important outcomes considered?
• Are the benefits worth the costs and potential risks?

Table 2

The Process of Conducting a Systematic Review
Define the question • Specify inclusion and exclusion criteria Population Intervention or exposure Outcome Methodology • Establish a priori hypotheses to explain heterogeneity
Conduct a literature search • Decide on information sources: databases, experts, funding agencies, pharmaceutical companies, personal files, registries, citation lists of retrieved articles
• Determine restrictions: time frame, unpublished data, language
• Identify titles and abstracts
Apply inclusion and exclusion criteria • Apply inclusion and exclusion criteria to titles and abstracts
• Obtain full articles for eligible titles and abstracts
• Apply inclusion and exclusion criteria to full articles
• Select final eligible articles
• Assess agreement among reviewers on study selection
Create data abstraction • Assess methodologic quality (validity of the study)
• Assess agreement among reviewers on validity decisions
• Data abstraction: participants, interventions, comparison interventions, study design
• Results
Conduct analysis • Determine method for pooling results
• Pool results (if appropriate)
• Decide on handling missing data
• Explore heterogeneity
Sensitivity and subgroup analysis • Explore possibility of publications bias

single estimate of the impact on fracture union rates. Clinicians would not find this type of review useful: they would conclude that it is "too broad." What makes a systematic review too broad? We believe the underlying question that clinicians ask themselves when considering whether a review is excessively broad is as follows: Across the range of patients and interventions included, and ways the outcome was measured, can I expect more or less the same magnitude of effect?

The reason that clinicians reject "all therapies for all fracture types" is that they know that some fracture therapies are extremely effective and others are harmful. Pooling across such therapies would yield an intermediate estimate of effect inapplicable to either the highly beneficial or harmful interventions. Clinicians also know that fracture types differ in their biology and response to treatment, again making estimates of average treatment effects inapplicable to all fractures.

The task of the reader, then, is to decide whether the range of patients, interventions or exposures and the outcomes chosen make sense. To help them with this decision, reviewers need to present a precise statement of what range of patients, exposures and outcomes they have decided to consider; in other words, they need to define explicit inclusion and exclusion criteria for their review. Explicit eligibility criteria not only facilitate the user's decision regarding whether the question is sensible but make it less likely that the authors will preferentially include studies supporting their own prior conclusion. Bias in choosing articles to cite is a problem for both systematic reviews and original reports of research.

There are good reasons to choose broad eligibility criteria. First, one of the primary goals of a systematic review, and of pooling data in particular, is to obtain a more precise estimate of the treatment effect. The broader the eligibility criteria, the greater are the number of studies and number of patients, and the narrower are the confidence intervals (CIs). Second, broad eligibility criteria lead to more generalizable results. If it is true that the results apply to a wide variety of patients with a wide range of injury severities, the surgeon is on stronger ground applying the findings to a particular patient.

Was the search for relevant studies detailed and exhaustive?

It is important that authors conduct a thorough search for studies that meet their inclusion criteria. Their search should include the use of bibliographic databases such as MEDLINE, EMBASE, the Cochrane Controlled Trials Register (containing more than 250 000 randomized clinical trials); checking the reference lists of the articles they retrieved; and personal contact with experts in the area (Table 3). It may also be important to examine books of recently published abstracts presented at scientific meetings, and less frequently used databases, including those that summarize doctoral theses. With all these sources, it becomes evident that a MEDLINE search alone will not be satisfactory. Previous meta-analyses in orthopedics have variably reported a comprehensive search strategy.[4]

Unless the authors tell us what they did to locate relevant studies, it is difficult to know how likely it is that relevant studies were missed. There are 2 important reasons why authors of a review should use personal contacts. The first is to identify published studies that might have been missed (including studies that are in press or not yet indexed or referenced). The second is to identify unpublished studies. Although controversies remain about including unpublished studies,[1,2,5,6] their omis-

Table 3

Potential Information Resources

Resource	URL*
Synopsis of systematic reviews	
• ACP Journal Club	www.acpjc.org
• Clinical evidence	www.clinicalevidence.com
Databases of systematic reviews	
• The Cochrane Library	www.cochranelibrary.com
• University of York/NHS Centre for Reviews and Dissemination	www.york.ac.uk/inst/crd/
• Bandolier	www.jr2.ox.ac.uk/bandolier/
Databases of systematic reviews and primary studies	
• MEDLINE	www.PubMed.gov
• Searching for systematic reviews in MEDLINE (clinical queries filter in PubMed)	www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html
Critical appraisal tools to practise evidence-based medicine	
• The Users' Guides to the Medical Literature	http://ugi.usersguides.org
• The Centre for Health Evidence	www.cche.net
• Centre for Evidence-Based Medicine, University of Toronto	www.cebm.utoronto.ca
Collection of evidence-based medicine resources	
• Netting the evidence	www.shef.ac.uk/uni/academic/R-Z/scharr/ir/netting/
• The McMaster Evidence Based Practice Resource	www-hsl.mcmaster.ca/ebcp

*Accessed 2003 Sept. 30.

sion increases the chances that studies with positive results will be overrepresented in the review (leading to a systematic overestimation of the treatment effect, referred to as publication bias).[7] The tendency for authors to differentially submit — and journals to differentially accept — studies with positive results constitutes a serious threat to the validity of systematic reviews.

If investigators include unpublished studies in a review, they should obtain full written reports and appraise the validity of both published and unpublished studies, and they may use statistical techniques to explore the possibility of publication bias. Reviews based on a small number of small studies with weakly positive effects are the most susceptible to publication bias.[2,8] The assessment of potential publication bias can be explored visually using a *funnel plot*.[2] This method uses a scatter plot of studies that relates the magnitude of the treatment effect to the weight of the study. An inverted funnel-shaped, symmetrical appearance of dots suggests that no study has been left out, whereas an asymmetrical appearance of dots, typically in favour of positive outcomes, suggests the presence of publication bias.

The authors of the systematic review of alternative management strategies for Achilles tendon ruptures identified articles with MEDLINE, the Cochrane Central Database of Randomized Trials (CENTRAL) and SCISEARCH, manual hand searches of orthopedic journals, textbooks and proceedings of annual orthopedic meetings. The investigators also contacted content experts. Ultimately, 11 potentially eligible studies were identified. Five of the 11 potentially eligible studies published in non-English journals (1 French, 4 German) were translated into English before additional eligibility review. After review of all 11 studies, 6 randomized trials (448 patients) were eventually included. The rigour of the reviewers' search methods reassures the clinician that omission of important studies is unlikely. Identifying articles published in non-English journals and articles outside North America strengthens the generalizability of the results.

Were the primary studies of high methodologic quality?

Even if a review article includes only randomized trials, it is important to know whether they were of good quality. Unfortunately, peer review does not guarantee the validity of published research. For exactly the same reason that the guides for using original reports of research begin by asking if the results are valid, it is essential to consider the validity of primary articles in systematic reviews. Differences in study methods might explain important differences among the results.[9] For example, studies with less rigorous methodology tend to overestimate the effectiveness of the intervention.[9,10] Consistent results from weak studies are less compelling than from strong ones. Consistent results from observational studies are particularly suspect. Physicians 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 is no single correct way to assess validity. Some investigators use long checklists to evaluate methodologic quality, others focus on 3 or 4 key aspects of the study.[11–14] Whether assessors of methodologic quality should be blinded remains a subject of debate.[14,15] In an independent assessment of 76 randomized trials, Clark and colleagues[15] were unable to find significant effect of reviewer blinding on quality scores.

Two of the authors of the Achilles tendon rupture review independently assessed the methodologic quality of each study, focusing on 6 methodologic domains (randomization and blinding, population, intervention, outcomes, follow-up and statistical analysis) and a summary quality scale. Study quality ranged from 57 to 72 points out of a maximum possible 100 points. Use of 2 independent assessors provided greater assurance that the assessment of quality was unbiased and reproducible.

The approach, while rigorous, omits an important aspect of validity. Randomization may fail to achieve its purpose of producing groups with comparable prognostic features if those enrolling patients are aware of the arm to which they will be allocated. For example, using year of birth or hospital identification numbers allow investigators to uncover the treatment allocation of their patients before enrolling them in a study. In a randomized trial of open versus laparoscopic appendectomy, the residents responsible for enrolling patients selectively avoided recruiting patients into the laparoscopic appendectomy group at night.[2] To the extent that patients coming in at night were sicker, this practice would have biased the results in favour of the laparoscopic appendectomy group. Allocation concealment (i.e., ensuring that study investigators do not know the treatment to which the next patient will be allocated) is a particularly important issue in surgical trials. As it turns out, not 1 of the trials considered in this systematic review instituted safeguards to ensure concealed randomization. Such safeguards require a separation of the roles of enrolment into the study and allocation into the study arms. In the laparoscopic appendectomy study for example, the investigators could have had the residents call a randomization centre to enrol a patient, and the surgical procedure to which the patient was randomized, communicated only after enrolment had been confirmed.

Were assessments of studies reproducible?

As we have seen, authors of review articles must decide which studies to include, how valid they are and what data to extract from them. Each of

these decisions requires judgement by the reviewers, and each is subject to both mistakes (random errors) and bias (systematic errors). Having 2 or more people participate in each decision guards against errors, and if there is good chance-corrected agreement among the reviewers, the clinician can have more confidence in the results of the review.[16,17]

The authors of the systematic review that addressed the management of Achilles tendon rupture assessed the reproducibility of the identification and assessment of study validity using the κ statistic and intraclass correlations (ICCs). Both of these estimate chance-corrected agreement, range between 0 and 1, with values closer to 1 representing better agreement.

The estimated κ statistic for the identification of potentially eligible studies was high (κ = 0.81, 95% CI 0.75–0.88). The ICC coefficient for rating of study quality was also very high (ICC = 0.85, 95% CI 0.70–0.97).

Summary of the validity guide to the meta-analysis of operative versus nonoperative treatment

The authors of the review specified explicit eligibility criteria. Their search strategy was comprehensive and reproducible. The primary studies had serious methodologic limitations. However, because these randomized trials represent the best available evidence, the results merit further consideration.

What are the results?

Were the results similar from study to study?

One aim of a systematic review, and in particular of a meta-analysis, is to increase the sensitivity of the primary studies to detect an effect by combining them as if all patients were part of a larger study. The validity of this assumption is confirmed if the magnitude of effect is similar across the range of patients, interventions and ways of measuring outcomes.

We have argued that the fundamental assumption is that across the range of patients, interventions and ways of measuring outcome, we anticipate more or less the same magnitude of effect. We have also noted that goals of increasing the precision of estimates of treatment effect, and the generalizability of results, provides reviewers with strong, legitimate reasons for selecting relatively wide eligibility criteria. Broad selection criteria, however, also increase the heterogeneity of the patient population, so systematic reviews often document important differences in patients, exposures, outcome measures and research methods from study to study. Fortunately, investigators can address this unsatisfactory situation by presenting their results in a way that allows clinicians to check the validity of the initial assumption. That is, did results prove similar from study to study? The remaining challenge is, then, to decide how similar is similar enough.

There are 3 criteria to consider when deciding whether the results are sufficiently similar to warrant a single estimate of treatment effect that applies across the populations, interventions and outcomes. First, how similar are the best estimates of the treatment effect (that is, the *point estimates*) from the individual studies. The more different they are, the more clinicians should question the decision to pool results across studies.

Second, to what extent do the CIs overlap? The greater the overlap between CIs of different studies, the more powerful is the rationale for pooling the results of those studies. The reviewers can also look at the point estimates for each individual study and determine if the CI around the pooled estimate includes each of the primary study point estimates.

Finally, reviewers can test the extent to which differences among the results of individual studies are greater than would be expected if all studies were measuring the same underlying effect and the observed differences were due to chance. The statistical analyses that are used to do this are called *tests of homogeneity*.[18] When the p value associated with the test of homogeneity is small (e.g., < 0.05), chance becomes an unlikely explanation for the observed differences in the size of the effect. Unfortunately, a higher p value (0.1 or even 0.3) does not necessarily rule out important heterogeneity. The reason is that when the number of studies and their sample sizes are small, the test of heterogeneity is not very powerful. Hence, large differences between the apparent magnitude of the treatment effect among the primary studies (i.e., the point estimates) dictate caution in interpreting the overall findings, even in the face of a nonsignificant test of homogeneity.[18] Conversely, if the differences in results across studies are not clinically important, then heterogeneity is of little concern, even if it is statistically significant.

Reviewers should try to explain between-study differences by looking for apparent explanations (i.e., sensitivity analyses). Heterogeneity in the current review of Achilles tendon ruptures may be attributable to differences in the surgical technique (e.g., simple v. Kessler v. Bunnell stitches), postoperative rehabilitation protocols (e.g., cast v. boot), methodologic features (methodologic quality scores), whether studies were full papers or abstracts, or whether studies were published in English or non-English-language journals.

What are the overall results of the review?

In clinical research, investigators collect data from individual patients. In systematic reviews, investigators collect data from individual studies rather than patients. Reviewers must also summarize these data and, increasingly, they are relying on quantitative methods to do so.

Simply comparing the number of positive studies to the number of negative studies is not an adequate way to summarize the results. With this sort of vote counting, large and small studies are given equal weight, and (unlikely as it may seem) one investigator may interpret a study as positive, while another investigator interprets the same study as negative. For example, a clinically important effect that is not statistically significant could be interpreted as positive with respect to clinical importance and negative with respect to statistical significance.[19] There is a tendency to overlook small but clinically important effects if studies with statistically nonsignificant (but potentially clinically important) results are counted as negative. Moreover, a reader cannot tell anything about the magnitude of an effect from a vote count, even when studies are appropriately classified using additional categories for studies with a positive or negative pattern.

Typically, meta-analysts weight studies according to their size, with larger studies receiving more weight.[1] Thus, the overall results represent a weighted average of the results of the individual studies. Occasionally studies are also given more or less weight depending on their quality; poorer-quality studies might be given a weight of 0 (excluded) either in the primary analysis or in a secondary analysis testing the extent to which different assumptions lead to different results (a sensitivity analysis). A reader should look to the overall results of a meta-analysis the same way one looks to the results of primary studies. In a systematic review of a question of therapy, one should look for the relative risk and relative risk reduction, or the odds. In reviews regarding diagnosis, one should look for summary estimates of the likelihood ratios.

Sometimes the outcome measures that are used in different studies are similar but not exactly the same. For example, different trials might measure functional status using different instruments. If the patients and the interventions are reasonably similar, it might still be worthwhile to estimate the average effect of the intervention on functional status. One way of doing this is to summarize the results of each study as an *effect size*. The effect size is the difference in outcomes between the intervention and control groups divided by the standard deviation. The effect size summarizes the results of each study in terms of the number of standard deviations of difference between the intervention and control groups (rather than using the conventional — and differing — units of measure). Investigators can then calculate a weighted average of effect sizes from studies that measured an outcome in different ways.

Readers are likely to find it difficult to interpret the clinical importance of an effect size (if the weighted average effect is one-half of a standard deviation, is this effect clinically trivial or is it large?). Once again, one should look for a presentation of the results that conveys their clinical relevance (e.g., by translating the summary effect size back into conventional units). For instance, if surgeons have become familiar with the significance of differences in functional outcome scores on a particular questionnaire, such as the Musculoskeletal Functional Assessment,[20] investigators can convert the effect size back into differences in score in this particular questionnaire.

Although it is generally desirable to have a quantitative summary of the results of a review, this is not always appropriate. In this case, investigators should present tables or graphs that summarize the results of the primary studies, and their conclusions should be cautious.

How precise were the results?

In the same way that it is possible to estimate the average effect across studies, it is possible to estimate a CI around that estimate, that is, a range of values with a specified probability (typically 95%) of including the true effect. The CI combines the average effect of the intervention, its standard deviation and the sample size to give us a range within which there is a 95% probability that the true effect falls. If the sample size is small the CI is wider; if the standard deviation is wide the CI is wide.

Results of the meta-analysis of operative versus nonoperative treatment of acute Achilles tendon ruptures

The mean age of patients in trials included in the current meta-analysis ranged from 36.5 to 41 years. Between 69% and 92% of the Achilles tendon ruptures were the result of sports-related injuries.

The authors tested the appropriateness of pooling data from 6 trials by examining trial-to-trial variability in results. When examining their primary outcome of repeat rupture rates, they found essentially similar point estimates, widely overlapping CIs and a nonsignificant test of heterogeneity ($p > 0.1$). However, they conducted a series of secondary analyses (sensitivity analyses) to explore their most questionable pooling decisions: pooling across publication status (published or unpublished), study quality score (< 50 v. ≥ 50) and language of publication.

In relation to the pooled analysis across all studies, operative treatment reduced the relative risk of repeat rupture by 68% (95% CI 29%–86%). However, operative fixation did significantly increase the risk of infection (relative risk 4.6%, 95% CI 1.2% –17.8%). Return to normal function and spontaneous complaints did not differ between the 2 groups.

Will the results help me in caring for my patients?

How can I best interpret the results to apply them to the care of my patients?

Although the pooled point estimate suggests a substantial reduction

in the relative risk of repeat rupture (68%) with surgery, the 95% CI ranges from 29% to 86%. If one accepts the point estimate as accurate, in patients at average risk for repeat rupture (say 30%), for every 10 patients treated with surgery, surgery would prevent 1 repeat rupture (number needed to treat = 1 ÷ 0.10 = 10).

The most obvious drawback to surgical repair is the increased risk of infection. In the current group of trials, there was a 4.7% rate of infection after surgery and no infections after conservative treatment. Therefore, for every 21 patients who receive surgical treatment, surgery would cause 1 wound infection (number needed to harm = 1 ÷ 0.047 = 21.2, 95% CI 17–59).

Were all clinically important outcomes considered?

Although it is a good idea to look for focused review articles because they are more likely to provide valid results, this does not mean that one should ignore outcomes that are not included in a review. For example, the potential benefits and harms of operative repair of Achilles tendon rupture include reduced risk of reoperation and increased risk of infection. Focused reviews of the evidence for individual outcomes are more likely to provide valid results, but a clinical decision requires consideration of all of them.[21] It is not unusual for systematic reviews to neglect the adverse effects of therapy. For example other outcomes of interest include magnitude and duration of pain, timing and extent of return to full function, and costs.

Are the benefits worth the costs and potential risks?

Finally, either explicitly or implicitly, when making recommendations to their patients surgeons must weigh the expected benefits against the potential harms and costs. Although this is most obvious for deciding whether to use a therapeutic or preventive intervention, providing patients with information about causes of disease or prognosis can also have benefits and harms. For example, a patient may benefit from decreased risk of infection with cast treatment of an Achilles tendon rupture at the cost (i.e., potential harm) of an increased risk of repeat rupture. A valid review article provides the best possible basis for quantifying the expected outcomes, but these outcomes still must be considered in the context of your patient's values and preferences about the expected outcomes of a specific decision.[2] For instance, one could recommend nonoperative management of Achilles tendon rupture to a patient who places a higher value on preventing infection and a lower value on preventing re-rupture.

Resolution of the scenario

The meta-analysis of operative versus nonoperative treatment of Achilles tendon ruptures meets most of the criteria for study validity, including explicit eligibility criteria, a comprehensive search strategy, and assessment and reproducibility of study validity.[2] The authors found a very large benefit of operative repair on re-rupture rates at the cost of greater infection risk. Furthermore, pooling of study results seems justified by the nonsignificant tests of heterogeneity, reasonable similarity of results (point estimates) and widely overlapping CIs around those point estimates. On the other hand, the quality of studies was relatively poor, including a failure to conceal randomization in all studies. Our interpretation is that the magnitude of the effect is sufficiently large that, despite the limitations in study quality, the inference that operative repair provides substantially lower repeat rupture rates in patients with Achilles tendon ruptures is secure. Thus, surgeons who manage patients with Achilles tendon ruptures similar to those presented in this meta-analysis (younger, athletic, acute ruptures) can reassure them that current evidence favours operative treatment. When patients seem different from those included in a meta-analysis, clinicians should consider whether they are really so different that the results cannot be applied to their patients.

In every situation, physicians should try to find out the general decisional preferences of their patients, such as the favoured decision-making model, the amount of information desired and their ideal degree of involvement in deliberation and decision-making. Clinicians should also be aware that their patients' preferences might vary with the nature of the decision, the choices and the outcomes. While researchers find answers to these questions, practising clinicians should try their best to make sure that important decisions remain as consistent as possible with the values and preferences of informed patients, the women and men who will live (and die) with the outcomes.

The current increase in the number of small randomized trials in the field of orthopedic surgery provides a strong argument in favour of meta-analysis. However, it remains essential that those who are planning future meta-analyses adhere to accepted methodologies and provide the best available evidence to address sharply defined clinical questions.[4] Although the quality of the primary studies will always be the major limiting factor in drawing valid conclusions, the quality of the meta-analysis is also important in ensuring that the pooling of these results is as valid and free of bias as possible.

Competing interests: None declared.

References

1. Oxman A, Cook DJ, Guyatt GH. User's guide to the medical literature: how to use an overview. *JAMA* 1994;272:1367-71.
2. Guyatt GH, Rennie D, editors. *Users' guides to the medical literature: a manual for evidence-based clinical practice.* Chicago: AMA Press; 2001.
3. Bhandari M, Guyatt GH, Siddiqui F, Morrow F, Busse J, Leighton RK, et al.

Treatment of acute Achilles tendon ruptures: a systematic overview and meta-analysis. *Clin Orthop* 2002;400:190-200.
4. Bhandari M, Morrow F, Kulkarni A, Tornetta P 3rd. Meta-analyses in orthopædic surgery: a systematic review of their methodologies. *J Bone Joint Surg Am* 2001;83: 15-24.
5. Dickerson K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-9.
6. Dickerson K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Control Clin Trials* 1987;8:343-53.
7. Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. *Mayo Clin Proc* 2000;75:1284-8.
8. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe A. Incorporating variation in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255-65.
9. Moher D, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, et al. Does the quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
10. Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996;156:661-6.
11. Cook DJ, Sackett DL, Spitzer WO. Methodological guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995; 48:167-71.
12. Cook DJ, Mulrow CD, Haynes RB. Synthesis of best evidence for clinical decisions. In: Mulrow C, Cook DJ, editors. *Systematic reviews.* Philadelphia: American College of Physicians; 1998. p. 5-15.
13. Turner JA, Ersek M, Herron L, Deyo R. Surgery for lumbar spinal stenosis: attempted meta-analysis of the literature. *Spine* 1992;17:1-8.
14. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials* 1996;17:1-12.
15. Clark H, Wells G, Huet C, McAlister F, Salmi LR, Fergusson D, et al. Assessing the quality of randomized trials: reliability of the Jadad scale. *Control Clin Trials* 1999;20:448-52.
16. Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics* 1975;31:651-9.
17. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995;345:772-6.
18. Cooper RM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. *Pyschol Bull* 1980;87:442-9.
19. Breslow NE, Day DE. Combination of results from a series of 2×2 tables; control of confounding. In: *Statistical methods in cancer research: the analysis of case–control studies.* Vol. 1. Lyon (France): International Agency for Research on Cancer; 1980. p. 136-46.
20. Engelberg R, Martin D, Agel J. Musculoskeletal function assessment instrument: criterion and construct validity. *J Orthop Res* 1996;14:182-92.
21. Colton C. Statistical correctness. *J Orthop Trauma* 2000;8:527-8.

CMA CENTRE FOR
PHYSICIAN HEALTH AND WELL-BEING
HEALTHY PHYSICIANS, VIBRANT PROFESSION

If you are feeling overwhelmed by professional and personal demands, you are not alone. Research has found that 45.7% of Canadian physicians are in an advanced state of burnout*.

CMA believes in a culture of medicine that values the health of its practitioners and does not stigmatize those who seek personal assistance. CMA's new Centre for Physician Health and Well-being provides national leadership and advocacy on issues affecting the health and morale of Canadian physicians.

To learn more about the work of the centre or to access information and resources, visit the centre's new Web section at ***cma.ca*** or call **1 877 CMA-4-YOU** (1 877 262-4968).

The centre supports and complements the work of the provincial and territorial physician health programs but does not provide individual assessment, support or treatment advice. Contact numbers for the programs are available at ***cma.ca***.

*2003 CMA *Physician Resource Questionnaire*.

ASSOCIATION MÉDICALE CANADIENNE
CANADIAN MEDICAL ASSOCIATION
cma.ca