

META-ANALYSES IN ORTHOPAEDIC SURGERY

A SYSTEMATIC REVIEW OF THEIR METHODOLOGIES

BY MOHIT BHANDARI, MD, MSC(EPID), FARRAH MORROW, BSC,
ABHAYA V. KULKARNI, MD, MSC(EPID), AND PAUL TORNETTA III, MD

Investigation performed at the Department of Clinical Epidemiology and Biostatistics, and the Department of Surgery, McMaster University, Hamilton, Ontario, Canada; the Department of Surgery, the University of Toronto, Toronto, Ontario, Canada; and the Department of Orthopaedic Surgery, Boston Medical Center, Boston, Massachusetts

Background: The number and quality of well-designed scientific studies in the orthopaedic literature are limited. The purpose of this review was to determine the methodological qualities of published meta-analyses on orthopaedic-surgery-related topics.

Methods: A systematic review of meta-analyses was conducted. A search of the Medline database provided lists of meta-analyses in orthopaedics published from 1969 to 1999. Extensive manual searches of major orthopaedic journals, bibliographies of major orthopaedic texts, and personal files identified additional studies. Of 601 studies identified, forty met the criteria for eligibility. Two investigators each assessed the quality of the studies under blinded conditions, and they abstracted relevant data.

Results: More than 50% of the meta-analyses included in this review were published after 1994. We found that 88% had methodological flaws that could limit their validity. The main deficiency was a lack of information on the methods used to retrieve and assess the validity of the primary studies. Regression analysis revealed that meta-analyses authored in affiliation with an epidemiology department and those published in nonsurgical journals were associated with higher scores for quality. Meta-analyses with lower scores for quality tended to report positive findings. The meta-analyses that focused upon fracture treatment and degenerative disease (hip, knee, or spine) had significantly lower mean quality scores than did meta-analyses that examined thrombosis prevention and diagnostic tests ($p < 0.05$).

Conclusions: The majority of meta-analyses on orthopaedic-surgery-related topics have methodological limitations. Limitation of bias and improvement in the validity of the meta-analyses can be achieved by adherence to strict scientific methodology. However, the ultimate quality of a meta-analysis depends on the quality of the primary studies on which it is based. A meta-analysis is most persuasive when data from high-quality randomized trials are pooled.

Systematic literature reviews are useful for synthesizing the results of multiple primary investigations with use of strategies to limit bias and random error¹⁻⁴. A quantitative systematic review, or meta-analysis, is a review in which statistical methods are used to combine the results of two or more studies. All systematic reviews are retrospective and observational. Therefore, they are subject to systematic and random error¹. Thus, the quality of a systematic review, and accordingly its validity, is dependent upon the scientific methods that have been used to minimize error and bias.

A well-conducted meta-analysis is invaluable for surgeons since it is unusual for single studies to provide definitive answers to clinical questions. Moreover, a well-conducted quantitative review may resolve discrepancies between studies with conflicting results³. Guiding principles in the conduct of meta-analyses include use of a specific health-care question, use of a comprehensive search strategy, assessment of the reproducibility of study selection, assessment of the

study validity, evaluation of heterogeneity (differences in effect across studies), inclusion of all relevant and clinically useful measures of treatment effect, and tests of the robustness of the results relative to features of the primary studies (sensitivity analysis)⁵.

The popularity of systematic reviews has resulted in a 500-fold increase in the number of published meta-analyses in the past decade^{6,7}. Unfortunately, the increased use of this research tool has not always been accompanied by an appreciation of the importance of scientific methodology. Without such methodology, meta-analyses can produce inaccurate, biased, and misleading estimates of the effectiveness of a particular surgical or medical intervention⁸⁻¹¹, which may have serious implications in terms of the quality and cost of patient care. In an effort to improve the quality of reporting, many authors have described sources of bias in the conduct of meta-analyses and an instrument has been developed to grade the scientific quality of such studies^{12,13}.

Orthopaedic surgeons must be aware of the limitations and risks of meta-analyses and must strive to limit bias. Similarly, journal editors must ensure that the meta-analyses that they publish adhere to accepted scientific methodology.

Given the increased use of meta-analyses in orthopaedics, we performed a systematic review of the literature to identify meta-analyses on orthopaedic-surgery-related topics. Our purpose was threefold: (1) to assess the evolution of the scientific quality of meta-analytic research in orthopaedic surgery, (2) to evaluate potential prognostic variables that are associated with the quality of a meta-analysis, and (3) to assess our ability to reliably score meta-analyses with use of a quality index.

Materials and Methods

Eligibility Criteria

In order to be included in our study, each meta-analysis had to meet the following criteria: (1) the study had to be described as a "meta-analysis," or, if not, statistical pooling of the results had to have been conducted; (2) primary studies included in the meta-analysis had to have direct relevance to the practice of orthopaedic surgery (that is, they had to involve subjects such as the prevention of thromboembolism, arthroplasty, the spine, trauma, pediatrics, sports medicine, and the upper extremity); and (3) the study had to have been published or accepted for publication.

Study Identification

A computerized Medline search was conducted for the period from 1969 to 1999 with use of the following terms and Boolean operators: "meta-analysis" OR "meta-anal: (text-word)" OR "quantitativ: review:" OR "quantitativ: overview:" AND "orthopaedics" OR "spine" OR "fractures" OR "arthroplasty" OR "pediatrics" OR "hip" OR "knee." The Cochrane Database for Systematic Reviews was also searched to identify any additional studies that may have been published in the orthopaedic literature. The bibliography of each meta-analysis was reviewed by two of us for additional relevant studies. In addition to bibliographic searches, three of us manually searched the last five years of issues published by five major orthopaedic journals (*The Journal of Bone and Joint Surgery: American and British Volumes*, *Clinical Orthopaedics and Related Research*, *Spine*, and *Acta Orthopaedica Scandinavica*). This sample of journals was thought to adequately represent the general sources of information used by most orthopaedic surgeons in North America and Europe. The proceedings for selected specialist meetings (the American Academy of Orthopaedic Surgeons, the Orthopaedic Trauma Association, and the Canadian Orthopaedic Association) and textbooks also were searched manually. Finally, content experts (that is, those with an interest in meta-analysis) were asked to identify additional studies that may have been missed by our search strategy. Any relevant meta-analysis identified from the proceedings was rechecked to ensure that it had been accepted for publication. Whenever a meta-analysis appeared to be eligible by its title alone, the complete article was retrieved.

Assessment of Methodological Quality

Each eligible meta-analysis was independently reviewed by two of us, surgeons with training in epidemiology, for methodological quality. Both authors were blinded to all specific meta-analytic information except the methods section. The Oxman and Guyatt index was utilized to score the methodology of the meta-analyses^{12,13} (Appendix). Briefly, this index contains ten items, the last of which is an overall interpretation of the study that rates it as one that contains minimal flaws, minor flaws, major flaws, or extensive flaws. This index was designed to evaluate the scientific quality (that is, adherence to scientific principles) of research overviews, including meta-analyses, published in the literature. It is not intended to measure literary quality, importance, relevance, originality, or other attributes of overviews. No specific training in the use of this instrument was obtained; however, the guidelines for scoring with use of this index were carefully reviewed by two of us. Any discrepancies in scoring between the reviewers were resolved by consensus. Additional information from the meta-analyses was occasionally requested to resolve disagreements in scoring.

Data Extraction

For each of the eligible meta-analyses, the relevant data were abstracted by one of us and were rechecked for accuracy by another. Specifically, we abstracted the following information: (1) the affiliation of the first author (surgical department, department of epidemiology, or medical), (2) citation of a degree (MSc, PhD, or MPH) in epidemiology or biostatistics as a surrogate of affiliation in an epidemiology department for any author (yes or no), (3) the name of the journal, (4) the year of publication, (5) the number of primary studies included in the review, (6) the total number of cases in the meta-analysis, (7) the type of study (a comparison of interventions, a report on a single intervention, an assessment of a diagnostic tool, or other), (8) the name of the intervention, (9) the category of the intervention (fracture treatment, treatment of degenerative disease of the spine or joints, evaluation of a diagnostic test, thrombosis prevention, or miscellaneous), (10) the region (spine, hip, knee, femur, tibia, or other), (11) financial support (none stated, non-peer-reviewed grant, government grant, charity, or internal funds), (12) the design of the primary studies (randomized double-blind, quasi-randomized, observational, or mixed), (13) the description of the methods used to identify the primary studies, (14) the rationale for statistical pooling (described or not described), (15) the method of statistical pooling, and (16) the direction of the results (positive if the findings of the meta-analysis were significant or negative if no significant differences between variables were reported).

We arbitrarily defined five major categories of meta-analyses: fracture treatment, treatment of degenerative disease of the spine or joints, thrombosis prevention, evaluation of a diagnostic test, and miscellaneous.

Assessment of Reviewer Agreement

The kappa statistic, a measure of chance-corrected agreement, provided most estimates of agreement between reviewers for

titles and methods sections of potentially relevant meta-analyses. Studies by Fleiss¹⁴ and by Donner and Klar¹⁵ provided persuasive arguments favoring the use of this statistic over other measures of agreement. For variables with more than two categories, we used weighted kappa with quadratic weights, which yields values identical to intraclass correlation coefficients. We chose an *a priori* criterion of kappa = 0.65 or greater for adequate agreement¹⁶.

Data Analysis

Prior to analyzing the data, we developed hypotheses regarding the association between the overall quality and the results of the meta-analysis. Specifically, we hypothesized that meta-analyses with lower quality scores would be more likely to produce a positive result. The extents to which the meta-analyses fulfilled each item on the Oxman and Guyatt index were compared with use

of the chi-square test. Moreover, relationships between the overall quality score and the results of the meta-analyses were evaluated with use of the chi-square test. The mean quality scores of the five categories of meta-analyses were compared with analysis of variance. A univariable regression analysis was used to identify the important factors influencing the methodological quality of a meta-analysis. We examined the effect of a number of independent variables (affiliation with an epidemiology department, type of journal [surgical or nonsurgical], date of publication, financial support, design of the primary studies, and category of intervention) on the dependent variable (an overall quality score of 1 to 7 points). The variables that revealed a significant association with the quality of the meta-analysis in the univariable analysis were used in a multiple regression model. The results from this analysis were reported as coefficients

with 95% confidence intervals. For all statistical analyses, a p value of less than 0.05 was considered significant.

Results

Study Identification

Six hundred and one potentially relevant citations were identified: 577 (96%) were identified from computerized searches; nineteen (3%), from reviews of bibliographies; three (less than 1%), from content experts; and two (less than 1%), from reviews of proceedings. The application of the criteria for eligibility eliminated 410 studies that were not meta-analyses, 150 studies that did not focus on the field of orthopaedic surgery, and one study in a proceeding of an annual orthopaedic meeting that was not accepted for publication. Thus, forty meta-analyses met all of the inclusion criteria: thirty-one (78%) were identified from computerized database searches; six (15%), from bibliography searches; three (7.5%), from content experts; and one (2.5%) was from a search of proceedings. Agreement between reviewers with respect to the eligibility of the meta-analyses was substantial (kappa = 0.75).

Characteristics of the Meta-Analyses

We were unable to identify any meta-analyses published prior to 1984. The number of meta-analyses in orthopaedic surgery increased from 1984 to 1999, and twenty-six (65%) of the forty studies were published, or had been accepted for publication, within the last five years (Fig. 1). The forty meta-analyses were published in twenty-nine different journals, 50% of which were surgical journals (Table I). Fourteen of the first authors were affiliated with surgical departments, whereas twenty-six were not. In nineteen (48%) of the forty meta-analyses, at least one author had cited training in epidemiology (MSc, MPH, or PhD). The meta-analyses pooled a mean of forty-three primary studies (range, two to 130 studies). In twenty-three meta-analyses (58%), the total number of patients pooled was reported, and this number ranged from fifty-five to 13,478. In most (thirty; 75%) of the meta-analyses, two or more interventions were compared. Five cate-

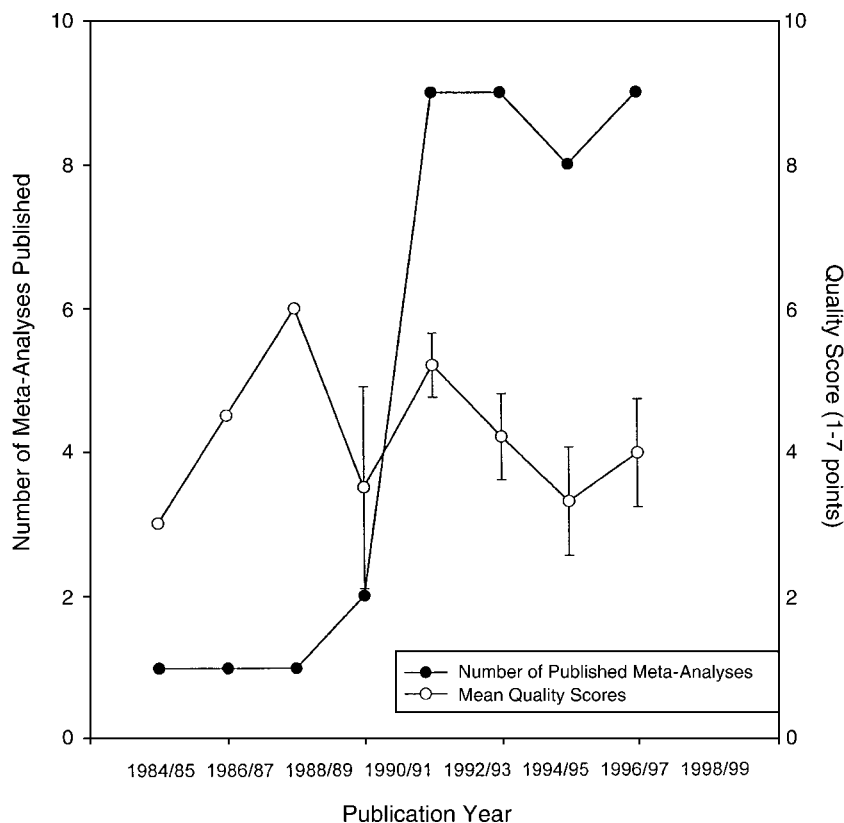


Fig. 1

Graph showing the number and quality of published meta-analyses over time. A dramatic increase occurred in the number of meta-analyses published during the last ten years, but the mean quality scores for these meta-analyses did not change. The data for quality scores are given as the mean and the standard error of the mean.

TABLE I Number of Studies per Journal

Journal	No. of Meta-Analyses
Acta Orthopaedica Scandinavica	3
AJR: American Journal of Roentgenology	1
American Journal of Emergency Medicine	1
American Journal of Medicine	1
American Surgeon	1
Annales Chirurgiae et Gynaecologiae	1
Annals of Internal Medicine	1
Archives of Internal Medicine	1
Australian and New Zealand Journal of Medicine	1
BMJ: British Medical Journal	1
Chest	1
Clinical Journal of Sports Medicine	1
Clinical Orthopaedics and Related Research	1
European Spine Journal	1
Haemostasis	1
International Orthopaedics	1
JAMA: Journal of the American Medical Association	2
Journal of Arthroplasty	2
Journal of Bone and Joint Surgery (American)	2
Journal of General Internal Medicine	2
Journal of Hand Surgery	1
Journal of Orthopaedic Trauma	1
Lancet	1
New England Journal of Medicine	1
Orthopedics	1
Physical Therapy	1
Seminars in Thrombosis and Hemostasis	1
Spine	5
Thrombosis and Haemostasis	2
Total	40

gories of meta-analyses were identified: (1) fracture treatment (eleven studies; 28%), (2) treatment of degenerative joint disease (eleven), (3) prevention of deep venous thrombosis in orthopaedic patients (eleven), (4) evaluation of a diagnostic test (five; 13%), and (5) miscellaneous (two; 5%). The primary studies included in the forty meta-analyses were most often randomized (in eighteen meta-analyses; 45%) or a mix of randomized and observational studies (in sixteen meta-analyses; 40%). Exclusive inclusion of observational primary studies or the inclusion of quasi-randomized primary studies occurred less frequently (in three meta-analyses [7.5%] each).

Methods of Statistical Pooling

Thirteen (33%) of the meta-analyses incorporated simple summation as the pooling method (Table II). Moreover, only seventeen meta-analyses (43%) evaluated the appropriateness of the pooling of the primary studies with a test of heterogeneity.

Scientific Quality (Item 10)

The level of agreement between reviewers in assessing the quality of the meta-analyses was substantial ($\kappa = 0.71$; 95% confidence interval, 0.41 to 0.85). The mean score (and standard error of the mean) for the overall quality (item 10) of the forty meta-analyses was 4.2 ± 1.78 points. We found that 88% (thirty-five) of the meta-analyses were observed to have methodological flaws, and thirteen (37%) of the thirty-five were considered to have major-to-extensive flaws. The main deficiency was the lack of information on the methods used to retrieve and assess the validity of the primary studies. Figure 2 illustrates the distribution of the meta-analyses with respect to the quality scores, which ranged from 1 to 7 points, and Figure 3 demonstrates the mean quality scores for each category of meta-analysis. Significant differences were observed among the categories with respect to the mean quality scores. The meta-analyses that focused on fracture treatment and degenerative disease (hip, knee, or spine) had significantly lower mean quality scores than did the meta-analyses involving thrombosis prevention or the evaluation of diagnostic tests ($p < 0.05$). Moreover, while the number of meta-analyses published from 1984 to 1999 increased, the mean quality score did not change over time (Fig. 1). When the overall quality score was compared with the results of the meta-analysis, it was found that higher-quality meta-analyses were less likely to produce a "positive" conclusion (Table III).

Association with Scientific Quality (Item 10)

We performed a multiple regression analysis to determine the association between potential prognostic variables (affiliation of an author with an epidemiology department, type of journal, date of publication, financial support, design of the primary studies, and category of the intervention) and the quality of the study (Table IV). Univariable analysis, in which each variable is examined independently of the others, revealed that the affiliation of an author with an epidemiology department, type of journal, category of the intervention, and design of the primary studies were significantly associated with the quality of a meta-analysis. However, only the journal type was shown to have a significant association with scientific quality on multivariable analysis, in which all of the variables were considered together. Affiliation with an epidemiology department approached significance in predicting overall scientific quality. Overall, the four variables accounted for more than 47% ($r = 0.68$, $p < 0.01$) of the total variation in the dependent variable, meta-analysis quality.

Individual Item Scores (Items 1 through 9)

The kappa statistic for interobserver agreement between the reviewers with respect to the scoring of items 1 through 9 ranged from 0.48 to 0.84. Disagreements were generally related to an oversight by one of the reviewers and were easily resolved. The percentage of meta-analyses that received a full score for each item ranged from 43% to 83% (Table V). A full score meant that, for a particular item, the reviewer answered "yes" (Appendix). Alternatively, if the meta-analysis did not fulfill the item, or if the reviewer could not ascertain the information

TABLE II Methods of Pooling in Forty Meta-Analyses

Method	No. (%) of Studies
Random effects	7 (18)
Odds ratios or Mantel-Haenszel test	7 (18)
Weighted means	5 (13)
Simple addition	13 (33)
Weighted proportions	4 (10)
Log-rank test	1 (3)
Receiver operating curves	2 (5)
Not reported	1 (3)

from the meta-analysis, it was labeled as “no” or “can’t tell.” Items 4, 5, and 6, which focused on bias in the selection of primary studies and validity assessments, were fulfilled by less than 50% of the meta-analyses. Meta-analyses with negative or uncertain conclusions were more likely to have avoided bias in the selection of primary studies (item 4, $p = 0.008$) and to have appropriately combined their results (item 8, $p = 0.01$).

Discussion

We conducted a systematic review of the literature to identify meta-analyses that focused on issues relevant to orthopaedic surgery. The relatively large number of meta-analyses across a variety of journals indicates the widespread influence of orthopaedic surgery in both surgery and medicine. We limited bias in the selection of meta-analyses by conducting the process in duplicate and by scoring the quality of the meta-analyses in an independent, blinded fashion. Given the meta-analyses assessed in this systematic review, there is evidence to suggest that most meta-analyses in orthopaedic surgery are limited by methodological flaws.

Strengths of Inference from Meta-Analysis

Organizing information about surgical treatments in a useful way is a major challenge for all who are involved in health care. Much of this information comes from observational research methods. To reduce bias, the randomized, controlled trial has been developed as a more valid method for comparing treatment effects¹⁷⁻¹⁹. However, even randomized trials may not answer specific questions because of weaknesses in their design or, more commonly, because they are not of adequate statistical power to detect a clinically important treatment effect. Meta-analyses of randomized trials combine data from different studies that address a similar question, with use of accepted statistical methods, to obtain more reliable estimates of treatment effects^{7,20}. The methodology of meta-analyses differs from that of narrative literature reviews in a number of ways: (1) meta-analyses often address a focused clinical question, (2) they involve a comprehensive and explicit search strategy, (3) the selection of articles is based upon a set of eligibility criteria, (4) the validity of the included studies is assessed, and (5) a quantitative summary of the data (or a meta-analysis) is conducted¹. These are, in essence, the same steps that we followed in our systematic review of meta-analyses.

The strength of inference from a meta-analysis is only

as good as the quality of the primary studies and the scientific rigor with which the meta-analysis was conducted. Meta-analyses that pool data from nonrandomized trials are subject to all of the limitations of the primary studies. Thus, in effect, combining the results of nonrandomized studies may result in a grossly biased pooled estimate of effect. However, prospective cohort studies, when rigorously conducted, can provide useful information when randomized trials are not feasible or available. Since the majority of the orthopaedic literature is derived from observational (nonrandomized and retrospective) studies, the inferences from meta-analyses that pool such results may be limited.

The effect of the quality of the primary studies included in meta-analyses has been well reported^{21,22}. Moher et al. found that lower-quality primary studies tend to demonstrate larger estimates of treatment effect than do those of higher quality²². Only 43% of the meta-analyses included in the current systematic review pooled exclusively randomized trials. In most cases, a mix of randomized trials and observational primary studies were pooled to obtain an overall treatment effect. Thus, adherence to a rigorous methodology in the conduct of a meta-analysis is important to limit additional bias, especially when the quality of the primary studies is questionable.

Appropriateness of Statistical Pooling in Meta-Analysis

As the term meta-analysis implies a systematic review with statistical pooling of results, it was interesting that less than half of the forty meta-analyses reported the rationale for pooling. Statistical tests cannot compensate for lack of common sense, clinical acumen, and biological plausibility in the design of the protocol of a meta-analysis²³. Combining poor-quality data, overly biased data, or data that do not make sense can easily produce misleading results. The homogeneity of

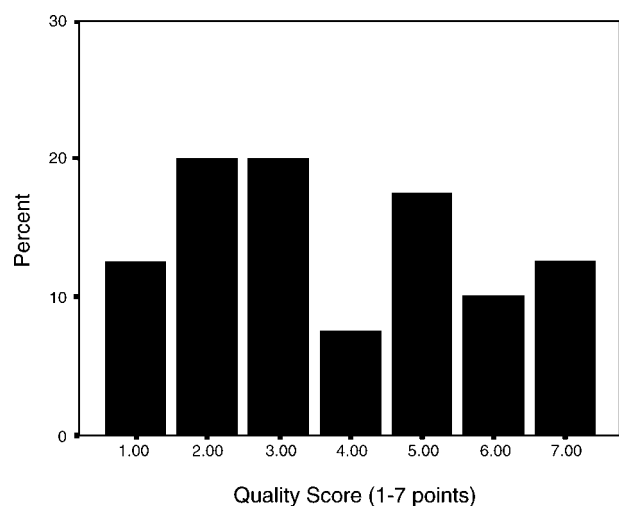


Fig. 2

Graph showing the percentage of meta-analyses by overall quality score. A score of 1 to 4 points indicated extensive or major flaws; 5 or 6 points, minor flaws; and 7 points, minimal flaws. Approximately half of all of the eligible meta-analyses were found to have major methodological flaws.

TABLE III Quality of Meta-Analyses and Conclusions

Overall Score* (points)	Study Conclusions	
	Positive (N = 28)	Negative or Uncertain† (N = 12)
1	1	1
2	5	1
3	4	1
4	5	0
5	4	2
6	5	5
7	4	2

*The overall score is from item 10 of the Oxman and Guyatt index^{12,13}. †A negative conclusion indicates that no difference between interventions was reported or that the findings were uncertain.

data from different primary studies can be assessed with use of a statistical test of homogeneity²⁴. Increasing confidence in the homogeneity of the results of primary studies comes from a similarity of their point estimates and widely overlapping confidence intervals. For example, if several small randomized trials all appear to favor one technique over another, and the confidence intervals are widely overlapping, it is probable that the studies in question are homogeneous. Statistical tests, however, do not replace “clinical sense.” Thus, pooling is reasonable if one would expect the same treatment effect, more or less, across the range of populations, interventions, and methodologies of the primary studies.

The most common method of pooling in our sample was simple addition. Some investigators have advocated against simple collapsing of data from multiple primary studies into one two-by-two table as the results may be misleading²⁵. As an example, assume that two primary studies that compared the risks of nonunion associated with two interventions (A and B) demonstrated widely divergent results (for example, a relative risk of 3.0 compared with a relative risk of 0.3). If the results of these primary studies are collapsed together into one two-by-

two table, the final result may suggest that there is no difference between the two treatments (relative risk = 1.0). However, this is very misleading given the fact that one study reported an increased risk of nonunion with intervention A (relative risk = 3.0) and the other reported a decreased risk of nonunion with intervention A (relative risk = 0.3).

Scientific Quality of Meta-Analyses

The use of meta-analyses to answer clinically important questions in orthopaedics has increased dramatically in the last several years^{6,7}. However, despite the availability of guidelines to limit bias²⁻⁵, most of the meta-analyses included in this systematic review had methodological deficiencies that may limit the validity of their conclusions. Only 10% of the meta-analyses satisfied all of the categories in the Oxman and Guyatt quality index, and 13% were given the lowest possible score. These findings are consistent with those of Jadad and McQuay, who reported that seventy-two (90%) of eighty meta-analyses of analgesic interventions were flawed²⁶.

It was interesting, but not unexpected, that meta-analyses focusing on trauma (fracture treatment) and degenerative disease (hip, knee, or spine) scored lower in quality than did meta-analyses of studies on thrombosis prevention and evaluation of a diagnostic test. We explored the potential reasons for these differences in quality. The meta-analyses on thrombosis prevention generally pooled data from randomized trials, were generally authored by nonsurgeons with some training in epidemiology, and were published primarily in nonsurgical journals. Moreover, limiting bias in drug trials (that is, those for the prevention of thrombosis) is less difficult than it is in trials in which surgical interventions are compared. Thus, the higher scores for quality likely reflected the fact that drug trials were most often published in nonsurgical journals. The few meta-analyses focusing on fracture treatment and degenerative disease, which scored high, either pooled data only from randomized trials or were authored by at least one person with training in epidemiology. Meta-analyses in the Cochrane Database of Systematic Reviews have been shown to be higher in scientific quality

TABLE IV Factors Associated with Study Quality

Variable	Univariable Analysis			Multivariable Analysis		
	Coefficient*	Standard Error of Mean	P Value	Coefficient*	Standard Error of Mean	P Value
Affiliation with epidemiology department	1.92 (0.85, 2.98)	0.52	0.001	1.04 (-2.8, 3.8)	0.29	0.07
Date of publication	-6.5 × 10 ⁻³ (-0.19, 0.18)	0.09	0.94	—	—	—
Journal type	2.28 (1.29, 3.21)	0.49	<0.001	1.86 (0.02, 3.7)	0.91	0.05
Category	0.65 (0.23, 1.08)	0.21	0.004	0.08 (-0.48, 0.64)	0.27	0.77
Funding	0.30	0.22	0.19	—	—	—
Design of primary studies	-0.46 (-8.77, -0.04)	0.21	0.03	0.10 (-0.44, 0.65)	0.27	0.71

*The values are given as the beta coefficient, with the 95 percent confidence interval in parentheses.

than were meta-analyses published in other sources²⁷.

It was not surprising that the regression analysis showed the most important predictors of the quality of a meta-analysis to be affiliation with an epidemiology department and journal type. The design of the primary studies and the category of the intervention were significantly associated with quality on univariable analysis but not on multivariable analysis. This likely was due to the fact that there was an association between the type of journal and the design of the primary studies included in the meta-analyses. Primarily nonsurgical journals (*Lancet*, *New England Journal of Medicine*, *Journal of the American Medical Association*, and *Archives of Internal Medicine*) tended to publish meta-analyses that pooled data from randomized trials, whereas surgical journals tended to publish meta-analyses that pooled data from a mix of observational primary studies and randomized trials.

There was a trend toward negative or uncertain conclusions in the meta-analyses with higher scores for quality. This observation is consistent with that of Jadad and McQuay in their review of meta-analyses of studies on analgesic interventions²⁶.

During the validation process, Oxman et al. reported that their index could be scored consistently by trained assessors^{12,13}. Our findings suggest that training may not be necessary to obtain reliable scores between observers. Kappa values ranged between 0.48 and 0.84 in the assessment of items 1 through 9 when the index was used by surgeons with training in epidemiology. This observation is consistent with that of Jadad and McQuay, who reported consistent scores among assessors who were not trained in scoring meta-analyses that evaluated analgesic interventions²⁶. Since this index is simple and has been extensively developed, we recommend it as a tool with which to evaluate the scientific methodology of systematic reviews, including meta-analyses, in orthopaedic surgery.

Limitations of the Current Study

While a comprehensive search of the literature was performed, there is a possibility that potentially relevant meta-analyses were omitted for the following reasons: (1) only meta-analyses published in the English-language literature were identified, (2) only published meta-analyses (or those accepted for publication) were retrieved, and (3) there is a publication bias

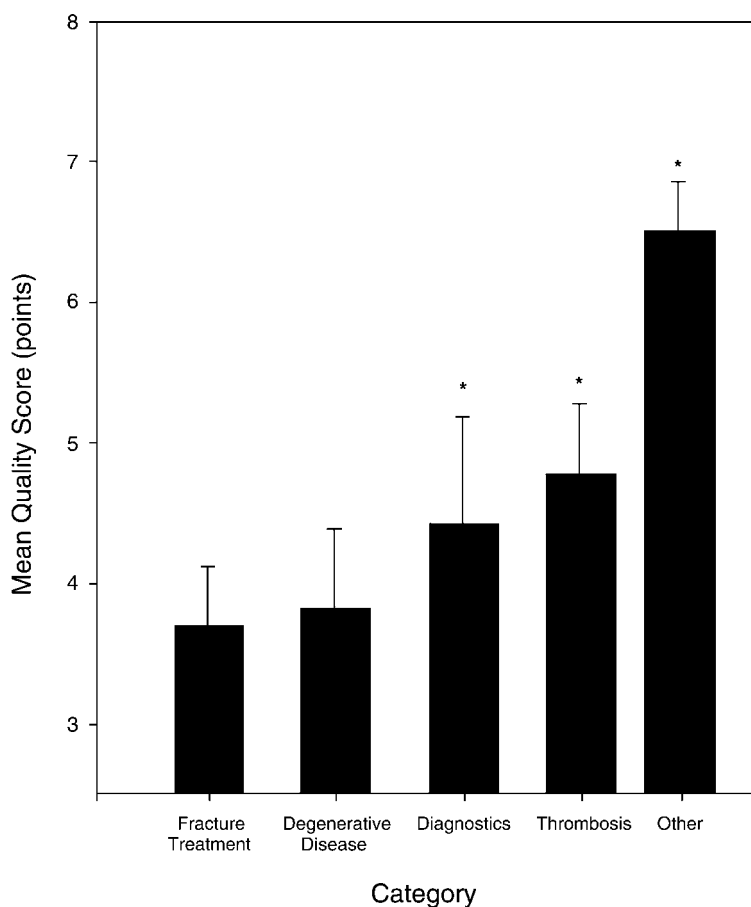


Fig. 3

Graph showing the mean quality scores among categories of meta-analyses. The quality scores for meta-analyses that focused upon fracture treatment and degenerative joint disease (hip, knee, or spine) were found to be significantly lower than those for the meta-analyses that examined thrombosis prevention and diagnostic tests ($p < 0.05$). An asterisk indicates that the p value was less than 0.05 when the category was compared with either the fracture-treatment or the degenerative-disease category. The data are expressed as the mean and the standard error of the mean.

TABLE V Quality Scores (Items 1 through 9) and Study Conclusions*

Item	Interobserver Agreement	Total No. (%) of Meta-Analyses Fulfilling Each Criterion	No. (%) of Meta-Analyses Fulfilling Each Criterion		P Value
			Positive Conclusion* (N = 28)	Negative or Uncertain Conclusion† (N = 12)	
1	0.75	33 (83)	22 (79)	11 (92)	0.30
2	0.70	29 (73)	21 (75)	8 (67)	0.64
3	0.78	31 (78)	21 (75)	10 (83)	0.53
4	0.48	17 (43)	8 (29)	9 (75)	0.008
5	0.84	19 (48)	13 (46)	6 (50)	0.75
6	0.58	18 (45)	13 (46)	5 (42)	0.96
7	0.65	28 (70)	18 (64)	10 (83)	0.20
8	0.55	25 (63)	14 (50)	11 (92)	0.01
9	0.48	26 (65)	18 (64)	8 (67)	0.82

*The percentages were derived by dividing the number of studies fulfilling each criterion by the total number of studies with a positive conclusion. †The percentages were derived by dividing the number of studies fulfilling each criterion by the total number of studies with a negative conclusion.

against meta-analyses that do not have “significant” findings. However, the meta-analyses in our study likely are a representative sample of the total number of meta-analyses in this field that would be readily accessible to most orthopaedic surgeons. An additional limitation, as with most meta-analyses, is the ability to extract data from the primary studies.

In our systematic review, we used the declaration of graduate degrees and departmental affiliations as the source for determining the epidemiological training of the authors. Although this represented the most feasible method, it potentially missed those with training in epidemiology who do not have a primary appointment to a department of epidemiology or a department of public health. Similarly, the authors may not have declared all of the sources of funding for their study.

Future Considerations

The current “respectability” of the methodology may, in part, have facilitated the increasing use of meta-analyses. Although it looks easy in concept, the production of a high-quality systematic review is extremely demanding. Sometimes, individuals with only limited knowledge of a treatment, the biology of a disease, or the clinical circumstances relating to a specific question may perform a meta-analysis that has little clinical relevance. Therefore, while the statistical methodology of meta-analysis has advanced greatly over the past decade, there also has been an increase in meta-analyses performed with use of suboptimal methods.

The methodological deficiencies identified in most meta-analyses can be easily avoided in the future by ensuring that investigators address the issues raised by each item in the Oxman and Guyatt scoring system^{12,13} (Appendix). The perpetuation of methodological flaws identified in this systematic review will not only devalue meta-analysis as a tool to facilitate decision-making processes but will provide justification for those who consider meta-analyses to be statistical trickery. The current increase in the number of small randomized trials in the field of orthopaedic surgery provides a strong argument

in favor 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. While 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. However, given the importance of the quality of the primary studies, the issue of whether one should even consider performing meta-analyses of nonrandomized, observational studies is controversial²⁸. The results obtained from such meta-analyses need to be approached with great caution and with an awareness of the potential limitations of the primary study designs. ■

NOTE: The authors are grateful to Dr. J. Hirsh, Director of the Hamilton Civic Hospitals Research Centre, for his suggestions regarding the manuscript.

Mohit Bhandari, MD, MSc(Epid)
 Farrah Morrow, BSc
 Department of Clinical Epidemiology and Biostatistics, Room 2C12 (M.B.), and Department of Surgery (F.M.), McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada. E-mail address for M. Bhandari: bhandari@netinc.ca

Abhaya V. Kulkarni, MD, MSc(Epid)
 Department of Surgery, University of Toronto, Toronto ON M5G 1X5, Canada

Paul Tornetta III, MD
 Department of Orthopaedic Surgery, Boston Medical Center, 720 Harrison Avenue, Boston, MA 02118

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article. No funds were received in support of this study. Dr. M. Bhandari was supported by the Ronald K. Fraser Foundation Research Fellowship, Hamilton, Ontario, Canada. Dr. A.V. Kulkarni was supported by the Duncan L. Gordon Fellowship from the Hospital for Sick Children Foundation, Toronto, Ontario, Canada.

10. **Goldman L, Feinstein AR.** Anticoagulants in myocardial infarction. The problems of pooling, drowning, and floating. *Ann Intern Med.* 1979;90:92-4.
11. **Thompson SG, Pocock SJ.** Can meta-analysis be trusted? *Lancet.* 1991;338:1127-30.
12. **Oxman AD, Guyatt GH.** Validation of an index of the quality of review articles. *J Clin Epidemiol.* 1991;44:1271-8.
13. **Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, Streiner DL.** Agreement among reviewers of review articles. *J Clin Epidemiol.* 1991;44:91-8.
14. **Fleiss JL.** Measuring agreement between two judges on the presence or absence of a trait. *Biometrics.* 1975;31:651-9.
15. **Donner A, Klar N.** The statistical analysis of kappa statistics in multiple samples. *J Clin Epidemiol.* 1996;9:1053-8.
16. **Sackett DL, Haynes RB, Guyatt GH, Tugwell P.** *Clinical epidemiology. a basic science for clinical medicine.* 2nd ed. Boston: Little, Brown; 1991. p 30.
17. **Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith, PG.** Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer.* 1976;34:585-612.
18. **Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG.** Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer.* 1977;35:1-39.
19. **Pocock SJ.** *Clinical trials. a practical approach.* New York: Wiley; 1983.
20. **Yusuf S.** Obtaining medically meaningful answers from an overview of randomized clinical trials. *Statist Med.* 1987;6:281-6.
21. **Khan KS, Daya S, Jadad A.** The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med.* 1996;156:661-6.
22. **Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP.** Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352:609-13.
23. **Lau J, Ioannidis JPA, Schmid CH.** Quantitative synthesis in systematic reviews. In: Mulrow C, Cook D, editors. *Systematic reviews: best evidence for health care decisions.* Philadelphia: American College of Physicians; 1998. p 91-101.
24. **Breslow NE, Day DE.** Combination of results from a series of 2x2 tables; control of confounding. In: *Statistical methods in cancer research. Volume 1, The analysis of case-control studies.* Lyon, France: International Agency for Research on Cancer; 1980. p 136-46.
25. **Greenland S, Salvan A.** Bias in the one-step method for pooling study results. *Statist Med.* 1990;9:247-52.
26. **Jadad AR, McQuay HJ.** Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol.* 1996;49:235-43.
27. **Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D.** Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA.* 1998;280:278-80.
28. **Egger M, Schneider M, Davey Smith G.** Spurious precision? Meta-analysis of observational studies. *BMJ.* 1998;316:140-4.