Role of Technology Assessment in Orthopaedics

Charles Turkelson PhD, Joshua J. Jacobs MD

Published online: 30 April 2009
© The Association of Bone and Joint Surgeons® 2009

Abstract A technology assessment is a literature-based research project that seeks to determine whether a medical device, drug, procedure, or biologic is effective or to summarize literature on a given technology. A well-conducted assessment is a form of secondary research that employs the same steps used in primary research studies (ie, well-designed clinical trials). The primary difference is that in technology assessment the investigator does not collect the raw data. Rather, (s)he must use data collected by someone else. Nevertheless, a well-designed assessment, like a well-designed study, employs the scientific method, which is a method designed to combat bias. When there is little available information, such as with new technologies, unbiased examinations can typically show that enthusiasm for that technology is not backed by much data. When there is more information, assessments can not only determine whether a technology is effective, but also how effective it is. Technology assessments can provide busy orthopaedic surgeons (who do not have the time to keep up with and critically evaluate current literature) with succinct information that enables them to rapidly determine

what is and what is not known about any given medical technology.

Introduction

The volume of the peer-reviewed literature is growing rapidly. Between 1965 and 2001, there was an exponential increase in the number of published randomized controlled trials (RCTs) [45] and, in the area of osteoarthritis treatment, the number of published RCTs tripled between 1987 and 1988 and 2001 and 2002 [6]. Keeping up with this literature is very difficult for many physicians, who simply do not have enough time for it.

Critically reading this literature is even more time-consuming. Unfortunately, critical reading is particularly important in orthopaedics because the validity of the results of many studies reported in orthopaedic journals is questionable [16, 26]. Making it even more difficult to determine whether a study's conclusions are valid is that claims from highly cited studies continue to be repeated in the literature, even after subsequent evidence suggests that these claims are untrue [22, 42]. For example, the first studies published on a topic tend to find larger effects than later trials [16, 44], yet the earlier, more positive effects continue to be cited [22]. Indeed, "contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes" [22].

In the face of such challenges, it is tempting to consult experts for their opinions. However, Antman et al. [1] have provided evidence suggesting that expert opinion, as expressed in traditional review articles and texts, often fails to note important medical advances, and that experts sometimes recommend treatments even after those treatments are shown to be potentially harmful. Further

One or more of the authors (JJJ) has received funding from Zimmer, Medtronic, Wright Medical, Spinal Motion, and Advanced Spine Technologies. Each author certifies that he or she has or may receive payments or benefits from a commercial entity related to this work (JJJ: Zimmer, Medtronic).

C. Turkelson
Department of Research and Scientific Affairs, American
Academy of Orthopaedic Surgeons, Rosemont, IL, USA

J. J. Jacobs (✉)
Department of Orthopaedic Surgery, Rush University Medical
Center, Chicago, IL 60612, USA
e-mail: joshua.jacobs@rushortho.com

undermining the validity of expert opinion is more recent evidence suggesting that slightly over half of all professional advice in one specialty area is not in agreement with the research literature [35].

Technology assessment provides the potential to be an accurate and objective way for physicians to reduce the time they must spend reading the literature. To ensure there is no confusion, we must first clarify that the term “technology assessment” as used in this article specifically refers to health technology assessment. Also, we are using the Institute of Medicine’s (IOM) definition of “technology assessment,” which is “any process of examining and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended” [21]. In this context, “technology assessment” does not exclusively refer to computer technology or any other branch of information technology. Similarly, “technology” does not refer solely to medical devices. Rather, as defined by the former Office of Technology Assessment (OTA) medical technologies are “the drugs, devices, and medical and surgical procedures used in medical care, and the organizational and supportive systems within which such care is provided” [5]. This definition embraces both innovations in medicine (including new drugs, biologics, medical devices, and procedures) and existing ones.

High-quality technology assessments are systematic reviews like those that appear in the published literature. As such, the activity of performing a technology assessment is secondary research based on data reported in the peer-reviewed literature. Technology assessment does not directly involve laboratory or materials testing or any other kind of primary data collection. That technology assessment is secondary research should not lull one into thinking that it is easy to do. Depending upon the amount of available literature, it can take anywhere from a few months to 1 or 2 years to complete an assessment. This implies that performing a technology assessment can be expensive. We know of no formal studies on this expense but the assessments we have been associated with have cost between approximately \$50,000 and \$300,000 to prepare.

Despite the fact that technology assessments (and systematic reviews) are literature-based, they are not traditional review articles. Most traditional reviews have many weaknesses, including lack of transparency, lack of rigor in preparation, and the potential for bias. Indeed, the fact that many traditional reviews are initiated precisely because of an author’s expert opinion only increases their potential for bias.

We highlight the steps that technology assessments follow to combat bias. In doing so, we will discuss how the scientific method can be applied to evaluating the

published literature and draw parallels between this method and the design and conduct of a clinical trial. Specifically we will discuss (1) framing questions for an assessment, (2) developing “rules of evidence” for including and excluding articles, (3) locating studies, (4) evaluating study quality, and (5) synthesizing study results.

How Technology Assessment is Conducted

Technology assessments (and systematic reviews) are prepared according to widely used (and widely accepted) protocols. We describe here the basic aspects of such protocols. A more comprehensive description appears in the Cochrane Handbook for Systematic Reviews of Interventions [20]. The AMSTAR instrument [36], which reflects much of what is in the Cochrane Handbook, is a useful checklist for determining the quality of a technology assessment of interventions. Finally, the QUOROM statement [30] (recently renamed the PRISMA statement) provides recommendations for reporting meta-analyses of randomized controlled trials of interventions. There are also reporting guidelines for reporting of observational studies and for the meta-analyses of them [39, 47] (a library of these and other reporting guidelines can be found at <http://www.equator-network.org/index.aspx?o=1032>). In considering these reporting guidelines, it is important to note that quality (as indexed by AMSTAR) and reporting (as indexed, for example, by QUOROM) are different concepts. Although it is often difficult to determine the quality of a poorly reported systematic review, well-reported technology assessments and systematic reviews can be poorly designed and conducted, and vice-versa. (For example, one item in the QUOROM statement relates to the title of an article, and several other items are about the contents of the article abstract. However, a poorly worded title and a less than optimal abstract do not necessarily indicate that a meta-analysis was poorly conducted. Similarly, even perfectly reported observational studies are usually of less than optimal quality.)

Technology assessments of screening and diagnostic technologies follow the same protocol as assessments of interventions (and the same processes described in this article). However, appraisal of the quality of screening and diagnostic studies differs from that of interventional studies, as does the statistical analysis of the data. The QUADAS instrument can be used to evaluate the quality of individual studies [48], and the STARD instrument [9, 10] can be used to assess the quality of reporting. However, even a study that achieves a perfect score on both instruments may be reporting biased results [4]. Readers of meta-analyses of screening and diagnostic articles should also be aware that these meta-analyses are more technical than

meta-analyses of interventions. Meta-analysis of screening and diagnostic tests requires use of advanced statistical techniques such as hierarchical regression or a bivariate random effects model [27].

The American Academy of Orthopaedic Surgeons uses the methodology described in the present article when preparing its clinical practice guidelines and its technology overviews. These documents are available for free at <http://www.aaos.org/research/>, and they provide orthopaedic-related examples of the processes we describe.

Ask the Right Questions the Right Way

Technology assessments begin by asking a series of well-focused questions that serve the same role as do hypotheses in laboratory experiments or clinical studies. Specifying these questions at the outset not only focuses an assessment, it also helps to combat bias by preventing someone from asking new questions if the answers to the original ones were not to their liking.

Good assessment questions are empirical. They are best answered by conducting a study that yields numerical results. They are not answered by opinion, and well-framed questions never ask whether a technology is appropriate (for example, they never ask whether it is appropriate to use Treatment A for a patient with Disease X). “Appropriate” is a subjective term that is difficult to quantify and, what one person considers appropriate might be something that another person thinks of as less appropriate. Similarly, well-framed questions do not ask whether a technology is “medically necessary,” or whether it is “experimental” or the “standard of care.” Questions that employ these terms are subjective, and ask for an opinion, not a data-driven answer.

Phrased in another way, the questions in a technology assessment essentially ask, “Does it work?”, “How well does it work?”, “Does it work better than the alternative?”, and “In whom does it work?”. Typically, the questions asked in a technology assessment are about the impact of a technology on patient outcomes, and not on technical measures. This keeps their focus on the benefits (and harms) that patients experience as a result of using the technology.

Develop Rules of Evidence

Just as a well-designed primary research study has inclusion criteria that determine which patients will be enrolled in it, a well-conducted technology assessment has inclusion criteria that determine which articles it will include. These criteria are the assessments “rules of evidence,” and studies that do not fit these rules do not provide admissible evidence.

Because the focus of an assessment is usually on how well patients fare, many assessments consider only studies of humans and either exclude or deemphasize animal studies and biomechanical studies. Often, older studies are not considered because they do not reflect current medical practice. Traditional review articles are almost never included because they are not a source of primary data. Finally, meeting abstracts are usually not considered because the quality of reporting in them is poor and because the study results that appear in the subsequent, full published report may be different from those reported in the abstract [7, 19].

Regardless of which specific inclusion criteria are adopted, their most important feature is that they are adopted before any articles are considered. This helps prevent bias because it helps to keep writers of an assessment from primarily including articles that support a particular point of view.

Find All the Data

In their efforts to be unbiased, technology assessments attempt to locate all of the literature relevant to a particular question. A clinical study that examined only part of the data collected in it would be highly suspect, particularly if the authors chose to consider only data that agreed with their expert opinion. It seems more than reasonable to hold a literature review to the same standard.

Literature searches in a technology assessment should not be confined to MEDLINE. Suarez-Almazor et al. [40] found that MEDLINE searches identified only 73% of the controlled trials on rheumatoid arthritis, osteoporosis, and low back pain, whereas EMBASE searches identified 85% of such trials. Wilkins et al. [49] have suggested even more dramatic differences, with MEDLINE identifying only 40% of the unique citations in osteoarthritis and EMBASE identifying the remaining 60%. Because an incomplete dataset might also be a biased dataset, authors of good assessments search, at a minimum, for articles in both PubMed and EMBASE. Often, CIHAHL and the Cochrane Central Register of Controlled Clinical Trials are also searched, as well as other databases. To further ensure that no relevant articles are missed, searches of electronic databases are typically supplemented by examining the bibliographies of all of the articles retrieved for an assessment.

Evaluate Study Quality

Evaluating the quality of the studies included in a technology assessment is extremely important. The results of

poor quality studies are difficult to trust and, unfortunately, poor quality studies are common in the medical literature. The situation is similar to the one we would expect if we were conducting a clinical trial and found that those who collected the data for us did not do so in the way we hoped they would. The result is that some of data cannot be trusted.

The challenges posed by suboptimal literature are particularly keen when one moves from the study of pharmacological to nonpharmacological interventions. Randomized controlled trials of the latter are of poorer quality than randomized controlled trials of the former in the area of hip and knee osteoarthritis, and reports of randomized controlled trials of surgery, arthroscopy, joint lavage, rehabilitation, and intra-articular injection are of the lowest quality [11]. Studies without control groups are even more challenging, both because they are so common in the orthopaedic literature and because their results are often unreliable. Such studies do not take account of the fact that patients' symptoms may wax and wane as part of the natural course of their disease or condition. Therefore, improvements observed in such studies might not be due to treatment, but to this "natural" improvement.

A crude appreciation of whether a study's results are trustworthy is provided by evidence hierarchies, of which the "levels of evidence" used in the *Journal of Bone and Joint Surgery* and *CORR* is one example. In general, higher levels of evidence (eg, well-designed randomized controlled trials denoted as Level I studies) tend to be evidence one can trust, while lower levels of evidence (eg, Level IV case series) tend to be suspect. There are many systems for describing levels of evidence, and the conclusions one reaches about the effectiveness of a technology can be markedly affected by which system one uses [15]. Methodologists (eg, the Cochrane Collaboration) tend not to use levels of evidence, and have developed other ways to evaluate how trustworthy the results of a study are, ranging from simple [24] to complex [13] methods for gauging the quality of randomized controlled trials, and to instruments for gauging whether case series are reliable [12]. In general, these instruments ask more detailed questions about the design and conduct of a study than are asked in a "levels of evidence" approach. A report by the Agency for Healthcare Research and Quality [3] (available online at <http://www.ahrq.gov/clinic/epcsums/strengthsum.htm>) summarizes the features of many of these quality assessment tools. Nevertheless, which tool one uses is largely a matter of personal preference. The important thing is that the assessment of study quality be performed according to a priori, transparent, and reproducible rules. Nonstructured critiques of study quality may be open to bias.

Synthesize Study Results

With rare exceptions, no study, regardless of its design, should be interpreted in isolation [18]. We would likely not use the results of a clinical study that reported the results from only one patient as an index of how most patients would fare. After all, case reports are usually published precisely because there is something unique about the case. In the same way, we should be cautious of using the results of a single study as if it was representative of all studies. Its results could have been influenced by the skill and experience of the participating physicians, unique institutional practices, enrollment of a group of patients not representative of those usually seen in clinical practice, use of unique diagnostic methods, or a host of other factors. That the first randomized controlled trials published on a topic tend to find larger effects than later trials [17, 44] illustrates why one should not overly rely on the results of a single trial.

Assessments that are conducted well do not rely on the conclusions of the authors of the included studies. This point is often missed by those who prepare assessments. Not relying on the conclusions of article authors recognizes the fact that these authors may be biased. One documented source of bias is manufacturer funding. In orthopaedics, published commercially funded studies tend to report a greater number of positive outcomes and larger positive effects than independently funded studies [14, 28]. The best assessments examine only the Methods section of an article to determine whether the results are valid and, if they are valid, the Results section. Little weight is placed on a study's Discussion section.

After evaluating the individual studies, their results are synthesized. A well-formulated synthesis looks across the results of all the available relevant studies and determines if their results are similar. If they are not, the synthesis attempts to explain why they are different. The term "synthesis" implies that examining a body of literature by providing study-by-study descriptions of the results of each study is often not helpful.

Synthesis can be either narrative (ie, qualitative) or meta-analytic (ie, quantitative). Qualitative syntheses are most suitable when data are sparse. Reviews of sparse data provide an overview of the "state of the art," but usually are unable to come to robust conclusions about the effectiveness of a technology. This inability to come to conclusions in the face of sparse data particularly affects assessments of new technologies. Nevertheless, assessments of new technologies can be useful to orthopaedists who wish to provide patients with information that supplements direct-to-consumer advertising. The *Cochrane Handbook* contains a description of how to conduct a qualitative review [32] but, in general, the processes for

conducting qualitative and quantitative syntheses are the same.

Quantitative analyses consist of meta-analyses, the statistical combining of the results of different studies. Meta-analyses can provide conclusions about whether a technology is effective and about how effective it is. For assessments of interventions, meta-analyses are typically performed when the results of randomized controlled trials are available. Meta-analyses of screening and diagnostic technologies require diagnostic cohort studies. Although meta-analysis can provide the most useful literature syntheses (because, for example, it provides an estimate of the size of a treatment's effect), a lack of appropriate studies deters them from being performed in many orthopaedic subject areas.

Another dimension of synthesis evaluates the costs of a technology. These analyses take the form of cost-effectiveness or cost benefit analysis. Cost-effectiveness analysis seeks to compare the relative value of diagnostics and/or interventions in creating better health and/or a longer life. Results of cost-effectiveness analyses are expressed as a ratio, where the numerator expresses the cost of obtaining that a health gain and the denominator represents the gain in health afforded by a technology. As such, the results the results of a cost-effectiveness analysis might be that implementation of a new technology "costs" \$100/life saved, which would yield a ratio of 100. Although it is theoretically possible to empirically derive a cost-effectiveness ratio, choosing a specific ratio that causes one technology to be preferred over another is a subjective decision. Recommendations for how to perform a cost-effectiveness on healthcare-related topics have been published [46].

Cost-benefit analysis is similar to cost-effectiveness except that its results are expressed solely in terms of dollars. Therefore, when conducting a cost-benefit analysis, one must place a dollar value on a particular health state or a human life. The advantage of cost-benefit analysis is that it can be used to compare interventions whose benefits are measured in different units. For example, cost-benefit analysis can be used to inform decisions about how much one should spend on education in relation to healthcare. Such analyses are not possible with cost-effectiveness analyses because it requires that the benefits of the interventions being compared be measured in the same units.

Cost analyses in technology assessments do not (and should not) take the form of cost-minimization analyses. These analyses look only at costs as part of an effort to do whatever is least expensive regardless of the health outcome.

Cost analyses only rarely appear in technology assessment. This may partly be because they require a great deal of effort. They can easily expand the time (and budget) it

takes to perform a technology assessment by a year. The consequent risk is that the technology assessment is becoming outdated even as it is published.

Discussion

Technology assessments have the potential to be valuable tools for busy orthopaedic surgeons. They can distill large amounts of information and provide information about the validity of available evidence. The latter can be particularly useful when orthopaedic surgeons and others are faced with information that comes from sources that are not objective.

This does not imply, however, that all technology assessments are well-conducted. Like any other research study, they should be critically read. One study estimated that 88% of all orthopaedic meta-analyses had methodological flaws that could limit their validity [8]. In orthopaedics, readers also need to ensure that a technology assessment is not overstating its conclusions. This is because they are based upon a literature that is, in general, not of high quality. Another study estimated that less than 2% of the orthopaedic literature is comprised of randomized controlled trials and [26] and another [33] estimated that only about 11% of the orthopaedic literature is comprised of Level I evidence while 58.1% of it is Level IV. Even some of the Level I evidence may be suspect. For example, the quality of randomized controlled trials of nonpharmacological interventions (eg, surgery, arthroscopy) for hip and knee osteoarthritis is less than that of trials of pharmacological interventions [11], and smaller randomized controlled trials in the orthopaedic literature tend to find larger effects than larger ones, rendering the results of these smaller trials suspect [41]. That only about half of all systematic reviews take quality into account in their analysis and interpretation of results or in their discussion sections [31] highlights the need for critical reading of assessments and reviews, as does the fact that there can be a subjective component to interpreting results [38]. Formal systems for incorporating the quality (as well the number of studies and the consistency of their results) of the body of relevant literature (as opposed to individual studies) into conclusions have been developed [2, 43], but they are more commonly used in clinical practice guidelines than in systematic reviews and technology assessments.

Readers of systematic reviews and technology assessments should also understand that these documents can come to different conclusions. Reviews and assessments often use different article inclusion criteria [29, 34], and this could explain why. In considering article inclusion criteria employed in a technology assessment or systematic review, it is important to remember that they usually

include only the best (ie, the most reliable) evidence. Therefore, it is only rarely appropriate to include a large case series when data from well-designed randomized controlled trials are available. It is also worth noting that industry-supported reviews tend to have more favorable conclusions than reviews performed by the Cochrane Collaboration [25]. A published algorithm that assists readers in critically evaluating each of the steps described in the present article is helpful in interpreting discordant systematic reviews [23]. Finally, one needs to understand that technology assessments are perishable items. Information suggesting the need for updates appears for almost 60% of systematic reviews, and the median time for it to appear is 5.5 years. For 7% of reviews, such information is present at the time of publication [37].

Demand for technology assessment could increase in the near future due to increased scrutiny of the costs of healthcare. Demand may be particularly great among insurers. However, even though insurers consider assessments in their coverage decisions, how frequently they do so, how heavily their decisions are influenced by an assessment's conclusions, or whether insurers are equally influenced by positive and negative conclusions is not known. We are unaware of any published study that has addressed these aspects of insurer's use of technology assessment in the United States. Similarly, there are no rigorously collected data on the extent to which physicians use assessments or how (or if) they use them in decision making.

There will also likely be an increased desire for assessments of new technologies. However, while technology assessments can be definitive when there is a substantial body of literature, they usually cannot come to conclusions when the literature is sparse, which is often the case with new technologies. If one again thinks of technology assessment as a form of research, for an assessment to come to a conclusion when there are only a few published studies is like coming to a conclusion about the results of a clinical trial after only the first few patients have been enrolled. This does not imply that there is no role for assessments of new technologies. They may still be useful for illustrating where data are lacking and for illustrating where there is more enthusiasm than data for a technology.

Despite the fact that poorly conducted technology assessments exist, there are also a number of good assessments. One can again draw the parallel between assessments and primary research; despite the fact that there are poorly conducted clinical studies, there are also a number of good ones. Well-conducted technology assessments and good clinical studies both overcome bias. Given that physicians, patients, insurers, and policy makers are exposed to a considerable amount of biased information,

this represents a major accomplishment. To achieve freedom from bias, technology assessments are developed using processes similar to those used in conducting a good clinical study. In short, good technology assessments, like good all research studies, are prepared using the scientific method.

Acknowledgments We thank Kristen Hitchcock, MIS, for library assistance in preparing this article.

References

1. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA*. 1992;268:240–248.
2. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S, GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.
3. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr, GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res*. 2004;4:38.
4. Bachmann LM, ter RG, Weber WE, Kessels AG. Multivariable adjustments counteract spectrum and test review bias in accuracy studies. *J Clin Epidemiol*. 2009;62:357–361.
5. Banta HD, Behney CJ, Sisk JS. *Toward Rational Technology in Medicine*. New York, NY: Springer; 1981.
6. Bayat N, Keen HI, Hill CL. Randomized clinical trials of osteoarthritis: a review. *APLAR J Rheum*. 2005;8:171–176.
7. Bhandari M, Devereaux PJ, Guyatt GH, Cook DJ, Swiontkowski MF, Sprague S, Schemitsch EH. An observational study of orthopaedic abstracts and subsequent full-text publications. *J Bone Joint Surg Am*. 2002;84:615–621.
8. Bhandari M, Morrow F, Kulkarni AV, Tornetta P, III. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am*. 2001;83:15–24.
9. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003;326:41–44.
10. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med*. 2003;138:40–44.
11. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA*. 2003;290:1062–1070.
12. Carey TS, Boden SD. A critical guide to case series reports. *Spine*. 2003;28:1631–1634.
13. Chalmers TC, Smith H, Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials*. 1981;2:31–49.
14. Ezzet KA. The prevalence of corporate funding in adult lower extremity research and its correlation with reported results. *J Arthroplasty*. 2003;18:138–145.

15. Ferreira PH, Ferreira ML, Maher CG, Refshauge K, Herbert RD, Latimer J. Effect of applying different "levels of evidence" criteria on conclusions of Cochrane reviews of interventions for low back pain. *J Clin Epidemiol*. 2002;55:1126–1129.
16. Gartland JJ. Orthopaedic clinical research. Deficiencies in experimental design and determinations of outcome. *J Bone Joint Surg Am*. 1988;70:1357–1364.
17. Gehr BT, Weiss C, Porzolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Med Res Methodol*. 2006;6:25.
18. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328:39–41.
19. Guryel E, Durrant AW, Alakeson R, Ricketts DM. From presentation to publication: the natural history of orthopaedic abstracts in the United Kingdom. *Postgrad Med J*. 2006;82:70–72.
20. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Oxford, UK: Wiley-Blackwell; 2008.
21. Institute of Medicine. *Assessing Medical Technologies*. Washington, DC: National Academy Press; 1985.
22. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294:218–228.
23. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ*. 1997;156:1411–1416.
24. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1–12.
25. Jorgensen AW, Hilden J, Gotzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ*. 2006;333:782.
26. Kiter E, Karatosun V, Gunal I. Do orthopaedic journals provide high-quality evidence for clinical practice? *Arch Orthop Trauma Surg*. 2003;123:82–85.
27. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;149:889–897.
28. Leopold SS, Warne WJ, Fritz BE, Shott S. Association between funding source and study outcome in orthopaedic research. *Clin Orthop Relat Res*. 2003;293–301.
29. Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med*. 2003;96:17–22.
30. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*. 1999;354:1896–1900.
31. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ*. 2005;330:1053.
32. Noyes J, Popay J, Pearson A, Hannes K, Booth A. Qualitative research and Cochrane reviews. In Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Oxford, UK: Wiley-Blackwell; 2008:571–591.
33. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P, III, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am*. 2005;87:2632–2638.
34. Peinemann F, McGauran N, Sauerland S, Lange S. Disagreement in primary study selection between systematic reviews on negative pressure wound therapy. *BMC Med Res Methodol*. 2008;8:41.
35. Schaafsma F, Verbeek J, Hulshof C, van Dijk F. Caution required when relying on a colleague's advice; a comparison between professional advice and evidence from the literature. *BMC Health Serv Res*. 2005;5:59.
36. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
37. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007;147:224–233.
38. Shrier I, Boivin JF, Platt RW, Steele RJ, Brophy JM, Carnevale F, Eisenberg MJ, Furlan A, Kakuma R, Macdonald ME, Pilote L, Rossignol M. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak*. 2008;8:19.
39. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA*. 2000;283:2008–2012.
40. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Control Clin Trials*. 2000;21:476–487.
41. Sung J, Siegel J, Tornetta P, Bhandari M. The orthopaedic trauma literature: an evaluation of statistically significant findings in orthopaedic trauma randomized trials. *BMC Musculoskelet Disord*. 2008;9:14.
42. Tatsioni A, Bonitsis NG, Ioannidis JP. Persistence of contradicted claims in the literature. *JAMA*. 2007;298:2517–2526.
43. Treadwell JR, Tregear SJ, Reston JT, Turkelson CM. A system for rating the stability and strength of medical evidence. *BMC Med Res Methodol*. 2006;6:52.
44. Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, Ioannidis JP; EU-PSI project. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol*. 2004;57:1124–1130.
45. Tsay MY, Yang YH. Bibliometric analysis of the literature of randomized controlled trials. *J Med Libr Assoc*. 2005;93:450–458.
46. United States Public Health Service Panel on Cost-effectiveness in Health and Medicine. *Cost-effectiveness in Health and Medicine: Project Summary: From the Report of the Panel on Cost-effectiveness in Health and Medicine*. Washington, DC: Office of Public Health and Science, U.S. Public Health Service; 1996.
47. Vandenbroucke JP, von EE, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007;147:W163–W194.
48. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
49. Wilkins T, Gillies RA, Davies K. EMBASE versus MEDLINE for family medicine searches: can MEDLINE searches find the forest or a tree? *Can Fam Physician*. 2005;51:848–849.