



Editorial

Editorial: Getting the Most from What You Read in Orthopaedic Journals

Seth S. Leopold MD

Many orthopaedic surgeons tell me they are uneasy about reading scientific journals. This is perhaps not surprising; the Accreditation Council for Graduate Medical Education requires that orthopaedic residents do research, but it has no specific requirements for journal clubs or critical-reading programs for trainees [1]. Training programs allocate resources accordingly, and the result is that some practitioners find reading scientific papers a real challenge.

No one is born a good reader of science; this skill improves with practice, and it can be taught and learned. A few principles can help, including a screening approach to help busy clinicians decide which articles to read in depth, a few key questions one can “ask” of a study’s methods sections, and available critical-reading tools that can take an interested reader even further.

Deciding What to Read

Journals publish far more articles than even the most-avid reader could possibly consume. Nearly 7000 orthopaedic articles were published in 2009 and the compound rate of growth was 10.2% [8], suggesting that this year more than 15,000 orthopaedic articles will see daylight. Thomson Reuters indexed 74 orthopaedic journals for its calculation of the 2015 Impact Factor (the latest year for which data are available [3]); Scimago listed 225 [13]. There probably are more. We need to be choosy about what we read, since it’s impossible to digest it all.

One possible screening approach is to identify a small number of trusted sources—perhaps skim the contents of

a high-quality general-interest orthopaedic journal or two, a subspecialty journal, and a leading medical or science publication. Since even three or four journals will deliver more content than most of us can reasonably consume before the next issue comes out, we need to be still-more selective. Start by reading the article titles, and for those that seem worthwhile, skip to the Conclusions section and ask, “If it were true, would it matter to me?” Do not, at this point, try to determine whether it is true, since without a deeper dive it’s impossible to know. Just use that question as a screening test.

If an article passes that test, skip directly to the Methods section of the paper, and begin the real operation.

Key Questions for the Methods Section

The thoughtful consumer of orthopaedic research will focus on two main themes while reading an article’s Methods section: Validity and bias. Validity comes in two flavors: External validity (generalizability) and internal validity (methodological rigor). Criteria for external validity—whether a reader can generalize the results of a

The author certifies that neither he, nor any members of his immediate family, have any commercial associations (such as consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article. All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request. The opinions expressed are those of the writers, and do not reflect the opinion or policy of *CORR*® or The Association of Bone and Joint Surgeons®.

S. S. Leopold MD (✉)
Clinical Orthopaedics and Related
Research®, 1600 Spruce St.,
Philadelphia, PA 19103, USA
e-mail: sleopold@clinorthop.org

Editorial

study done by other people in other places to one's own practice—are the same regardless of the type of article, while the criteria for internal validity vary by study type. One needs to consider both external and internal validity to ascertain whether a study's findings are applicable and trustworthy.

External validity (generalizability) checklist

- Does it apply to my practice?
- Do the patients in this article look like my patients?
- Do my skills and experience compare with those of the surgeons in the study?
- Are the technology and facilities used in the study available to me?

The key factors in determining whether the findings from a study might generalize well to one's own practice pertain principally to the patients and practice setting, the skills and experience of the provider performing the intervention, and special technology, tools, or facilities that might be needed to do the job well.

Patient-specific factors like age, activity level, and workplace demands (or workplace injuries) can differ dramatically from one study to another, and warrant consideration. Findings from tertiary-care practices do not always generalize well to community orthopaedic settings and vice versa. As

in sports and musical performance, the background and experience of the performer can matter tremendously. Often, authors write about procedures they have done hundreds or even thousands of times; it may not be reasonable to assume that one's first attempt will go as well as the authors' 4000th. Finally, the availability of specific tools or facilities may make a big difference. Sometimes these technological differences are obvious—computer navigation, robotics, or devices. But sometimes they are subtle—experienced anesthesiologists for procedures that might run long, or expert pathologists to make difficult diagnoses from intraoperative pathology specimens.

While those factors apply to a wide variety of study designs, the assessment of methodological rigor depends on design-specific questions. If a reader determines that a study is sufficiently applicable to his or her own practice, the next step is to identify the study's design, and ask the questions that apply from the internal validity checklists below. For reasons of length, I've confined this discussion to tools for the most-common orthopaedic study designs: Studies on surgical treatments, new diagnostic tests, and systematic reviews. Book-length treatments of this topic—including toolkits like these for studies of many other designs—certainly are widely

available and worth having on the shelf [6].

Internal validity checklist: Retrospective studies about surgical treatments

- Which patients underwent the procedure in question (indications) and which patients were included in this study (selection bias)?
- Is the followup sufficiently long and complete to identify all of the outcomes a clinician might care about (transfer bias)?
- Are the endpoints assessed using validated, robust tools, and who is asking the questions (assessment bias)?

It is critically important to remember that the effects of these three kinds of biases are additive: They all tend to make novel interventions look better than they really are. Selection bias, transfer bias, and assessment bias do not offset one another.

A robust study about a therapeutic intervention will define, at least in general terms, what indications drove the decision to perform the procedure being studied, and will inform the reader whether the study reports results on all or only a select subset of the patients who underwent the intervention. Both the length and the completeness of followup can make a crucial difference in the robustness of a study. If an intervention produces

Editorial

short-term relief at the risk of some longer-term risk, then short-term followup cannot justify the adoption of an intervention. In a similar vein, a study reporting 97% success with 45% of the patients missing is not a study the reader generally can have confidence in, since patients who are missing may be more likely to have treatment-related complications than those who are accounted for [4, 11]. Finally, pay attention to who is asking the questions, and what questions they ask, in assessing the efficacy of a novel treatment. Subjective, provider-administered questionnaires may make inflate the apparent benefits of novel treatments, when compared to validated patient-reported outcomes tools.

Internal validity checklist: Prospective studies about surgical treatments

- Was randomization used?
- Were patients and providers blinded to the interventions performed?
- If randomization was used, were all patients analyzed in the groups to which they were assigned (intention-to-treat analysis)?

When evaluating prospective studies about surgical interventions, it is worth starting with the same checklist for retrospective studies on surgical interventions. But since prospective studies may further benefit from randomization, blinding, and (if randomization was used), intention-to-

treat analysis, it is worth considering those elements as well.

Because these features tend to minimize some of the more-severe kinds of bias so pervasive in retrospective research, randomized trials tend to report smaller effect sizes than are observed in retrospective studies.

New ideas presented in surgical journals tend to follow a similar life-cycle: Early studies present a novel technique in a case-series and it looks promising. Closer inspection finds that patients in those series were carefully selected, some were lost to followup, and assessment was rather subjective. Followup research in randomized trials—which eliminate selection bias, account for all or most patients, and use standardized assessment tools—determine the new treatment is much-less effective than was earlier believed.

Internal validity checklist: Studies about diagnostic tests

- Was the new test compared against a “gold standard”?
- Was the “gold standard” test applied to all patients, regardless of the results of the novel diagnostic tool?
- Was the study performed in a population that includes a reasonably broad spectrum of disease?

The randomized trial is, in fact, not the definitive study design for every kind of research. Studies evaluating the utility of novel diagnostic tests do

not particularly benefit from randomization unless two novel tests are being compared to one another, which rarely is the case. Rather, new tests should be compared to accepted and valid reference standards for diagnosis of the treatment in question. Without this, there is no obvious way to know whether a new test is accurate.

In particular, it is important to apply the “gold standard” for diagnosis to all patients receiving the new test; if the reference standard is applied only when the new test is positive, the new test’s diagnostic properties may look better than they actually are. Readers should consider carefully the population of patients involved in any study of diagnostic tests. Although this is to some degree an “external validity question,” the prevalence of the disease in question also influences the reported properties of a test. For example, the positive predictive value of a test will increase in populations where the condition of interest is more common, and negative predictive value goes up in populations where the prevalence is lower—even if the same diagnostic test is used [10].

Internal validity checklist: Systematic reviews and meta-analyses

- Is the search adequately described?
- Do the authors address the fact that not all studies are similarly well designed or convincing?

Editorial

- Do the authors describe the differences in effect sizes and directions across the studies they included (heterogeneity)?

A good meta-analysis is a scientific undertaking that should be as rigorous as any laboratory experiment; the Methods section of a systematic review should allow the reader to replicate the work, starting with the search terms and databases used, so the reader can see that no important topics were missed. Systematic reviews also should explain how the “grey literature” such as non-English-language studies, conference proceedings, and unpublished work—was handled. The savvy reader (and the thoughtful meta-analyst) recognizes that different study designs are likely to suffer from different kinds of bias. Lumping case series in with randomized trials in a meta-analysis is likely to result in the accumulation of the very kinds of bias that these studies should help us to minimize.

Heterogeneity in meta-analysis can be an intimidating topic, because the statistical approaches to dealing with it involve some heavy mathematical sledding. But it is a key topic to understand at least on an intuitive level, since it will result in quick (and appropriate) disqualification of any meta-analysis that fails to mention it. Imagine a meta-analysis that pools the results from five studies showing a

treatment to be toxic with five others showing it to be curative. Summing the effects of those 10 studies may result in a finding of no effect, which would be both misleading and risky. Statistical approaches can help deal with this, but in essence it is better not to lump together that which is better off split apart. Readers might ask (and authors should report): Why might there have been such differences among the studies? Did the patient populations differ in the source studies? Were the treatments administered differently? Was followup insufficiently long in the five “positive” studies to detect harms that accrue over longer periods of followup?

Tools for Thoughtful Readers

Many of these ideas have been around for some time; the concept of selection bias, for example, was identified at least as far back as 1843 [5], although the term itself was not coined until much later. An early and thoughtful treatment of this topic, which heavily informs both my own approach and these checklists, first was published in *JAMA* in the early 1990s as a series called “Users’ Guides to the Medical Literature” [12]. Over time, that series came to include dozens of articles, providing suggestions on how best to consume a wide variety of article

types. We at *Clinical Orthopaedics and Related Research*® are proud to have published a pair of them recently [2, 7]. Although the clinical examples in the original *JAMA* series are not musculoskeletal, those classic articles remain well worth a look. A book-length treatment of these critically important topics is available [6], which I heartily recommend.

Beyond that, *CORR*® offers an online tool (<http://www.clinorthop.org/reviewertool>) that we originally designed for our peer reviewers, but which serves equally for journal clubs and even for the busy clinician who would like some guided practice at applying these sorts of checklists in practice [9]. We encourage readers of all journals to read with these kinds of questions in mind, as they can help us determine which articles should inform and influence our practices, and which ones may be less trustworthy or applicable.

References

1. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in orthopaedic surgery. Approved October 1, 2011; last revised September 27, 2015. Available at https://www.acgme.org/Portals/0/PFAAssets/ProgramRequirements/260_orthopaedic_surgery_2016.pdf. Accessed on April 10, 2017.

Editorial

2. Chaudhry H, Foote CJ, Guyatt G, Thabane L, Furukawa TA, Petrisor B, Bhandari M. Network meta-analysis: Users' guide for surgeons part II – certainty. *Clin Orthop Relat Res.* 2015;473:2172–2178.
3. Clarivate Analytics. Science citation index expanded – orthopedics – journal list. Available at: <http://science.thomsonreuters.com/cgi-bin/jmlst/jlresults.cgi?PC=D&SC=TC>. Accessed April 24, 2017.
4. Farrokhyar F, Karanicolas PJ, Thoma A, Simunovic M, Bhandari M, Devereaux PJ, Anvari M, Adili A, Guyatt G. Randomized controlled trials of surgical interventions. *Ann Surg.* 2010;251:409–416.
5. Guy WA. Contributions to the knowledge of the influence of employments upon health. *J Roy Stat Soc.* 1843;6:197–211.
6. Guyatt G, Rennie D, Meade MO, Cook DJ. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd Ed.* Columbus, OH: McGraw-Hill Education; 2015.
7. Foote CJ, Chaudhry H, Bhandari M, Thabane L, Furukawa TA, Petrisor B, Guyatt G. Network meta-analysis: Users' guide for surgeons part I – credibility. *Clin Orthop Relat Res.* 2015;473:2166–2171.
8. Lee KM, Ryu MS, Chung CY, Choi IH, Kwon DG, Kim TW, Sung KH, Seo SG, Park MS. Characteristics and trends of orthopaedic publications between 2000 and 2009. *Clin Orthop Surg.* 2011;3:225–229.
9. Leopold SS. Editorial: *CORR's* new peer-reviewer tool—Useful for more than peer reviews. *Clin Orthop Relat Res.* 2016;474:2321–2322.
10. Maxim LD, Niebo, Utell MJ. Screening tests: A review with examples. *Inhal Toxicol.* 2014;26:811–828.
11. Murray DW, Britton AR, Bulstrode CJ. Loss to follow-up matters. *J Bone Joint Surg Br.* 1997;79:254–257.
12. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA.* 270:2093–2095.
13. Scimago Journal & Country Rank. What is Scimago JR for? Available at: <http://www.scimagojr.com/>. Accessed April 18, 2017.