

TOPICS IN TRAINING

Levels of Evidence Are Low for Clinical Management Questions on the Orthopaedic In-Training Examination

By Joseph Bernstein, MD, Julia A. Kenniston, MD, Jason A. Nydick, DO, Miltiadis H. Zgonis, MD, and Pedro K. Beredjikian, MD

The Orthopaedic In-Training Examination (OITE) has been administered to orthopaedic surgery residents throughout the United States since 1963 and, over the past forty-five years, has become “an integral part of orthopaedic education.”¹ The test was designed as an educational tool for residents, but it also is used to assess the performance of residents relative to that of others.

Additionally, an implicit function of the test is to help to define the corpus of core knowledge for orthopaedic surgery residency. As noted by the participants of the Academic Orthopaedic Society symposium on musculoskeletal education², “Testing organizations are influential in shaping the curriculum. Faculty are correct to teach what students need to know to pass their [tests].”

The Evaluation Committee of the American Academy of Orthopaedic Surgeons (AAOS) writes the questions, amalgamates the examination, and scores the OITE. As such, test takers can reasonably infer that the material tested on the OITE reflects the Academy’s vision of the orthopaedic surgery core curriculum for residents in training, and, owing to the central position of the AAOS, the OITE is an important influence on the orthopaedic sur-

gery curriculum for residents in the United States. For this reason, among others, it is worthwhile to examine the OITE for the purposes of studying what exactly residents are expected to master.

One aspect of that analysis centers on the levels of evidence on which the OITE questions are based. Although the principles of evidence-based medicine³ do not limit the practitioner to the randomized trial (or any particular study type), there has been particular recent emphasis on the level of evidence associated with individual studies. In their editorial introducing the policy of explicitly listing this level, for example, the editors of this journal stated: “Orthopaedic surgeons have always based their clinical care on evidence. . . . Higher levels of evidence should be more convincing to surgeons attempting to resolve clinical dilemmas.”⁴

Accordingly, we asked, what is the level of evidence supporting the credited answers of clinical management questions on the OITE? We hypothesized that orthopaedic surgery residents are tested on treatment principles that are supported by low levels of evidence. To test this hypothesis, we undertook a review of OITE examinations and evaluated the level of evidence of studies

that provide the credited answer to clinical management questions.

Materials and Methods

All questions from the 2006 and 2007 Orthopaedic In-Training Examinations were reviewed by two readers. Each reader independently composed a list of the clinical management questions from the two examinations, with such questions defined as those in which the examinee was provided a clinical scenario and was asked about the management of a patient. The lists prepared by the two readers were then compared to derive a consensus list. All other questions were excluded.

For each clinical management question, we reviewed the recommended readings cited in the AAOS answer key. If the reading was a textbook or monograph, this material was examined to find the primary source to generate a master list of primary literature sources. From these sources, we then identified the evidence supporting the credited answer.

Evidence was characterized as a prospective randomized trial, prospective cohort study, retrospective comparative study, case series, case report, or (when no research data were cited) expert opinion. Answers based on

Disclosure: The authors did not receive any outside funding or grants in support of their research for or preparation of this work. Neither they nor a member of their immediate families received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity.

basic-science studies were categorized as “expert opinion.” For questions that had multiple sources, we denoted the source with the highest level of evidence as the support for the particular question. When sources did not explicitly provide an answer, we categorized this as expert opinion.

To assess the possibility that the recommended reading list did not point us to the highest level of evidence, a random subsample of ten questions was created. For each such question, we performed a MEDLINE search of the topic of the question, recorded the nature of the evidence we found, and compared it with that in the recommended reading list.

After determining the level of evidence for each question, we aggregated the responses by subject area, such as hand, sports, musculoskeletal trauma, etc., and compared the rate of lower level (IV and V) evidence for each subject.

Results

A total of 550 questions were reviewed. From those, 115 clinical management questions, namely, the questions in which the examinee was provided a clinical scenario and asked about the management of a patient, were identified. One item on the consensus list of 115 was deleted by the AAOS in scoring, yielding 114 questions available for further analysis.

The types of studies supporting these 114 clinical management questions were distributed as follows: prospective randomized trials (nine), prospective cohort studies (three), retrospective comparative studies (twenty-four), case series (sixty-three), case reports (two), and expert opinion (thirteen). These study types were then classified, with use of *The Journal's* levels-of-evidence schema, yielding the distribution shown in Figure 1.

In our sample of ten questions, our independent review of the medical literature failed to detect a level of evidence above that which was found with use of the recommended reading list.

Distribution of Levels of Evidence

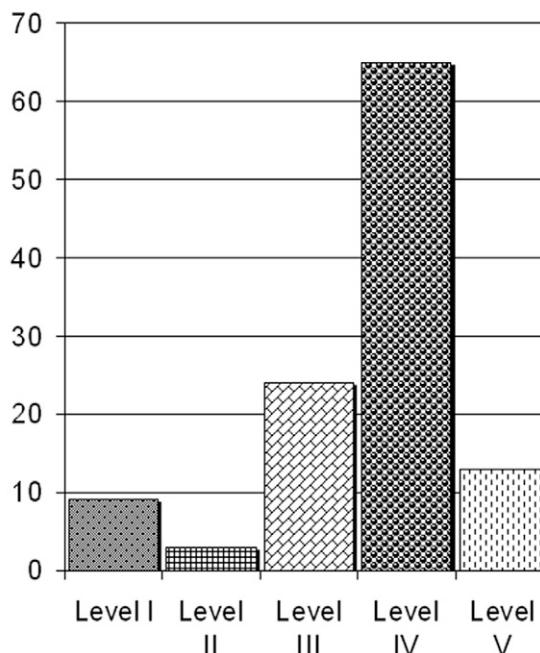


Fig. 1

The number of studies, categorized by levels of evidence, supporting clinical management questions in the OITE sample reviewed.

There were no significantly different rates of lower level (IV and V) evidence among subject areas.

Discussion

Our review of the recommended readings for clinical management questions on the OITE, supplemented by an independent sampling audit, revealed that more than two-thirds of these questions were supported by studies whose level of evidence was IV or lower.

Wright noted that levels of evidence “provide a concise and simple appraisal of study quality. The essence of levels of evidence is that, in general, controlled studies are better than uncontrolled studies, prospective studies are better than retrospective studies, and randomized studies are better than nonrandomized studies.”²⁵ As such, our finding may be particularly interesting to those who evaluate the performance of OITE test takers. It is also, to our knowledge, the first such study both in orthopaedic surgery as well as in other surgical domains.

The first issue to address when considering these results is the limita-

tions of our methods. For one thing, we studied only two years’ worth of examinations. Also, it is possible that some higher-level studies were overlooked both by our search and by the recommended reading lists. (After all, the recommended reading list is just that: recommended reading. It is not offered as a definitive reference list.) Last, it has been recently shown⁶ that level-of-evidence grades are applied imperfectly. Nevertheless, it is unlikely that whatever error is introduced by those factors would markedly alter the inferences from our work. After all, even if the number of Level-I and Level-II studies were doubled, <25% of the questions would be in that category.

Our data should not be used to disparage the OITE in general. For one thing, we examined only the clinical management questions, a small subset of the examination. Also, it appears that these clinical management questions reflect (or perhaps exceed) the state of the art of the orthopaedic literature. The AAOS guidelines on osteoarthritis of the knee⁷, for example, offered twenty-two recommendations, but only four

(18%) of the recommendations had Level-I studies with consistent findings to support (or refute) the recommended intervention. Freedman et al.⁸, in their review of all of the clinical papers in *The Journal of Bone and Joint Surgery (American and British Volumes)* and *Clinical Orthopaedics and Related Research* in one calendar year, reported that 82% of the clinical papers would be classified as Level-IV evidence. Obremsky et al.⁹ reviewed nine orthopaedic surgery journals over six months in 2003 and found that 58% of the articles offered Level-IV evidence. As such, the OITE cannot at once be representative of the orthopaedic surgery literature and yet all the while offer only questions resting on Level-I evidence. (If the examination were to be based on questions only supported by high-quality evidence, a large portion of orthopaedic practice would not be tested, and this, we suggest, would be worse than the current mode of testing.)

It must also be recalled, of course, that the level of evidence is not the final word on quality assessment. These levels are assigned in proportion to the study type's resistance to bias, which of course is only one measure of quality. Thus, while a Level-IV study is more prone to confounding and chance error than one rated Level I, a Level-IV study may be very informative and clinically helpful nonetheless—especially in cases when treatments are adopted before sufficient translational studies to evaluate efficacy and safety have been performed. Consider the example of thermal capsulorrhaphy. This was an operation with great promise not too long ago, yet the popularity of the thermal wand has waned. That change came about not because of a randomized trial showing its lack of effectiveness, but because isolated reports¹⁰ suggested its limitations and dangers. For studies describing harm, we do not need a Level-I study, and if we waited

for one, the damage inflicted might be great. In short, there are many things that make a study useful and valid. To use the level-of-evidence label as the sole metric—to adopt a “trust Level I; discount Level IV” mantra—is a grave mistake.

One must further reflect on the fact that some questions designated to be clinical management were only superficially so; their real purpose was to test judgment and the application of clinical knowledge. These questions were phrased as clinical management questions presumably because of the constraints of the multiple choice format. Consider a question that asked about the management of a patient with a distal interphalangeal joint dislocation, as shown on the radiograph provided. In point of fact, we were able to find no high-level studies on the management of these dislocations, but the correct answer to the question here was not predicated on the knowledge of the outcomes literature for distal interphalangeal joint dislocations. Rather, finding the correct answer hinged on an awareness that, given the radiograph, the volar plate was apt to block closed reduction. As such, one might say this is a high-level question associated peripherally with low-level evidence. (The examiner could have asked “What structure is blocking closed reduction?” in which case the question would not have been categorized as “clinical management,” but that attenuates the question's power, as the wording unnecessarily alerts the test taker to the very point being tested.)

Perhaps the most interesting finding in our study was that only 115 of the 550 questions, or approximately 20%, were identified as “clinical management” questions. This is a fairly low percentage, one may argue, for a test in clinical medicine. After all, the examinees are studying to become practitioners of orthopaedic surgery,

and one might therefore expect a majority of the questions to test management skills. We believe that this low percentage of management questions is appropriate, given that practice management standards are apt to evolve. Thus, what should be tested is not so much whether the examinee has mastered the standard of care as of the date of the examination, but rather whether the examinee has acquired the skill of knowledge acquisition.

We close by noting that the creators of the examination added a caveat in the examination instructions, stating that the OITE “is not intended to be used as a qualifying examination for determining resident promotions.” Given that we found that more than two-thirds of the OITE's clinical management questions were supported by studies whose level of evidence was IV or lower, that caveat is appropriate.

Joseph Bernstein, MD
Department of Orthopaedic Surgery,
University of Pennsylvania School of Medicine,
424 Stemmler Hall, 36th and Hamilton Walk,
Philadelphia, PA 19104. E-mail address:
orthodoc@post.harvard.edu

Julia A. Kenniston, MD
Miltiadis H. Zgonis, MD
Department of Orthopaedic Surgery,
University of Pennsylvania, 3400 Spruce Street,
2 Silverstein, Philadelphia, PA 19104.
E-mail address for J.A. Kenniston:
julia.kenniston@uphs.upenn.edu.
E-mail address for M.H. Zgonis:
Miltiadis.Zgonis@uphs.upenn.edu

Jason A. Nydick, DO
Department of Orthopaedic Surgery,
Philadelphia College of Osteopathic Medicine,
4170 City Avenue, Philadelphia, PA 19131.
E-mail address: jason31878@yahoo.com

Pedro K. Beredjikian, MD
The Rothman Institute, 925 Chestnut Street,
Philadelphia, PA 19107

References

1. Buckwalter JA, Schumacher R, Albright JP, Cooper RR. The validity of orthopaedic in-training

examination scores. *J Bone Joint Surg Am.* 1981; 63:1001-6.

2. Bernstein J, Alonso DR, DiCaprio M, Friedlaender GE, Heckman JD, Ludmerer KM.

Curricular reform in musculoskeletal medicine: needs, opportunities, and solutions. *Clin Orthop Relat Res.* 2003;415:302-8.

3. Bernstein J. Evidence-based medicine. *J Am Acad Orthop Surg.* 2004;12:80-8.

4. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am.* 2003;85:1-3.

5. Wright JG. A practical guide to assigning levels of evidence. *J Bone Joint Surg Am.* 2007;89:1128-30.

6. Schmidt AH, Zhao G, Turkelson C. Levels of evidence at the AAOS meeting: can authors rate their own submissions, and do other raters agree? *J Bone Joint Surg Am.* 2009;91:867-73.

7. American Academy of Orthopaedic Surgeons. Treatment of osteoarthritis of the knee (non-arthroplasty): full guideline. 2008 Dec 6. <http://www.aaos.org/Research/guidelines/OAKguideline.pdf>. Accessed 2009 Apr 5.

8. Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled

trials in orthopaedics. *J Bone Joint Surg Br.* 2001;83:397-402.

9. Obrebsky WT, Pappas N, Attallah-Wasif E, Tometta P 3rd, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am.* 2005;87:2632-8.

10. Levine WN, Clark AM Jr, D'Alessandro DF, Yamaguchi K. Chondrolysis following arthroscopic thermal capsulorrhaphy to treat shoulder instability. A report of two cases. *J Bone Joint Surg Am.* 2005;87:616-21.