

# Have Levels of Evidence Improved the Quality of Orthopaedic Research?

Brian P. Cunningham MD, Samuel Harmsen MD,  
Chris Kweon MD, Jason Patterson MD, Robert Waldrop MD,  
Alex McLaren MD, Ryan McLemore PhD

Received: 11 February 2013 / Accepted: 28 June 2013 / Published online: 12 July 2013  
© The Association of Bone and Joint Surgeons® 2013

## Abstract

**Background** Since 2003 many orthopaedic journals have adopted grading systems for levels of evidence (LOE). It is unclear if the quality of orthopaedic literature has changed since LOE was introduced.

**Questions/purposes** We asked three questions: (1) Have the overall number and proportion of Level I and II studies increased in the orthopaedic literature since the introduction of LOE? (2) Is a similar pattern seen in individual orthopaedic subspecialty journals? (3) What is the interobserver reliability of grading LOE?

**Methods** We assigned LOE to therapeutic studies published in 2000, 2005, and 2010 in eight major orthopaedic subspecialty journals. Number and proportion of Level I and II publications were determined. Data were evaluated using log-linear models. Twenty-six reviewers (13 residents and 13 attendings) graded LOE of 20 blinded therapeutic articles from the *Journal of Bone and Joint Surgery* for 2009. Interobserver agreement relative to the *Journal of Bone and Joint Surgery* was assessed using a weighted kappa.

**Results** The total number of Level I and II publications in subspecialty journals increased from 150 in 2000 to 239 in 2010. The proportion of high-quality publications increased with time ( $p < 0.001$ ). All subspecialty journals other than the *Journal of Pediatric Orthopaedics* and the *Journal of Orthopaedic Trauma* showed a similar behavior. Average weighted kappa was 0.791 for residents and 0.842 for faculty ( $p = 0.209$ ).

**Conclusions** The number and proportion of Level I and II publications have increased. LOE can be graded reliably with high interobserver agreement. The number and proportion of high-level studies should continue to increase.

## Introduction

The focus on evidence-based medicine (EBM) during the last decade has drawn attention to the scientific quality of clinical publications. The concept of interpreting medical data in a systematic, standardized fashion initially was proposed in the early 1970s [6, 22]. The first guidelines specifically designed for grading the quality of medical literature were released in 1986 [19] leading to the formation of the Evidence-based Working Group in the early 1990s. A paradigm shift occurred from “the unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision-making” to the EBM approach that “stresses the examination of evidence from clinical research” [10]. EBM has been refined during the last two decades [20, 22, 25], requiring physicians to gain new skills for evaluation of the clinical literature. Considerable effort is being invested in following the principles of EBM for medical decision-making; however, the quality of those decisions can be only as good as the data they are based on.

---

Work on this study was supported by the Herbert Louis Fund (ACM) at the Orthopaedic Research and Education Foundation and by Banner Good Samaritan Medical Center, Phoenix, AZ, USA. 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.

---

B. P. Cunningham, S. Harmsen, C. Kweon, J. Patterson,  
R. Waldrop, A. McLaren, R. McLemore (✉)  
Orthopaedic Residency, Banner Good Samaritan Medical  
Center, 901 East Willetta Street, Floor 2, Phoenix,  
AZ 85006-2511, USA  
e-mail: ryan.mclemore@gmail.com

To quantify the quality of clinical data, grading systems for the level of evidence (LOE) of clinical studies have been proposed by the Preventative Services Task Force in the United States and by the National Health Service in the United Kingdom [16, 26]. The LOE of clinical data has been used to quantify the strength of recommendation for clinical decisions [23] and to develop clinical practice guidelines [9, 12–14, 18, 30].

In 2003, guidelines for identification of LOE in therapeutic studies were introduced to the orthopaedic literature based on the recommendations from the Centre for Evidence-based Medicine [31]. LOE are assigned to therapeutic studies as follows: Level I is a randomized controlled trial, Level II is a prospective controlled study with some methodologic deficiencies, and Levels III and IV are retrospective studies, Level III with a control and Level IV case series with no control. Level V is expert opinion. There are slight variations to the LOE definitions for prognostic, diagnostic, and economic analyses, which were used in accordance with study design. Using the LOE grading system, previous studies have shown an improvement in the number of Level I, II and III publications during a 30-year period in the *Journal of Bone and Joint Surgery* [11]. A cross-sectional study of numerous orthopaedic articles in 2005 showed that higher LOE studies were associated with a higher journal Impact Factor [17]. Orthopaedic surgery is composed of subspecialties, each with a National Society that promotes education, research, and clinical excellence. The LOE of therapeutic studies published in orthopaedic subspecialty journals has not been studied.

The purpose of our study is to quantitatively analyze the LOE of therapeutic studies published in eight major orthopaedic subspecialty journals before, around, and after the introduction of LOE to orthopaedic literature. These journals did not list LOE for all issues during this period. Some of the journals introduced grading systems between 2005 and 2010, however because of the variation between grading systems of individual journals, all 2010 articles were graded similarly to previous years to maintain a consistent classification system. We asked three questions; (1) Have the overall number and proportion of Level I and II studies in orthopaedic literature in eight subspecialty journals increased since the introduction of LOE? (2) Is a similar pattern seen in each specific subspecialty journal? (3) What is the interobserver reliability of grading LOE?

## Materials and Methods

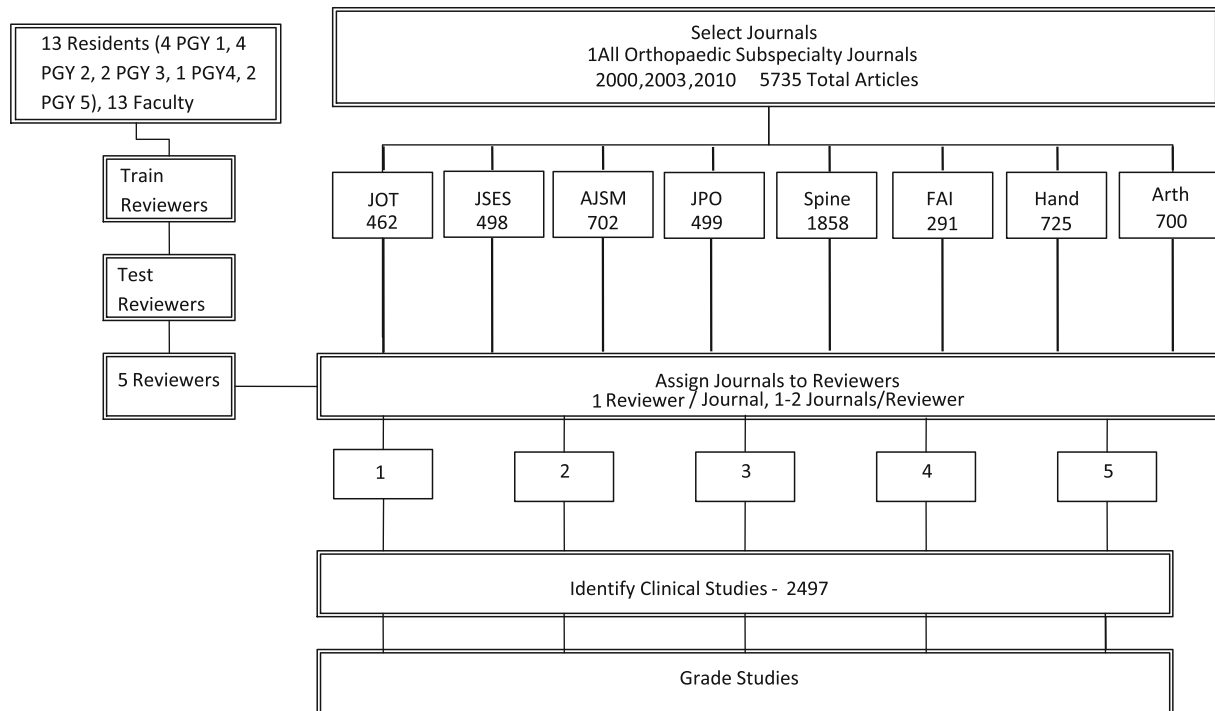
We sought to determine changes in publication behavior during 10 years. Eight subspecialty journals were identified. These journals were reviewed by five of the authors (BPC, SH, CK, JP, RW), with each author reviewing one to two journals (Fig. 1). Since the journals neither consistently

listed the LOE for each year nor used a consistent grading system, the LOE for each therapeutic publication in each journal was graded from 2000, 2005, and 2010 using a previously published LOE grading system [31]. Each journal was examined to determine if the total number and proportion of high-quality studies (Levels I and II) were changing with time. A subgroup analysis then was performed to compare the relative changes among the different journals. To ensure LOE assigned by the reviewers lacked subjectivity and was relatively accurate, interobserver reliability was assessed. Twenty-six reviewers with various levels of experience reading orthopaedic literature evaluated 20 blinded studies from the *Journal of Bone and Joint Surgery* and assigned a LOE to each article. The LOE assigned to the articles by each reviewer was compared with that assigned by the journal to determine the reviewer's relative accuracy in determining the appropriate LOE.

Eight major orthopaedic subspecialty journals were included in this review: (1) *Journal of Orthopaedic Trauma*; (2) *Journal of Shoulder and Elbow Surgery*; (3) *American Journal of Sports Medicine*; (4) *Journal of Pediatric Orthopaedics*; (5) *Foot & Ankle International*; (6) *Journal of Hand Surgery (American)*; (7) *Journal of Arthroplasty*; and (8) *Spine*. These journals represent most of the major orthopaedic subspecialties with the exception of orthopaedic oncology, which does not have a dedicated journal.

To assess trends, all articles published in 2000, 2005, and 2010 in each subspecialty journal selected were reviewed. All animal, cadaveric, and basic science studies, review papers, case reports, and expert opinions were excluded [17]. These studies were excluded because they are not therapeutic studies, even though they can provide information useful to clinical practice. Therapeutic studies then were graded for LOE by five reviewers (BPC, SH, CK, JP, RW) [31]. Each journal was assigned to one reviewer to minimize the chance for bias by avoiding different reviewers analyzing different years in the same journal. The ability of each reviewer to recognize appropriate LOE was emphasized to ensure accuracy and reliability throughout the grading process. The reviewers selected were orthopaedic residents in their third or fourth year of training that scored an average kappa of  $0.83 \pm 0.15$  compared with the LOE grades assigned by the journal in the interobserver agreement portion of the study.

The change in the number and proportion of Levels I and II studies in the eight subspecialty journals was assessed by first pooling the identifications made by each of the five reviewers in their assigned journals. This generated a count of published Level I, II, III, and IV studies in all journals in each year evaluated. These counts then were analyzed using a log-linear model which used journal, year, and LOE as



**Fig. 1** Our study methods are shown. PGY = postgraduate year; JOT = *Journal of Orthopaedic Trauma*; FAI = *Foot and Ankle International*; JHS = *Journal of Hand Surgery*; Arth = *Journal of*

*Arthroplasty*; JSES = *Journal of Shoulder and Elbow Surgery*; AJSM = *American Journal of Sports Medicine*; JPO = *Journal of Pediatric Orthopaedics*.

factors to explain the observed change in the frequency of articles of each LOE with time. To look at the change in total number of high-quality publications the model included only the Level I and II count data. To look at the change in proportion of high-quality publications, the relative change between pooled counts of Levels I and II versus Levels III and IV was assessed using the same model. Log-linear models were constructed and analyzed using CATMOD procedure in SAS 9.2(SAS Institute, Cary, NC, USA), which fits log-linear models to tables of count data.

To examine whether the observed change in behavior was consistent across each subspecialty journal evaluated, an individual log-linear model was constructed for the counts of number of high-level (I and II) and low-level (III and IV) articles for each journal. This model used year and LOE as factors. Statistical results then were compared among journals to determine if the behaviors were consistent. This analysis also used CATMOD in SAS 9.2.

Thirteen orthopaedic residents (Postgraduate Years 1–5) and 13 attending faculty, orthopaedic surgeons (3–30 years in practice), independently reviewed the LOE guidelines and author submission instruction on the *Journal of Bone and Joint Surgery* website. Reviewers then were tested by assigning a LOE grade to 20 randomly ordered therapeutic studies published in 2009 in the *Journal of Bone and Joint Surgery*. These 20 articles included five therapeutic studies for each LOE (I–IV), as assigned by the *Journal of Bone*

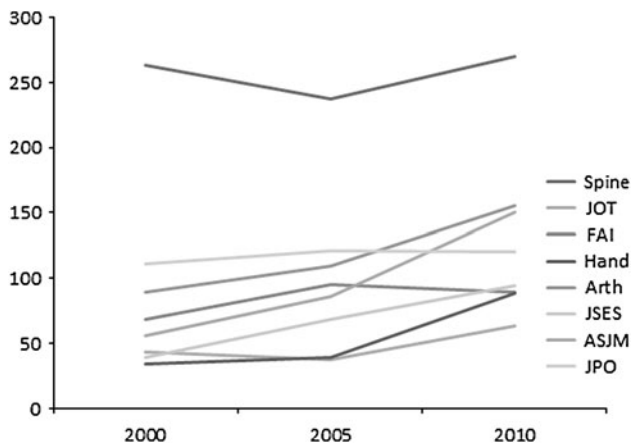
*and Joint Surgery*. All reviewers were given access to the full manuscripts but were blinded to the LOE of each article, which was blacked out. Grading was performed under controlled conditions. The accuracy of the LOE assigned by each reviewer was determined by calculating weighted kappa for each reviewer, relative to the LOE assigned to the 20 articles by the *Journal of Bone and Joint Surgery* [1], using Microsoft Excel (Microsoft, Redmond, WA, USA). Weights were assigned based on degree of error in identification (Table 1). To provide context to weighted kappa scores, overall agreement is reported. Finally, weighted kappa between residents and faculty and raters and faculty were compared using the Mann-Whitney U test in Minitab<sup>®</sup> 15 (Minitab Inc, State College, PA, USA).

## Results

Five thousand seven hundred thirty-five articles were reviewed with 2479 studies that met the inclusion criteria. The total number of articles published has increased during the 10-year period (Fig. 2). The total number of Level I and II publications significantly increased across the 10-year period from 150 in 2000, to 221 in 2005, to 239 in 2010 ( $p < 0.001$ ). Total numbers of Level III and IV studies also increased (Table 2). The change in overall proportion of

**Table 1.** Weights used in calculating weighted kappa

Correct score	Rater score			
	I	II	III	IV
I	1	0.66	0.33	0
II	0.66	1	0.33	0
III	0	0.33	1	0.66
IV	0	0.33	0.66	1



**Fig. 2** The changes in clinical study volume for each subspecialty journal between 2000 and 2010 are shown. JOT = *Journal of Orthopaedic Trauma*; FAI = *Foot and Ankle International*; JHS = *Journal of Hand Surgery*; Arth = *Journal of Arthroplasty*; JSES = *Journal of Shoulder and Elbow Surgery*; AJSM = *American Journal of Sports Medicine*; JPO = *Journal of Pediatric Orthopaedics*.

high-quality publications over time significantly increased from 22% in 2000, to 28% in 2005, to 23% in 2010 ( $p < 0.001$ ) (Fig. 3).

The total number of articles published in all subspecialty journals also increased with time (Fig. 2). Regarding change in total number of high-quality publications with time, the behavior in *Spine*, *Journal of Shoulder and Elbow Surgery*, *Journal of Hand Surgery*, and *Journal of Arthroplasty* were similar to the overall analysis: the total number of high-quality publications increased at every period. Total numbers of high-quality articles published in 2010 by *American Journal of Sports Medicine* and *Foot and Ankle International* were greater than in 2000, but were fewer than in 2005. The numbers of high-quality publications in the *Journal of Pediatric Orthopaedics* and *Journal of Orthopaedic Trauma* remained relatively unchanged throughout (28 in 2000, 24 in 2005, 24 in 2010; seven in 2000, four in 2005, and four in 2010, respectively). For changes in proportion of high-quality publications with time, the behavior of *Spine*, *Journal of Shoulder and Elbow Surgery*, *Journal of Arthroplasty*, *Journal of Hand Surgery*, *American Journal of Sports Medicine*, and *Foot and Ankle International* was

similar to the overall analysis ( $p < 0.001$ ,  $p < 0.001$ ,  $p < 0.001$ ,  $p < 0.001$ ,  $p < 0.001$ ,  $p < 0.001$ ,  $p < 0.001$ ,  $p = 0.09$ ), showing an increase with time. Although the change seen in *Foot and Ankle International* was not statistically significant, the proportion of high-quality studies changed from 18% to 33% to 24%, revealing a positive trend. The *Journal of Pediatric Orthopaedics* did not show significant changes ( $p = 0.77$ ) in behavior (26% to 19% to 20%). The *Journal of Orthopaedic Trauma* did show a significant change in behavior ( $p = 0.02$ ), however, the primary change seen was in increase the proportion of Level III and IV publications with time (84% to 89% to 97%) (Fig. 4).

The weighted kappa values were 0.791 for the orthopaedic residents and 0.842 for the attending surgeons. Residents and faculty had similar accuracy ( $p = 0.209$ , Mann-Whitney test). The subgroup of residents which rated the articles in the study also were not different from faculty ( $p = 0.6221$ , Mann-Whitney test). Residents correctly determined the LOE for  $77\% \pm 13\%$  of the 20 selected manuscripts (15.4/20 papers). Attending surgeons determined the correct LOE  $81\% \pm 13\%$  of the selected manuscripts (16.1/20 papers). Of the 260 article identifications performed by the 13 residents, 200 were performed correctly (77%). Only 1% (three articles) would have led to a Level I or II study being incorrectly identified as a Level III or IV study. Approximately 5% (13 articles) would have led to a Level III or IV article being identified as a Level I or II study. Of the 260 identifications performed by the 13 faculty members, 217 were identified correctly (83%). Of the identifications, 0.3% (one article) classified a Level I or II as a Level III or IV, and 4% (11 articles) classified a Level III or IV as a Level I or II.

## Discussion

Efforts to evaluate the quality of therapeutic studies published in medical and surgical specialty journals have been reported [3, 5, 11, 17, 24, 32]. The primary purpose of our study was to quantitatively analyze the orthopaedic subspecialty literature before, around, and after the introduction of LOE to orthopaedics to determine if the overall number and proportion of high-quality publications (Levels I and II) had improved. Additionally, we wanted to determine if a similar behavior was seen in each specific subspecialty journal. Finally we wanted to evaluate the interobserver reliability of grading LOE.

Our study has some limitations. First, LOE is not a comprehensive scale for research quality. It primarily reflects the inclusion of randomization, cohort studies, and control groups. Not all Level I studies are equally good or provide equally good evidence. High-quality articles (Levels I and II), however, generally do include some randomization, and Level III studies include control groups; including these

**Table 2.** Results

Journal	Level of evidence				Year total	Journal total
	I	II	III	IV		
<b>Spine</b>						
2000	19	42	47	128	236	743
2005	33	50	56	98	237	
2010	30	62	75	103	270	
<b>JOT</b>						
2000	5	2	11	25	43	143
2005	1	3	14	19	37	
2010	3	1	16	43	63	
<b>FAI</b>						
2000	2	10	15	41	68	252
2005	3	28	16	48	95	
2010	4	17	26	42	89	
<b>Hand</b>						
2000	0	4	7	23	34	161
2005	1	5	8	25	39	
2010	5	5	17	61	88	
<b>Arth</b>						
2000	6	12	5	66	89	353
2005	13	13	10	73	109	
2010	14	18	24	99	155	
<b>JSES</b>						
2000	4	2	2	31	39	201
2005	10	5	1	52	68	
2010	12	14	19	49	94	
<b>AJSM</b>						
2000	3	10	11	32	56	292
2005	12	20	12	42	86	
2010	10	20	46	74	150	
<b>JPO</b>						
2000	3	26	30	52	111	352
2005	2	22	37	60	121	
2010	0	24	39	57	120	
<b>Totals</b>						
2000	42	108	128	398	676	2497
2005	75	146	154	417	792	
2010	78	161	262	528	1029	

JOT = *Journal of Orthopaedic Trauma*; FAI = *Foot and Ankle International*; JHS = *Journal of Hand Surgery*; Arth = *Journal of Arthroplasty*; JSES = *Journal of Shoulder and Elbow Surgery*; AJSM = *American Journal of Sports Medicine*; JPO = *Journal of Pediatric Orthopaedics*.

practices in therapeutic publications is likely to lead to better, more reliable recommendations. Second, articles in journals that do not directly report the LOE on each article were rated by one observer. Thus, we cannot study either the validity of the rating or the validity of LOE as a rating system. Errors made by these observers could be reflected in the counts of articles analyzed for the study data. Based on the high

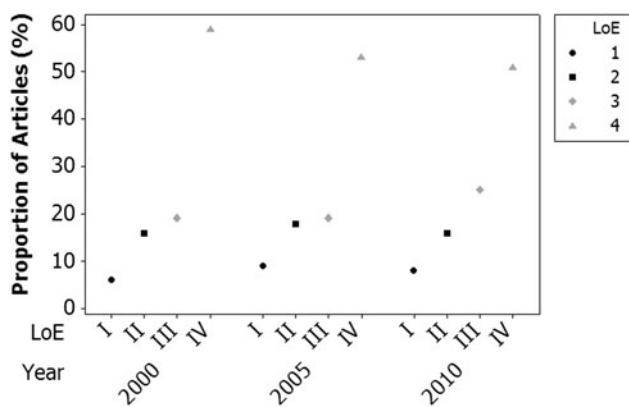
interobserver reliability in our study and others [2, 4, 17, 32], it is unlikely that this had a significant effect on our conclusions. Third, not all of the journals behaved similarly. Fourth, only three discrete years were chosen for review, introducing possible sampling bias. Fifth, it is not necessarily obvious what the correct ratio of studies should be. Generally, higher-level studies are required to answer comparative

effectiveness questions and address issues where biologic variability plays a large role in the outcome. It was expected, however, that recognition of the value and need for higher-level studies would in turn lead to a larger number and proportion of high-level studies being done.

The results of our study show that there have been increases in total number and proportion of high-quality studies in the orthopaedic literature. Other areas of medicine have shown similar trends [8, 15, 29, 33]. Recognition of the importance of high-quality research in performing reliable therapeutic studies seems to have produced a change in behavior. However, while the proportion of these studies has improved significantly with time, there remains a predominance of Level III and IV studies. This finding is consistent with results from other studies in the orthopaedic literature [4, 17, 27, 32]. Although there is value in these lower-quality studies, higher-quality studies generally are

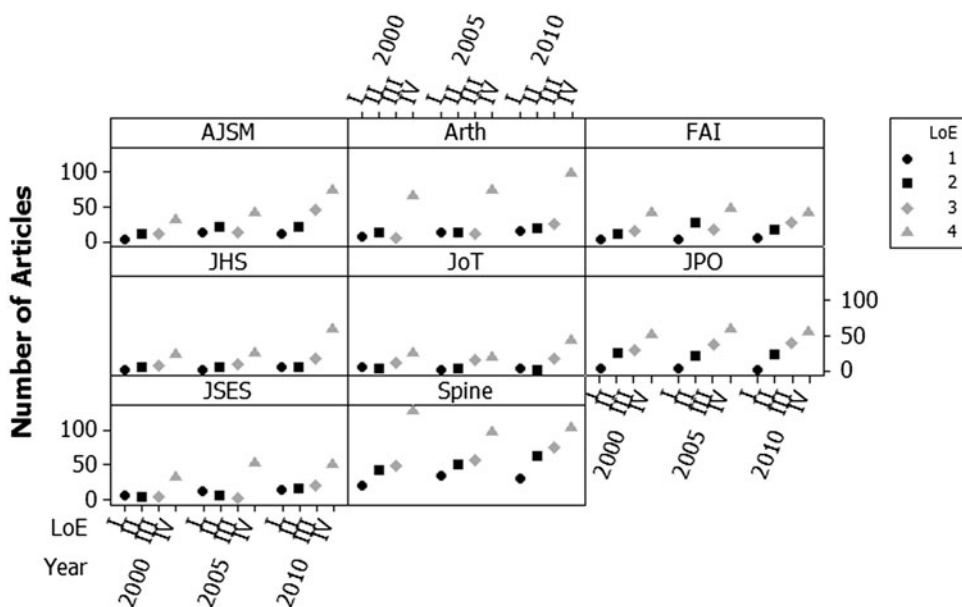
better at removing bias, isolating confounders, and providing reliable data for further clinical practice. As such, it benefits clinicians, researchers, and patients to attempt to produce the highest-quality data possible to answer questions related to clinical practice. Voleti et al. [27] witnessed a similar behavior when they evaluated papers and posters presented at the American Academy of Orthopaedic Surgeons (AAOS) annual meetings from 2001, 2003, 2007, and 2010. They saw an increase in the proportion of Level I and II presentations from 2001 to 2010 (17% to 36%). However, Level III and IV presentations continued to outnumber those of high quality in 2010 (824 to 465). Donegen et al. [7] reported that AAOS presentations have only a 49% publication rate in peer-reviewed journals after 5 years. Voleti et al. [28] found that at 5 years, 72% of Level I and II presentations and 46% of Level III and IV presentations were published. This would suggest that the proportion of high-quality publications will further increase with time. However, when translated to number of publications, this represented 44 high-quality articles and 105 lesser-quality papers [28]. Despite the AAOS Annual Meeting generally representing the current pulse of orthopaedic research, presentations made there do not represent the entirety of orthopaedic research and literature.

Similar behavior is seen in different subspecialty journals; however, not all showed significant trends toward improvement. The *Journal of Pediatric Orthopaedics* and *Journal of Orthopaedic Trauma* have seen little change in the number of Level I and II studies published during the last 10 years and a decrease in the proportion of high-quality studies. Cashin et al. [4] found similar trends in preLOE (before 2003) to postLOE (after 2003)



**Fig. 3** The overall proportions of level of evidence in publications in 2000, 2005, and 2010 are shown. LOE = level of evidence.

**Fig. 4** The numbers of publications of different levels of evidence by journal in 2000, 2005, and 2010 are shown. Data are plotted as number rather than proportion because of differences in publication volumes between subspecialties. Abbreviations are: LOE = level of evidence; JOT = *Journal of Orthopaedic Trauma*; FAI = *Foot and Ankle International*; JHS = *Journal of Hand Surgery*; Arth = *Journal of Arthroplasty*; JSES = *Journal of Shoulder and Elbow Surgery*; AJSM = *American Journal of Sports Medicine*; JPO = *Journal of Pediatric Orthopaedics*.



publications in the pediatric orthopaedic subspecialty [4]. They also noted that the proportion of high-quality LOE decreased from 11% to 9%. They attribute this in part to the difficulties encountered when enrolling children as study participants. This also could be a result of the high number of pediatric disorders with low prevalence. Orthopaedic trauma also has the constraints of enrolling patients owing to the urgent and often emergent nature of the disorders. Despite the apparent general trend toward higher-quality research, the predominance of Level III and IV research also holds true at the subspecialty level. AAOS practice guidelines for common orthopaedic conditions such as carpal tunnel syndrome, symptomatic osteoporotic spinal compression fractures, and distal radius fractures have few recommendations based on Level I or II data, and rely on Level III, IV, and V evidence [9, 12, 14]. Although treatment recommendations for other conditions such as osteoarthritis of the knee are based on a larger number of high-quality studies, the low prevalence of Level I and II studies remains evident [18].

Finally, we looked at the interobserver reliability of grading LOE using the *Journal of Bone and Joint Surgery* grading system. Using weighted kappa values, we found that residents and attendings had a substantial or better agreement with the LOE grading performed by the journal. It generally is accepted that kappa values greater than 0.5 indicate a high level of agreement [17, 21, 32]. An accurate identification rate (77% and 83%), with a low incidence of misclassification (4% to 5% leading to an inappropriate increase in quality) indicates that LOE is a concept with which faculty and residents are conversant. Our results are consistent with those of other studies that show substantial or better agreement when grading LOE, regardless of experience [2, 4, 17, 32]. In contrast, Schmidt et al. [21] found substantial inconsistency and poor agreement between different observers when grading the LOE of presentations at the 2007 AAOS Annual Meeting. In our study, reviewers had access to the entire published manuscript, whereas in their study reviewers often had access to the AAOS abstract alone, which was limited to 250 words. Furthermore, our study includes a substantially larger number of reviewers than other studies reviewed, and is the first study that we are aware of in orthopaedic literature to compare the ability of orthopaedic residents and attending surgeons to grade LOE.

Our study established that the quality and volume of orthopaedic literature has increased during the last 10 years across all eight studied subspecialties. Despite these changes, much of the literature still consists of Levels III and IV studies. Although high-level data may not be necessary to answer every clinical question, the field should attempt to produce the highest-quality evidence, improving the strength of recommendation of the resulting work. Finally, orthopaedic residents and attending surgeons can

consistently differentiate high- and low-level studies with high interobserver reliability.

**Acknowledgments** We thank Richard Gerkin MD, for assistance with the statistical analysis of the data and discussion of model results.

## References

1. Altman DG. *Practical Statistics for Medical Research*. 2nd ed. London, England: Chapman & Hall/CRC; 2006.
2. Bhandari M, Swiontkowski MF, Einhorn TA, Tornetta P 3rd, Schemitsch EH, Leece P, Sprague S, Wright JG. Interobserver agreement in the application of levels of evidence to scientific papers in the American volume of the *Journal of Bone and Joint Surgery*. *J Bone Joint Surg Am*. 2004;86:1717–1720.
3. Brophy RH, Gardner MJ, Saleem O, Marx RG. An assessment of the methodological quality of research published in *The American Journal of Sports Medicine*. *Am J Sports Med*. 2005;33:1812–1815.
4. Cashin MS, Kelley SP, Douziech JR, Varghese RA, Hamilton QP, Mulpuri K. The levels of evidence in pediatric orthopaedic journals: where are we now? *J Pediatr Orthop*. 2011;31:721–725.
5. Chang EY, Pannucci CJ, Wilkins EG. Quality of clinical studies in aesthetic surgery journals: a 10-year review. *Aesthet Surg J*. 2009;29:144–147; discussion 147–149.
6. Cochrane AL. Archie Cochrane in his own words. Selections arranged from his 1972 introduction to “Effectiveness and Efficiency: Random Reflections on the Health Services” 1972. *Control Clin Trials*. 1989;10:428–433.
7. Donegan DJ, Kim TW, Lee GC. Publication rates of presentations at an annual meeting of the American Academy of Orthopaedic Surgeons. *Clin Orthop Relat Res*. 2010;468:1428–1435.
8. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine. *Lancet*. 1995;346:407–410.
9. Esses SI, McGuire R, Jenkins J, Finkelstein J, Woodard E, Watters WC 3rd, Goldberg MJ, Keith M, Turkelson CM, Wies JL, Sluka P, Boyer KM, Hitchcock K. The treatment of symptomatic osteoporotic spinal compression fractures. *J Am Acad Orthop Surg*. 2011;19:176–182.
10. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268:2420–2425.
11. Hanzlik S, Mahabir RC, Baynosa RC, Khiabani KT. Levels of evidence in research published in *The Journal of Bone and Joint Surgery (American Volume)* over the last thirty years. *J Bone Joint Surg Am*. 2009;91:425–428.
12. Keith MW, Masear V, Amadio PC, Andary M, Barth RW, Graham B, Chung K, Maupin K, Watters WC 3rd, Haralson RH 3rd, Turkelson CM, Wies JL, McGowan R. Treatment of carpal tunnel syndrome. *J Am Acad Orthop Surg*. 2009;17:397–405.
13. Kuhn JE. Exercise in the treatment of rotator cuff impingement: a systematic review and a synthesized evidence-based rehabilitation protocol. *J Shoulder Elbow Surg*. 2009;18:138–160.
14. Lichtman DM, Bindra RR, Boyer MI, Putnam MD, Ring D, Slutsky DJ, Taras JS, Watters WC 3rd, Goldberg MJ, Keith M, Turkelson CM, Wies JL, Haralson RH 3rd, Boyer KM, Hitchcock K, Raymond L. Treatment of distal radius fractures. *J Am Acad Orthop Surg*. 2010;18:180–189.
15. Loiselle F, Mahabir RC, Harrop AR. Levels of evidence in plastic surgery research over 20 years. *Plast Reconstr Surg*. 2008;121:207e–211e.
16. NHS National Institute for Health and Clinical Excellence. Methods for development of NICE public health guidance. Available at:

- <http://www.nice.org.uk/nicemedia/pdf/CPHEMethodsManual.pdf>. Accessed June 20, 2013.
17. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P 3rd, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am*. 2005;87:2632–2638.
  18. Richmond J, Hunter D, Irrgang J, Jones MH, Levy B, Marx R, Snyder-Mackler L, Watters WC 3rd, Haralson RH 3rd, Turkelson CM, Wies JL, Boyer KM, Anderson S, St Andre J, Sluka P, McGowan R; American Academy of Orthopaedic Surgeons. Treatment of osteoarthritis of the knee (nonarthroplasty). *J Am Acad Orthop Surg*. 2009;17:591–600.
  19. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1986;89(2 suppl):2S–3S.
  20. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
  21. 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;91:867–873.
  22. Schwartz WB, Gorry GA, Kassirer JP, Essig A. Decision analysis and clinical judgment. *Am J Med*. 1973;55:459–472.
  23. Thornton J, Alderson P, Tan T, Turner C, Latchem S, Shaw E, Ruiz F, Reken S, Mugglestone MA, Hill J, Neilson J, Westby M, Francis K, Whittington C, Siddiqui F, Sharma T, Kelly V, Ayiku L, Chamberlain K. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol*. 2013;66:124–131.
  24. Turpen RM, Fesperman SF, Sultan S, Borawski KM, Norris RD, Klink J, Sur RL, Breau RH, Krupski TL, Dahm P. Levels of evidence ratings in the urological literature: an assessment of interobserver agreement. *BJU Int*. 2010;105:602–606.
  25. Ubbink DT, Legemate DA. Evidence-based surgery. *Br J Surg*. 2004;91:1091–1092.
  26. US Preventive Services Task Force. *Guide to Clinical Preventive Services: Report of the US Preventive Services Task Force*. Collingdale, PA: DIANE Publishing; 1989.
  27. Voleti PB, Donegan DJ, Baldwin KD, Lee GC. Level of evidence of presentations at American Academy of Orthopaedic Surgeons annual meetings. *J Bone Joint Surg Am*. 2012;94:e50.
  28. Voleti PB, Donegan DJ, Kim TWB, Lee GC. Level of evidence: does it change the rate of publication and time to publication of American Academy of Orthopaedic Surgeons presentations? *J Bone Joint Surg Am*. 2013;95:e2.
  29. Wasserman JM, Wynn R, Bash TS, Rosenfeld RM. Levels of evidence in otolaryngology journals. *Otolaryngol Head Neck Surg*. 2006;134:717–723.
  30. Wingfield K, Matheson GO, Meeuwisse WH. Preparticipation evaluation: an evidence-based review. *Clin J Sport Med*. 2004; 14:109–122.
  31. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85:1–3.
  32. Wupperman R, Davis R, Obremskey WT. Level of evidence in Spine compared to other orthopedic journals. *Spine (Phila Pa 1976)*. 2007;32:388–393.
  33. Xu CC, Côté DW, Chowdhury RH, Morrissey AT, Ansari K. Trends in level of evidence in facial plastic surgery research. *Plast Reconstr Surg*. 2011;127:1499–1504.