

What Are the Levels of Evidence on Which We Base Decisions for Surgical Management of Lower Extremity Bone Tumors?

Nathan Evaniew MD, James Nuttall MBBChBAO,
Forough Farrokhyar MPhil, PhD, Mohit Bhandari MD, PhD,
Michelle Ghert MD

Received: 22 June 2013 / Accepted: 19 September 2013 / Published online: 1 October 2013
© The Association of Bone and Joint Surgeons® 2013

Abstract

Background Benign and malignant lower extremity primary bone tumors are among the least common conditions treated by orthopaedic surgeons. The literature supporting their surgical management has historically been in the form of observational studies rather than prospective controlled studies. Observational studies are prone to confounding bias, sampling bias, and recall bias.

Each author certifies that he or she, or a member of his or her immediate family, has no funding or commercial associations (eg, 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. This work was performed at McMaster University, Hamilton, Ontario, Canada.

N. Evaniew, M. Bhandari
Division of Orthopaedics, Department of Surgery, McMaster University, 293 Wellington Street North, Suite 110, Hamilton, ON L8L 8E7, Canada

J. Nuttall, M. Ghert (✉)
Division of Orthopaedics, Department of Surgery, McMaster University, 711 Concession Street, Hamilton, ON L8V 1C3, Canada
e-mail: Michelle.Ghert@jcc.hhsc.ca

F. Farrokhyar
Department of Clinical Epidemiology and Biostatistics, McMaster University, 39 Charlton Ave. East, Hamilton, ON L8N 1Y3, Canada

M. Bhandari
Department of Clinical Epidemiology and Biostatistics, McMaster University, 293 Wellington Street North, Suite 110, Hamilton, ON L8L 8E7, Canada

Questions/purposes (1) What are the overall levels of evidence of articles published on the surgical management of lower extremity bone tumors? (2) What is the overall quality of reporting of studies in this field based on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist? (3) What are the most common pitfalls in reporting that authors might improve on?

Methods All studies describing the surgical management of lower extremity primary bone tumors from 2002 to 2012 were systematically reviewed. Two authors independently appraised levels of evidence. Quality of reporting was assessed with the STROBE checklist. Pitfalls in reporting were quantified by determining the 10 most underreported elements of research study design in the group of studies analyzed, again using the STROBE checklist as the reference standard. Of 1387 studies identified, 607 met eligibility criteria.

Results There were no Level I studies, two Level II studies, 47 Level III studies, 308 Level IV studies, and 250 Level V studies. The mean percentage of STROBE points reported satisfactorily in each article as graded by the two reviewers was 53% (95% confidence interval, 42%–63%). The most common pitfalls in reporting were failures to justify sample size (2.2% reported), examine sensitivity (2.2%), account for missing data (9.8%), and discuss sources of bias (14%). Followup (66%), precision of outcomes (64%), eligibility criteria (55%), and methodological limitations (53%) were variably reported.

Conclusions Observational studies are the dominant evidence for the surgical management of primary lower extremity bone tumors. Numerous deficiencies in reporting limit their clinical use. Authors may use these results to inform future work and improve reporting in observational studies, and treating surgeons should be aware of these limitations when choosing among the various options with their patients.

Introduction

Evidence-based medicine is the conscientious, explicit, and judicious integration of best available scientific evidence with clinical expertise and patient values to facilitate clinical decision-making [27]. Although well-designed randomized controlled trials (RCTs) and meta-analyses of RCTs sit atop the hierarchy of medical evidence, observational studies dominate the orthopaedic surgery literature and despite their limitations have some value for treatment decisions [6, 14, 15]. In fact, they dominate most of the medical literature and until very recently have formed the basis for most therapeutic decisions.

Common observational study designs include cohorts, case series, case-controls, and cross-sectional studies. Observational studies inform clinicians about disease etiology, natural history, prognostic factors, and treatment effectiveness, but they are prone to confounding bias, sampling bias (including selection bias), and recall bias [30]. Confounding bias is the possibility of an unknown variable being causally linked to both an exposure and an outcome, whereas sampling bias is a systematic error in participant selection that may inadequately represent a population of interest, and recall bias is the inaccuracy with which test subjects retrospectively recall prior events or health states. Transfer bias, or unequal loss to followup between groups, also is a common limitation in observational studies. These biases limit the validity and applicability of results from observational studies [30].

Benign and malignant bone tumors are among the least common conditions treated by orthopaedic surgeons. Osteosarcoma, for example, occurs with a prevalence of four to five patients per million each year [22]. The largest tumor centers in the world may treat only 25 to 30 patients with lower extremity bone sarcomas per year. A formal assessment of the quality of the evidence that guides the management of primary bone tumors, to our knowledge, has not previously been published. Such an assessment may be important, because the surgical literature includes many more observational studies than studies of other designs, a limitation that reflects the difficulties of conducting randomized trials in surgical patients [9].

The purpose of this systematic review was to answer the following questions: (1) What are the overall levels of evidence of articles published on the surgical management of lower extremity bone tumors? (2) What is the overall quality of reporting of studies in this field based on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist? (3) What are the most common pitfalls in reporting on which authors might improve? This study focused on lower extremity bone tumors because the orthopaedic oncology field is vastly heterogeneous in the type of surgical cases managed and therefore this study can act as a feasible starting point for further systematic reviews in the field.

Materials and Methods

We conducted a systematic review of all therapeutic studies reporting on the surgical management of lower extremity primary bone tumors from 2002 to 2012. This study was performed according to the Quality of Reporting of Meta-analyses (QUOROM) guidelines [18].

Search Strategy and Criteria

MEDLINE, EMBASE, and the Cochrane databases were systematically searched with the assistance of a health sciences librarian (MF) experienced in systematic reviews. MeSH and Emtree headings and subheadings were used in various combinations in OVID (ie, [exp bone neoplasms/su AND exp leg bones/] OR [exp bone neoplasms/AND exp leg bones/su]). These terms were supplemented with free text to increase sensitivity (ie, “bone” and [“cancer” or “tumor” or “tumour” or “neoplasm” or “osteosarcoma”] and [“surgery” or “surgical*” or “operat*” or “reoperat*” or “fixa*” or “endoprosth*”). Free-text terms were reviewed by an orthopaedic oncologist experienced in clinical research (MG). Results were limited to humans and the publication dates January 1, 2002, to December 17, 2012. These dates were selected to capture an approximately 10-year interval of the most current literature. The end date was extended past 10 years at the time of the literature search to include the most recently published articles on this topic.

Eligibility Criteria

All clinical studies of surgical interventions for lower extremity primary bone tumors in living humans of any age published from 2002 to 2012 were included. Upper extremity, spine, and craniofacial tumors were excluded to reduce heterogeneity. Studies of pelvic cases were excluded. Studies including lower extremity bone tumors in addition to bone tumors from other anatomic locations were included if their lower extremity data were reported separately. Studies that reported on diseases in addition to bone tumors (ie, infections) were included if their lower extremity bone tumor was reported separately. Animal and cadaveric studies were excluded. Studies of chemotherapy and/or radiotherapy were included only if patients also underwent surgery and the surgical interventions were analyzed separately. Otherwise, studies of chemotherapy and/or radiotherapy were excluded. Studies of perioperative care were included if patients underwent surgery. To characterize the entirety of the evidence describing the

surgical management of bone tumors, all types of surgical interventions were included. Studies of image-guided radiofrequency ablation were considered nonsurgical. Studies with purely radiographic, biomechanical, or histological outcomes were excluded. Nontherapeutic study designs (diagnostic, economic, prognostic) were excluded. Articles that did not present original clinical evidence such as letters to the editor, commentaries, expert opinions, and narrative reviews were also excluded. Original case reports were included. Studies of benign and malignant bone tumors were included, but studies of bony metastases from nonbone primaries and studies of soft tissue tumors that locally invaded bone were excluded. Studies were excluded if their full text was not available in English. Duplicate articles were excluded. We selected the interval 2002 to 2012 to focus on current trends in the literature.

Study Selection

Two independent reviewers (NE, JN) screened all titles and abstracts for eligibility using a piloted computerized Microsoft Excel database (Santa Rosa, CA, USA). All discrepancies were resolved by consensus. Duplicates were removed by automation in OVID and by manual review. All eligible studies underwent subsequent full-text review. Reviewers were not blinded to authors or publication information.

Levels of Evidence

The two reviewers independently assigned levels of evidence to each eligible study using the Centre for Evidence-Based Medicine in Oxford guidelines for therapeutic studies [36]. The reviewers used a sample of five studies graded by the senior author (MG) for training. All discrepancies were resolved by consensus or by review with the senior author when consensus was not reached. If a study was not clearly prospective or retrospective, authors were contacted by email for clarification. Reviewers were not blinded to authors, publication information, or any published level of evidence descriptions. Retrospective controlled studies were graded as Level III and retrospective uncontrolled studies were classified as “case series” and graded Level IV [36]. Case reports were included in the analysis and graded as Level V [13, 38].

Quality of Reporting

To address quality of reporting, we used the Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE) checklist. The STROBE checklist was developed by a multinational group of methodologists, researchers, and journal editors to guide authors of observational studies [35]. Whereas the reporting of observational studies is often not clear enough to allow assessment of strengths and weaknesses, STROBE specifically considers sources of bias and clarifies the interpretation and applicability of a study’s results and conclusions [34]. It has been adopted by many top journals and is readily available to authors, editors, and reviewers [5, 30]. Because the STROBE checklist can only be applied to observational study designs that have comparative groups, we examined the quality of reporting of the prospective and retrospective controlled observational studies but excluded the retrospective uncontrolled studies (case series), case reports, and any RCTs from analysis with STROBE. RCTs were evaluated with the Detsky Quality Assessment Scale [8].

The STROBE checklist contains 22 numbered items, but some items have multiple subitems [35]. There is no differential weighting of items or subitems, and there is no hierarchy of item importance. In total, there are 34 items that can be evaluated for each study. Under the supervision of a senior investigator (MG), two independent reviewers (NE, JN) learned to use the STROBE checklist by reviewing a published explanatory article and a published practical application guide [30, 35]. The published application guide was used to define the criteria required to score each STROBE item as adequately reported [30, 35]. The two reviewers then assessed each of the eligible studies for each of the 34 points and scored whether each point was satisfactorily reported. The numbers of studies satisfactorily reporting each STROBE point as evaluated by each reviewer were summed and converted to overall percentage scores.

The initial search identified 1387 potential studies. After removal of duplicates, 1123 studies remained eligible. Screening of titles and abstracts eliminated 506 studies (Fig. 1). Agreement between the two reviewers for eligibility was satisfactory ($\kappa = 0.838$). Upon screening the remaining 617 studies in full text, a further 10 were excluded; three were not available as full texts in English, four were not therapeutic designs, two were published before 2002, and one was a duplicate. A total of 607 studies were included for further analysis (Fig. 1).

Statistical Analysis

Cohen’s kappa coefficient of agreement was calculated for the reviewers’ assessments of study eligibility. A weighted kappa coefficient was calculated for the reviewers’ assessments of levels of evidence. Kappa values of > 0.65 were considered adequate [26]. The STROBE items were then ranked by percentage scores of reporting with the most underreported items having lower percentage scores

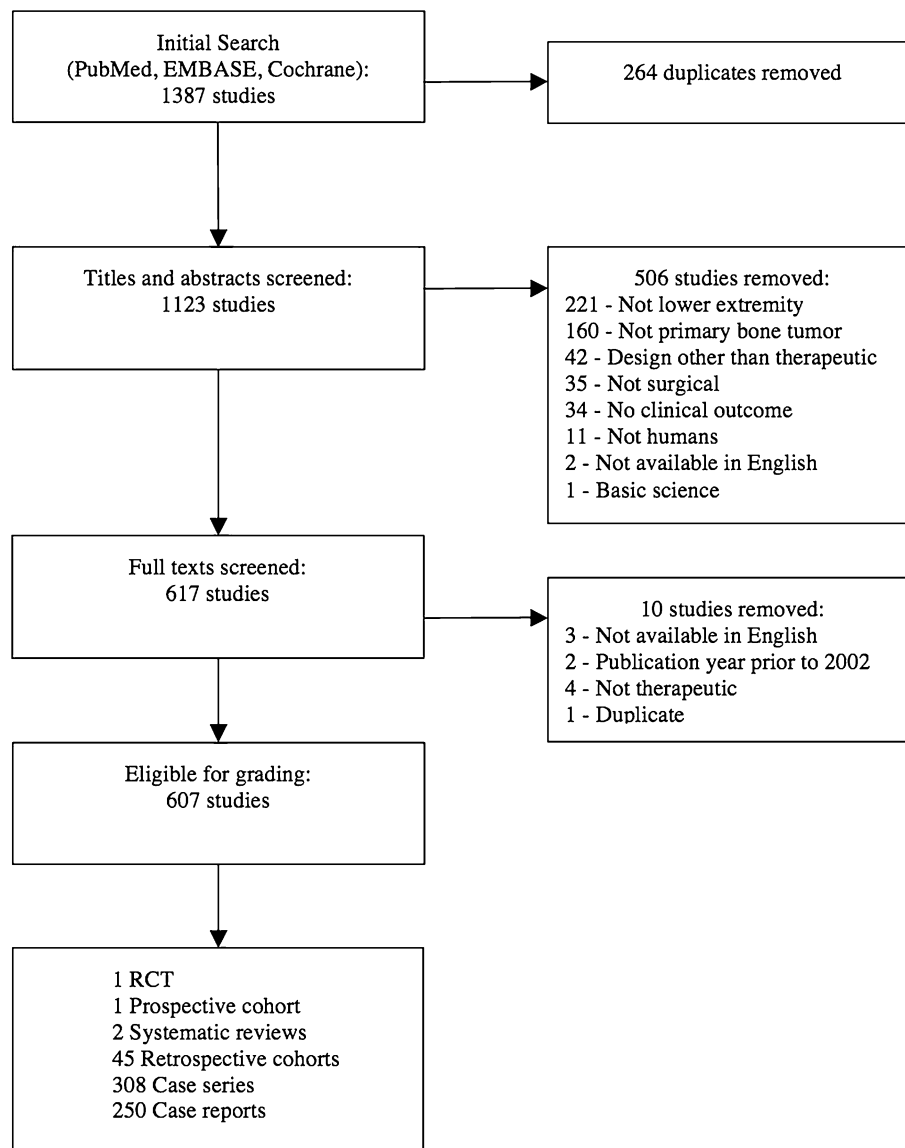


Fig. 1 The search results, screening, and reasons for exclusion of articles at each stage of the study are shown in a flow chart.

and the best reported items having higher percentage scores. Kappa coefficient with the corresponding confidence intervals (CIs) are reported. SPSS (Chicago, IL, USA) and Microsoft Excel (Redmond, WA, USA) were used for data analysis.

Results

Levels of Evidence

Of the 607 included studies, one was an RCT, one was a prospective cohort, two were systematic reviews, 45 were retrospective controlled studies, 308 were retrospective

uncontrolled studies (case series), and 250 were case reports. Both of the systematic reviews were graded by each of the two reviewers as Level III evidence on the basis of the quality of their included studies. The RCT was graded as Level II by both reviewers on the basis of its poor method of randomization, lack of sample size calculation, and lack of blinding. There was one prospective cohort study graded as Level II by both reviewers. The 45 retrospective controlled studies were all graded as Level III and the 308 retrospective uncontrolled studies (case series) were all graded as Level IV (Fig. 2). The 250 case reports were graded as Level V. Agreement between the two reviewers for levels of evidence was satisfactory ($\kappa = 0.87$; 95% CI, 0.80–0.95). Overall, 99% of the studies on this topic are Level III to V evidence.

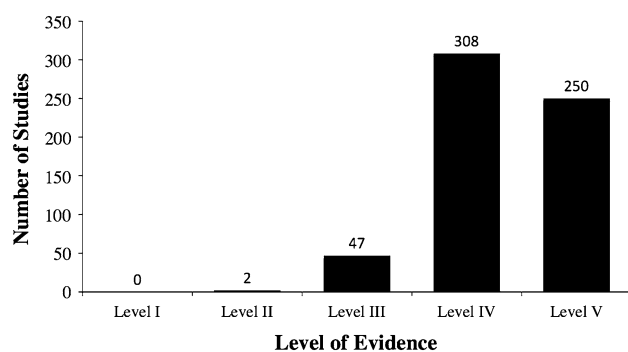


Fig. 2 The distribution of the levels of evidence for therapeutic studies in the surgical management of primary lower extremity bone tumors illustrates the high prevalence of level III–V studies.

Quality of Reporting

Both reviewers scored the single RCT 54% on the Detsky Quality Assessment Scale. The 45 retrospective controlled studies and the one prospective cohort were eligible for assessment using the STROBE checklist. The STROBE points are presented in descending order from best to least reported (Table 1). The mean percentage of STROBE points reported satisfactorily in each article as graded by the two reviewers was 53% (95% CI, 42%–63%). The 10 most consistently reported items were (1) provision of an informative and balanced abstract (99% reported); (2) reporting numbers of outcome events in the results section (99% reported); (3) explaining the scientific background and rationale for the investigation in the introduction (96% reported); (4) key results reviewed in the discussion section (96% reported); (5) giving a cautious overall interpretation of the results in the discussion (88% reported); (6) clearly defining all variables and outcomes in the methods (86% reported); (7) describing all statistical methods (78% reported); (8) reporting numbers of participants included at each stage of the study including those at final followup in the results (76% reported); (9) describing the setting, location, and relevant periods of recruitment, exposure, followup, and data collection in the methods (74% reported); and (10) reporting demographic, clinical, and social characteristics of the study participants, usually in the results (73% reported).

Common Pitfalls

The 10 most consistently underreported items were (1) description of any sensitivity analyses to verify the main results (2.2% reported). Sensitivity analyses are secondary statistical calculations that examine uncertainties in the primary analyses, such as quantifying the impact of losses to follow-up or missing data on the study outcomes [35];

Table 1. STROBE checklist points presented in descending order from best to least reported in the primary management of lower extremity bone tumors

Rank	STROBE item	Proportion of studies reporting each item
1	1b—abstract	98.9%
1	15—outcome data	98.9%
3	2—background	95.7%
3	18—key results	95.7%
5	20—interpretation	88.0%
6	7—variables	85.9%
7	12a—statistical methods	78.3%
8	13a—participants	76.1%
9	5—setting	73.9%
10	14a—characteristics	72.8%
11	4—study design	71.7%
12	3—objectives	69.6%
13	21—generalizability	68.5%
14	13b—non-participants	66.3%
14	14c—followup	66.3%
16	16a—estimates and precision	64.1%
17	17—other analyses	60.9%
18	12b—subgroups	58.7%
19	6b—matching	56.5%
20	6a—participants	55.4%
21	19 limitations	53.3%
22	8—measurement	44.6%
23	16b—boundaries	39.1%
24	16c—relative risk	38.0%
25	11—quantitative variables	29.3%
26	14b—missing data	26.1%
27	22—funding	18.5%
28	12d—losses	14.1%
29	12c—missing data	9.8%
30	9—bias	6.5%
31	13c—flow diagram	4.3%
32	1a—title	3.3%
33	10—study size	2.2%
33	12e—sensitivity analysis	2.2%

STROBE = Strengthening the Reporting of Observational Studies in Epidemiology.

(2) justification of a sample size or a description of how the study size was arrived at in the methods (2.2% reported); (3) indicating the study’s design with a commonly used term in the title or the abstract (3.3% indicated); (4) providing a flow diagram to illustrate numbers of individuals at each stage of the study (4.3% provided); (5) describing any efforts to address potential sources of bias in the methods (6.5% reported); (6) explaining how missing data were addressed (9.8% reported); (7) explaining how losses to followup were addressed (14% reported). This item

required a description of how losses were handled statistically and whether lost subjects may have systematically differed from non-lost subjects, thereby biasing the study results [30]; (8) describing whether there were any sources of funding and the role of any funders (19% reported). Standard disclosures of funding sources were considered adequate even if details about the role of the funders were not provided; (9) indicating the number of participants with missing data for each data point (26% reported); and (10) explaining how quantitative variables were handled in the analyses (29% reported). Simple descriptions of the statistical tests used were considered adequate.

Discussion

Despite their limitations, observational studies dominate the orthopaedic surgery literature and provide an important source of evidence to guide treatment decisions [6, 14, 15]. To our knowledge, a formal assessment of the quality of the evidence that guides the management of primary bone tumors has not previously been published. We systematically reviewed the literature to characterize the overall levels of evidence in studies reporting on the surgical management of lower extremity primary bone tumors and to evaluate the quality of reporting of observational studies in this field. There were no Level I studies and only two Level II studies. The best available evidence in this field otherwise comes from 47 Level III studies; the large majority of what is published in this discipline, however, is either case reports or retrospective uncontrolled studies (case series). We further found that there exist a number of important deficiencies in the quality of scientific reporting on the topic of surgical management of lower extremity primary bone tumors.

This study has several limitations. We did not manually search the bibliographies of included studies and did not solicit further articles from an expert panel; we suspect that we would have further included only a relatively small number of articles that would not have significantly changed our key results. We did not calculate kappa values or intraclass correlation coefficients between reviewers for total STROBE scores because the STROBE checklist has not been validated as a scoring tool. Our system of ranking each STROBE point by percentage scores of reporting could have overestimated quality through an assessment bias, which effectively only strengthens our conclusions. We did not examine trends in levels of evidence or reporting over time, but this report illustrates the state of the current literature over the most recent 12-year period. By nature of our systematic search, our analysis reflects the literature in this field as a whole. We did not assess whether the reporting trends of one or a few very productive

research groups biased our findings. Similarly, we did not compare reporting between journals that have formally adopted the STROBE checklist and those that have not. We considered all items as potentially eligible, although some were clearly less applicable to every article than others (such as descriptions of matching in nonmatched studies or the inclusion of sensitivity analyses). We scored items as “not reported” when it was absent, regardless of its applicability. This procedure may have introduced some imprecision to our analysis and likely depressed the percentage of articles reporting each item, but it is unlikely to have affected our overall rankings of best and least reported items. A process to modify the STROBE checklist based on applicability has not been established in the literature. Our search strategy did not specify free-text supplements for every possible type of specific bone tumor, but instead used broad MeSH and Emtree headings. These headings are broad and inclusive tags that include all subcategories of related items such as various types bone tumors by their design. It is possible that additional free-text supplements would have yielded further studies. We did not manually search for articles that are not electronically indexed.”

The most commonly used study design in this field was the retrospective uncontrolled study (case series) and the vast majority of studies were Level IV or Level V evidence. Our findings are not unique in the orthopaedic literature. In their review of the *Foot and Ankle* literature from 2000 to 2010, Zaidi et al. [38] showed a similar preponderance of observational studies with Level III to V data comprising over 89% of the published literature in that field. Samuelsson et al. [28] identified in their review of ACL reconstruction studies that Level III to V studies made up 81% of all reports between 1995 and 2011. Obremeskey et al. [21] examined the levels of evidence across nine orthopaedic journals from January to June 2003 and found that, after excluding case reports and expert opinions, the most common level of evidence was Level IV (58%). In their review of publications in a single orthopaedic journal over the last 30 years, Hanzlik et al. [13] showed a steady increase in the number of published Level I and II studies, but Level III and IV reports remained most common (64% combined). Proposed explanations for the relative paucity of RCTs in orthopaedic surgery include ethical barriers limiting the enrollment of patients and the use of sham procedures, difficulty applying RCT methodology to varied surgical problems, and the significant expense and time associated with performing large trials [10, 31, 32]. With rigorous design, however, many surgical questions can be addressed in expertise-based or sham-controlled studies with satisfactory blinding of participants and outcome assessors [11, 19, 29, 33]. Adequately powered, large RCTs require leadership and collaboration in

addition to infrastructure and funding [25, 31]. Investigative networks in which centers each recruit relatively fewer patients can increase overall recruitment and bring feasibility to the study of rarer diseases [37], and the recently launched Prophylactic Antibiotics in Tumor Surgery (PARITY) trial marks the first such undertaking for malignant primary bone tumors [12].

The mean percentage of STROBE points reported satisfactorily in each article as graded by the two reviewers was 53% (95% CI, 42%–63%). Parsons et al. [23] reported an overall STROBE compliance rate of 58% (95% CI, 56%–60%) in a random sample of 100 articles from selected general orthopaedic surgery journals, which is quite similar to the result we found for oncology. Others have investigated the quality of reporting of RCTs and meta-analyses in orthopaedic surgery; Audige et al. [1] demonstrated that only 20% of systematic reviews included a quality assessment for included observational studies [2–4, 7, 24]. Nonetheless, many top orthopaedic journals have adopted STROBE in their instructions to authors and preference for its use is evident in editorials, methodological reviews, and instructional articles [5, 14, 16, 17, 20, 30]. To our knowledge, this is the first report to systematically review the quality of reporting for studies in orthopaedic oncology. We have identified the need for improved reporting field in this field, which could potentially improve the quality of future studies.

Using the STROBE checklist, we identified several recurrent strengths and deficiencies in the quality of reporting for cohort studies. There are several STROBE items whose improved reporting could significantly improve the transparency of future observational studies: identification of the chosen study design in either the title or abstract; justified study sample size; accounting of missing data and losses to followup; and discussion of potential sources of bias. Some items such as sensitivity analyses are not always possible or indicated in small observational surgical studies. Parsons et al. also found deficiencies in study design identification and sample size justification, but did not report on the other items for which we found deficiencies [23].

We found that retrospective studies are the dominant form of evidence for the surgical management of primary lower extremity bone tumors. Given the rarity of these tumors and the challenges inherent to completing large prospective interventional studies, observational studies will likely continue to prevail as the best available evidence in this field, and treating surgeons should be aware of the quality of literature supporting his or her treatment decisions when deciding on an operation and discussing the options with patients. Authors may use the results of this study to minimize the ambiguity of selection bias, transfer bias, and assessor bias by emphasizing clear descriptions of participant selection, losses to followup, and whether there

were differences in followup between treatment arms, the amount and effect of any missing data, and by clarifying who assessed the primary outcomes and which instruments were used.

Acknowledgments We thank librarian Michael Fraumeni for his assistance in the conduct of our electronic literature search.

References

- Audige L, Bhandari M, Griffin D, Middleton P, Reeves BC. Systematic reviews of nonrandomized clinical studies in the orthopaedic literature. *Clin Orthop Relat Res*. 2004;427:249–257.
- Bhandari M, Guyatt GH, Lochner H, Sprague S, Tornetta P 3rd. Application of the Consolidated Standards of Reporting Trials (CONSORT) in the fracture care literature. *J Bone Joint Surg Am*. 2002;84:485–489.
- Bhandari M, Morrow F, Kulkarni AV, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am*. 2001;83:15–24.
- Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the journal of bone and joint surgery from 1988 through 2000. *J Bone Joint Surg Am*. 2002;84:388–396.
- Brand RA. Standards of reporting: the CONSORT, QUORUM, and STROBE guidelines. *Clin Orthop Relat Res*. 2009;467:1393–1394.
- Carr AJ. Evidence-based orthopaedic surgery: what type of research will best improve clinical practice? *J Bone Joint Surg Br*. 2005;87:1593–1594.
- Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. *J Bone Joint Surg Am*. 2007;89:1970–1978.
- Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol*. 1992;45:255–256.
- Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, Sackett DL, Cinà CS, Walter SD, Haynes B, Schünemann HJ, Norman GR, Guyatt GH. Need for expertise based randomised controlled trials. *BMJ*. 2005;330:88.
- Dowrick AS, Bhandari M. Ethical issues in the design of randomized trials: to sham or not to sham. *J Bone Joint Surg Am*. 2012;94(Suppl 1):7–10.
- Farrokhhyar 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.
- Ghert M, Deheshi B, Holt G, Randall RL, Ferguson P, Wunder J, Turcotte R, Werier J, Clarkson P, Damron T, Benevenia J, Anderson M, Gebhardt M, Isler M, Mottard S, Healey J, Evaniew N, Racano A, Sprague S, Swinton M, Bryant D, Thabane L, Guyatt G, Bhandari M; PARITY Investigators. Prophylactic Antibiotic Regimens In Tumour Surgery (PARITY): protocol for a multicentre randomised controlled study. *BMJ Open*. 2012;2:10.1136/bmjopen.2012-002197. Print 2012.
- 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.
- Hoppe DJ, Schemitsch EH, Morshed S, Tornetta P 3rd, Bhandari M. Hierarchy of evidence: where observational studies fit in and why we need them. *J Bone Joint Surg Am*. 2009;91(Suppl 3):2–9.

15. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996;347:984–985.
16. Konstantakos EK, Laughlin RT, Markert RJ, Crosby LA. Assuring the research competence of orthopedic graduates. *J Surg Educ*. 2010;67:129–134.
17. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. *J Bone Joint Surg Am*. 2009;91(Suppl 3):21–26.
18. 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. QUOROM group. *Br J Surg*. 2000;87:1448–1454.
19. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002;347:81–88.
20. Mundi R, Chaudhry H, Singh I, Bhandari M. Checklists to improve the quality of the orthopaedic literature. *Indian J Orthop*. 2008;42:150–164.
21. 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.
22. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. *Cancer Treat Res*. 2009;152:3–13.
23. Parsons NR, Price CL, Hiskens R, Achten J, Costa ML. An evaluation of the quality of statistical design and analysis of published medical research: results from a systematic survey of general orthopaedic journals. *BMC Med Res Methodol*. 2012;12:60.
24. Poolman RW, Struijs PA, Krips R, Sierevelt IN, Lutz KH, Bhandari M. Does a 'level I evidence' rating imply high quality of reporting in orthopaedic randomised controlled trials? *BMC Med Res Methodol*. 2006;6:44.
25. Rangan A, Brealey S, Carr A. Orthopaedic trial networks. *J Bone Joint Surg Am*. 2012;94(Suppl 1):97–100.
26. Sackett D, Haynes R, Guyatt G, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinicians*. 2nd ed. Boston, MA, USA: Little Brown; 1991.
27. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. 1996. *Clin Orthop Relat Res*. 2007;455:3–5.
28. Samuelsson K, Desai N, McNair E, van Eck CF, Petzold M, Fu FH, Bhandari M, Karlsson J. Level of evidence in anterior cruciate ligament reconstruction research: a systematic review. *Am J Sports Med*. 2013;41:924–934.
29. Scholtes VA, Nijman TH, van Beers L, Devereaux PJ, Poolman RW. Emerging designs in orthopaedics: expertise-based randomized controlled trials. *J Bone Joint Surg Am*. 2012;94(Suppl 1):24–28.
30. Sheffler LC, Yoo B, Bhandari M, Ferguson T. Observational studies in orthopaedic surgery: the STROBE statement as a tool for transparent reporting. *J Bone Joint Surg Am*. 2013;95:e14(1–12).
31. Shore BJ, Nasreddine AY, Kocher MS. Overcoming the funding challenge: the cost of randomized controlled trials in the next decade. *J Bone Joint Surg Am*. 2012;94(Suppl 1):101–106.
32. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery*. 1995;118:459–467.
33. Study to Prospectively Evaluate Reamed Intramedullary Nails in Patients with Tibial Fractures Investigators, Bhandari M, Guyatt G, Tornetta P 3rd, Schemitsch EH, Swiontkowski M, Sanders D, Walter SD. Randomized trial of reamed and unreamed intramedullary nailing of tibial shaft fractures. *J Bone Joint Surg Am*. 2008;90:2567–2578.
34. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. 2005;161:280–288.
35. Vandembroucke JP, von Elm E, Altman DG, Gotzsche 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. *Epidemiology*. 2007;18:805–835.
36. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85:1–3.
37. Wright JG, Yandow S, Donaldson S, Marley L, Simple Bone Cyst Trial Group. A randomized clinical trial comparing intralesional bone marrow and steroid injections for simple bone cysts. *J Bone Joint Surg Am*. 2008;90:722–730.
38. Zaidi R, Abbassian A, Cro S, Guha A, Cullen N, Singh D, Goldberg A. Levels of evidence in foot and ankle surgery literature: progress from 2000 to 2010? *J Bone Joint Surg Am*. 2012;94:e1121–1110.