Full Length Article

JHS(E)

Journal of Hand Surgery
(European Volume)
2018, Vol. 43(8) 801–807

DOI: 10.1177/1753193418780184
journals.sagepub.com/home/jhs
SAGE

Quality of surgical randomized controlled trials in hand surgery: a systematic review

Chao Long[1], Heather E. desJardins-Park[1], Rita Popat[1] and Paige M. Fox[2,3]

Abstract

We assessed the quantity, quality and trends of randomized controlled trials comparing hand surgical interventions. Study characteristics were collected for 125 randomized controlled trials comparing hand surgical interventions. The Jadad scale (0–5), which assesses methodological quality of trials, was calculated. Logistic regressions were conducted to determine associations with the Jadad score. The studies were published between 1981 and 2015, with an increase over time, most often in *Journal of Hand Surgery (European)*. Mean study size was 68 patients. Mean Jadad score was 2.1, without improvement over time. Thirty percent conducted a power analysis and 8% an intention-to-treat analysis. Studies conducted in the United Kingdom and with smaller sample sizes, power analysis and intention-to-treat analysis were associated with a higher Jadad score. The quantity of trials has increased over time while methodological quality has remained low, indicating a need to improve quality of trials in hand surgery literature.

Keywords

Clinical research, hand surgery, randomized controlled trials, systematic review

Date received: 8th January 2018; accepted: 10th May 2018

Introduction

As the now well-established paradigm of evidence-based medicine (EBM) took hold in recent decades (Evidence-Based Medicine Working Group, 1992), randomized controlled trials (RCTs) rose to become the gold standard of evidence in clinical research and medical knowledge (Bothwell et al., 2016). Although many specialties have adopted the principles of EBM, its application in surgery, or evidence-based surgery (EBS), has been limited by the relative paucity of RCTs examining surgical, as compared to medical, interventions (McCulloch et al., 2002; Slim, 2005). This, in turn, has been attributed to the unique challenges of conducting well-designed surgical RCTs, such as the inevitable variability in surgeons' techniques, the surgical learning curve, impracticalities associated with blinding and difficulties in recruiting patient cohorts large enough to sufficiently power the studies (Cook, 2009; Farrokhyar et al., 2010; McCulloch et al., 2002). While these challenges contribute to the low quantity of surgical RCTs, they may also compromise their quality. Since poor study design and execution undermine the superiority of evidence that RCTs promise (Farrokhyar et al., 2010), both the level of evidence of a study's research design and its methodological rigour must be taken into account in making an assessment of the validity of the conclusions drawn.

Although recent studies have looked at the quantity and quality of surgical RCTs in all surgery (Agha et al., 2007; Jacquier et al., 2006; Yu et al., 2017), general surgery (Ahmed Ali et al., 2013; Balasubramanian et al., 2006), plastic surgery (Veiga Filho et al., 2005; Voineskos et al., 2016; Voineskos et al., 2016), orthopaedic surgery (Chess and Gagnier, 2013; Karanicolas et al., 2008) and elbow surgery (Somford et al., 2015), there are limited data in the specialized field of hand

[1]Stanford University School of Medicine, Stanford, CA, USA
[2]Division of Plastic & Reconstructive Surgery, Stanford Health Care, Palo Alto, CA, USA
[3]Division of Plastic & Reconstructive Surgery, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA

Corresponding Author:
Paige M. Fox, Stanford University Medical Center, Department of Surgery, Division of Plastic & Reconstructive Surgery, 770 Welch Road, Suite 400, Palo Alto, CA 94304, USA.
Email: pfox@stanford.edu

surgery. While several studies have examined RCTs in hand surgery, they either included RCTs comparing elbow and/or shoulder interventions, or examined only RCTs published in a predetermined set of journals (Gummesson et al., 2004; Sugrue et al., 2016). For example, Kim et al. (2017) queried six journals, while Post et al. (2014) searched only two journals for RCTs for inclusion.

We conducted a systematic review of surgical RCTs published about hand surgery. Our primary goal was to assess the quantity and methodological quality of hand-surgery-related RCTs over time. Specifically, we aimed to ascertain the risk of bias, as measured by the appropriateness of randomization methods, whether the patient and/or study personnel were blinded, and adequate accounting of all recruited patients. Our secondary goal was to characterize the RCTs published in the hand surgery literature.

Methods

Search strategy and study selection

We conducted a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). In collaboration with a medical librarian, we constructed electronic search strategies to identify RCTs in PubMed, Cochrane, Scopus, Google Scholar and Clinicaltrials.gov comparing two or more hand surgical interventions. Our search included variations of the following terms: hand; surgery; and randomized controlled trial. The search strategy used for PubMed was as follows:

> (''hand/surgery'' [mesh] OR ''hand transplantation'' [mesh] OR ((surger* [ti] OR surgic* [ti]) AND (hand [ti] OR wrist* [ti] OR finger* [ti] OR hands [ti] OR thumb* [ti] OR metacarpus [ti])) OR ''hand injuries/surgery'' [mesh] OR ''hand joints/surgery'' [mesh] OR (((''Median Neuropathy'' [mesh] OR ''median nerve'' [ti] OR ''brachial plexus'' [mesh]) AND (''surgery'' [sh] OR surger* [ti] OR surgical* [ti])) OR ''hand deformities/surgery'' [mesh] OR (hand [mesh] AND ''surgery'' [sh])) AND (random*[tw] OR rcts [tw] OR rct [tw] OR placebo*[tw] OR control* [ti] OR ''double blind''[tw] OR blinded[tw] OR ''single blind''[tw] OR ((singl*[tw] OR doubl*[tw] OR trebl*[tw] OR tripl*[tw]) AND (mask*[tw] OR blind*[tw])) OR ''latin square''[tw] OR ''randomized controlled trial''[pt] OR ''controlled clinical trial''[pt] OR ''Random Allocation'' [mesh] OR ''Randomized Controlled Trials as Topic'' [mesh]) NOT (animals[mh] NOT humans[mh]) AND English [lang]
>
> The search strategies for the remaining databases can be provided upon request. The last search was performed on 19 May 2016.

Inclusion criteria included surgical RCTs comparing two or more surgical interventions to address the same hand condition, where surgical was defined as 'a procedure involving cutting, suturing, or physically changing body tissues' (Voineskos et al., 2016), RCT was defined per the Cochrane handbook (Higgins and Green, 2011) and hand was defined as any anatomical structure including or distal to the carpal bones. RCTs comparing a surgical intervention to no surgery were not included. Exclusion criteria were as follows: RCTs studying: (1) injections (e.g. steroid injections, botulinum toxin injections); (2) postoperative mobilization plans; (3) anaesthesia/pain management (e.g. nerve blocks); (4) laser therapy; and (5) pharmaceuticals, even if used in surgery. Animal studies and studies not written in English were also excluded. To capture as many hand surgery RCTs as possible, we had no exclusion criteria regarding date of publication.

Two independent reviewers (CL and HED) assessed manuscripts for inclusion. All studies first underwent title and abstract screening, and those that were included then underwent full text screening. Conflicts between the two reviewers were discussed until a consensus decision was reached. Covidence software (Melbourne, VIC, Australia), the standard production platform for Cochrane Reviews, was used to conduct this screening.

Data collection and data items

We collected study characteristics including year published, publishing journal, first author's country affiliation, funding status, hand condition studied, single vs. multi-centre collaboration, study size, randomization technique, patient and researcher blinding, appropriate accounting of recruited patients, primary outcome measure(s), power analysis, mention of adverse events, follow-up and intention-to-treat analysis. For RCTs that resulted in more than one publication, all publications were used for data extraction. When a given variable was reported in more than one manuscript (e.g. year of publication), data from the earliest manuscript were collected.

Risk of bias assessment

We calculated the Jadad scale, a 0–5 scale with a higher score representing higher quality. The Jadad

scale is a validated instrument for assessing RCTs' reported methodological quality that incorporates: (1) appropriateness of randomization methods; (2) double blinding; and (3) an adequate account of all patients (Jadad et al., 1996). A point is given if the study reports employing each of the above three methodological safeguards against bias. For randomization and double blinding, an additional point is given if reported methods are appropriate, and a point is deducted for inappropriate methods. The Cochrane Handbook was used as the guideline for determining appropriateness (Higgins and Green, 2011).

Statistical analysis

Descriptive statistics were calculated for all collected data. Summary statistics are expressed as mean (standard deviation [SD]). The Kendall rank correlation was used to assess trends of RCTs over time. Because the Jadad score was normally distributed, as assessed by a box and whisker plot, a stem and leaf plot and a Q-Q plot, it was treated as a continuous variable and linear regression models were used to determine which study characteristics were associated with Jadad score. The principal summary measure that quantifies the association was slope (regression coefficient), which represents the mean change in Jadad score for a unit change in the study characteristic variable. Significance was set at $p < 0.05$.

Results

Study selection

Of 2253 identified studies, 135 manuscripts met full inclusion criteria (Figure 1, Supplementary Appendix). These 135 manuscripts represented 125 unique RCTs eligible for data extraction.

Study characteristics

The 125 RCTs were published between 1981 and 2015, with a significant increase over time ($p < 0.0001$) (Figure 2). They were published in 41 different journals, most often in *Journal of Hand Surgery (European)* (n = 34, 27%), *Journal of Hand Surgery (American)* (n = 21, 17%) or *Journal of Bone and Joint Surgery* (n = 14, 11%) (Table S1). The RCTs were affiliated with 29 countries, most often with the United Kingdom (n = 30, 24%), the United States (n = 20, 16%) or Sweden (n = 8, 6%) (Table S2). They were most often single-centre studies (n = 112, 90%). Of 13 studies that were multi-centre, two recruited patients internationally, with the remaining recruiting from only one country. The majority of manuscripts (n = 79, 63%) did not mention funding source, 16% (n = 20) were funded and 21% (n = 26) were unfunded. Of those that were funded, 13 received not-for-profit or academic funding, while seven received industry funding. Almost half of all RCTs were studying interventions to treat carpal tunnel syndrome (n = 58, 46%); the next most commonly studied hand conditions were osteoarthritis (n = 20, 16%) and rheumatoid arthritis (n = 10, 8%) (Table S3). Mean study size was 68 (SD 50.6) patients.

Reported methodology

Randomization methods were not specified in approximately one-quarter of studies (n = 34, 27%). When specified, the two most commonly used appropriate methods were computer random number generator (n = 24, 26%) and sealed envelopes (n = 23, 25%), while those of inappropriate methods were sequence generated by medical record identifier (n = 6, 6.6%) or by day of admission (n = 3, 3.3%) (Supplementary Fig S1).

Less than one-tenth (n = 12, 9.7%) of studies reported double blinding of both patients and assessors. In the vast majority of studies (n = 103, 82%), it was unknown whether assessors and/or patients were blinded. The remaining nine studies reported blinding, but methodology was insufficient: six studies did not blind patients; one study did not blind the outcome assessor; and two studies blinded neither.

In the majority of studies, the fate of some patients in the trial was unknown (n = 76, 61%). In the remaining studies, the fate of all patients in the trial was known and reasons given if data were missing (n = 49, 39%).

One-fifth of studies defined primary outcome measures (n = 25, 20%) and 16.8% (n = 21) defined secondary outcome measures. Of those with outcome measures, 15 studies defined these before initiation of the trial and the remaining (n = 10) did not specify whether outcome measures were defined pre-RCT or post-hoc. Less than one-third (n = 38, 30%) conducted a power analysis, and of those that did, 26 conducted *a priori* analyses, six conducted post-hoc analyses and six did not specify the type. The large majority did not use intention-to-treat analysis (n = 115, 92%).

Follow-up

Follow-up was as follows: 27% (n = 34) had follow-up on all patients, 42% (n = 53) had patients lost to follow-up, 28% (n = 35) had incomplete reporting so that follow-up status could not be determined and 2.4% (n = 3) had a study design without or not requiring follow-up. Of the 53 (42%) studies with loss to

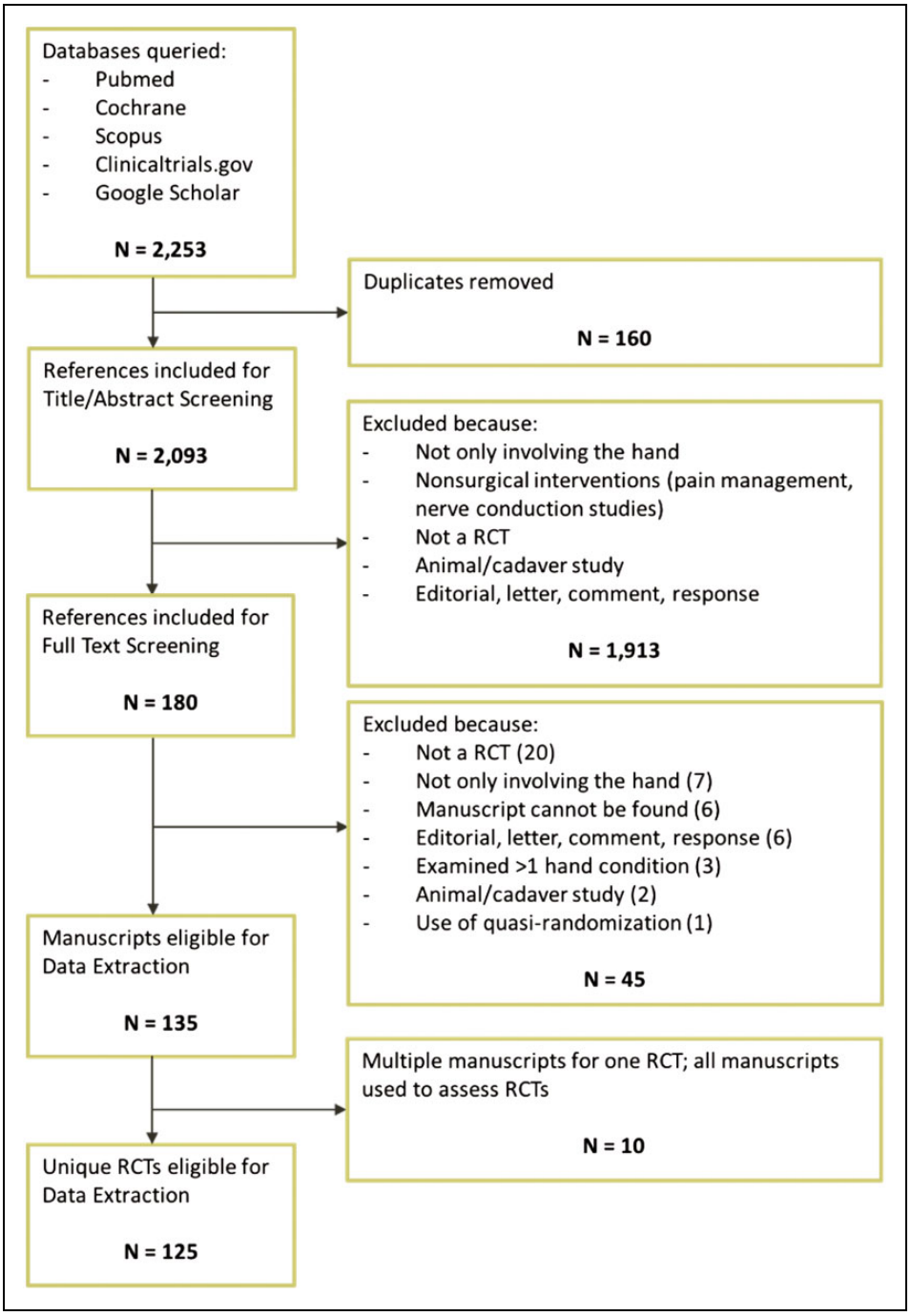

Figure 1. PRISMA flow chart of selection of randomized controlled trials in the hand surgery literature.

follow-up, almost two-thirds (n = 33, 62%) did not provide an explanation for each patient.

Jadad score

Mean Jadad score was low at 2.1 (SD 1.2), with the distribution as seen in Figure 3. There was no significant increase in Jadad score over time ($p > 0.05$) (Figure 4).

Study characteristics associated with Jadad score

Linear regression results suggested that studies with power analysis (slope = 0.85, $p = 0.0002$), with a smaller sample size (slope = –0.005, $p = 0.02$), conducted in the UK (slope = 0.95, $p = 0.005$) and that used intention-to-treat analysis (slope = 1.07, $p = 0.006$) were more likely to have a higher Jadad score (Table S4). Year

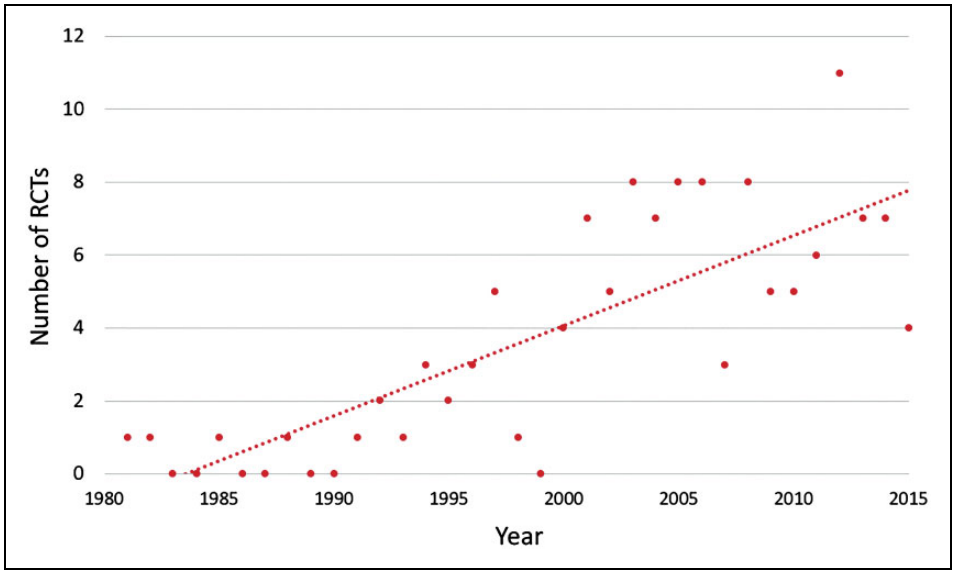

Figure 2. Trend over time for numbers of randomized controlled trials in hand surgery.

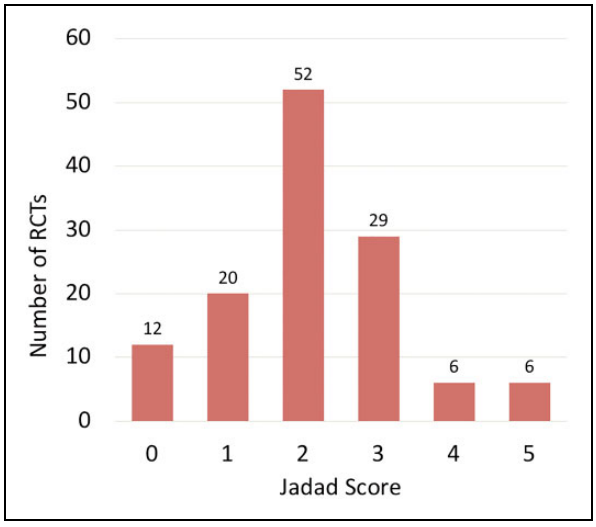

Figure 3. Histogram of Jadad Scores for randomized controlled trials in hand surgery.

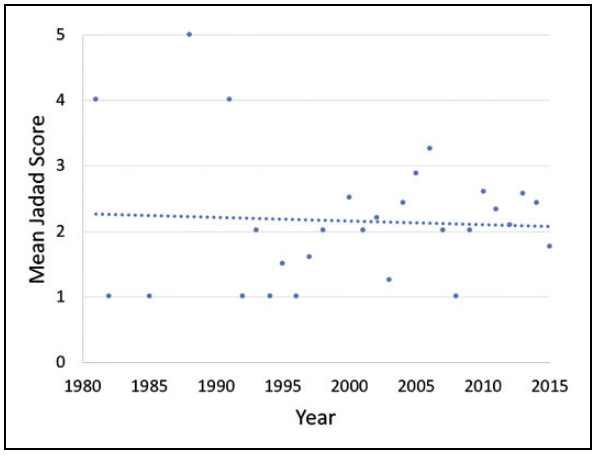

Figure 4. Trend over time of mean Jadad Score for randomized controlled trials in hand surgery.

published, publishing journal, single- vs. multi-centre study design, funding status and type of funding (industry vs. not-for-profit) were not associated with Jadad score ($p > 0.05$), and whether a primary outcome was defined trended towards significance ($p = 0.06$).

Discussion

While RCTs are considered level 1b evidence second only to systematic reviews of RCTs (level 1a evidence) (Oxford Centre for Evidence-based Medicine, 2009), the superiority of the evidence cannot be assumed within surgery (McCulloch et al., 2002). Instead, as with all research findings, the weight and grade of recommendation of the generated evidence must depend on not only research design, but also on the appropriateness and rigour of the study's methodology and the patient cohort recruited. Despite the challenges associated with surgical RCTs, creative solutions can be devised to overcome them. For example, with regard to blinding, although it is impractical to blind the surgeon, using the same dressing that covers the hand and/or incision site for all arms of the study is one potential way to blind the patient and outcome assessor.

Several studies have reported on the level of evidence within hand surgery. They demonstrated that most available studies are case series, or level 4

evidence, with only 11% comprised of high-quality evidence (Rosales et al., 2012; Sugrue et al., 2016). They also demonstrated that there has been a significant increase in the average level of evidence of articles (Sugrue et al., 2016), which is consistent with our findings of increased numbers of RCTs over time. Although the level of evidence in hand surgery was shown to be correlated with journal (Rosales et al., 2012), we found that journal was not predictive of Jadad score. Together, these findings suggest that although journals are publishing studies with a higher level of evidence, the methodological quality of these studies has remained poor over time.

The hand surgery community has already made strides towards addressing this issue. For example, the CONSORT (Consolidated Standards of Reporting Trials) Statement, an evidence-based set of recommendations regarding best practices of reporting RCTs, was first issued in 1996 (Begg et al., 1996). By 2004, *Journal of Hand Surgery (European)* formally endorsed CONSORT, recommending authors to ensure that manuscripts fulfil the CONSORT criteria before submission (Sauerland and Davis, 2004). More broadly speaking, leaders in the field have stressed the importance of ensuring the rigour of clinical research within hand surgery, and professional societies such as the American Society for Surgery of the Hand (ASSH) have created task forces to improve clinical research in hand surgery (Chung, 2010; Chung et al., 2006). Despite these critical efforts, our findings suggest that there is much work that remains to be done.

There are several limitations to our study. First, because only RCTs reported in English were included, language bias may have been introduced into our review. Second, the Jadad score is a limited metric for quality and risk of bias in RCTs, as it incorporates only three aspects of a good RCT. Although using appropriate randomization methods, double blinding and an account of all patients are three essential components of a well-designed RCT, there are admittedly others. For example, failure to apply intention-to-treat analysis, which compares patients in the groups to which they were originally randomly assigned, may overestimate clinical effectiveness (Hollis and Campbell, 1999; Ruiz-Canela et al., 2000). Failure to conduct a power analysis leaves insufficient power undetected, which increases the likelihood of making a type II error (Chung et al., 1998). However, our data showed that these two study design components were associated with higher Jadad score, hence validating the Jadad score as an appropriate proxy for these other components.

To the best of our knowledge, there is not yet a better metric for assessing the quality of surgical RCTs. In a review of 21 scales for assessing quality of RCTs, Olivo et al. (2008) concluded that the Jadad scale 'presented the best validity and reliability evidence'. The Jadad scale was, as of 2008, also the most frequently cited and used scale in medical literature, which allows for easy and direct comparison with other similar studies (Olivo et al., 2008). There may, however, be a need for a scale designed specifically for measuring quality of surgical RCTs, with consideration of challenges unique to surgical RCTs (e.g. blinding). Although the CONSORT statement has also been used, it is designed to be a checklist of best practices for reporting for investigators and 'is not intended as an instrument to evaluate the quality of a trial. Nor is it appropriate to use the checklist to construct a ''quality score''' (Schulz et al., 2010).

In conclusion, we found that there has been a significant increase in the number of RCTs over time. However, methodological quality of RCTs has remained low. These findings indicate a need to improve the quality of RCTs in hand surgery literature, which can be accomplished by employing appropriate randomization, blinding, accounting for all patients, power analysis and intention-to-treat analysis. These components of a well-designed study are essential to avoiding bias, ensuring the validity of conclusions drawn and ultimately bolstering evidence-based hand surgery to advance patient care.

Acknowledgements The authors thank Christopher Stave for his invaluable methodological guidance, specifically in regard to constructing our electronic search strategies.

Declaration of conflicting interests The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

References

Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: A systematic review. Int J Surg. 2007, 5: 413–22.

Ahmed Ali U, van der Sluis PC, Issa Y et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg. 2013, 258: 199–207.

Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: Can we do better? Ann Surg. 2006, 244: 663–7.

Begg C, Cho M, Eastwood S et al. Improving the quality of reporting of randomized controlled trials. The consort statement. JAMA. 1996, 276: 637–9.

Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard–lessons from the history of RCTs. N Engl J Med. 2016, 374: 2175–81.

Chess LE, Gagnier J. Risk of bias of randomized controlled trials published in orthopaedic journals. BMC Med Res Methodol. 2013, 13: 76.

Chung KC. Clinical research in hand surgery. J Hand Surg Am. 2010, 35: 109–20.

Chung KC, Burns PB, Sears ED. Outcomes research in hand surgery: Where have we been and where should we go? J Hand Surg Am. 2006, 31: 1373–9.

Chung KC, Kalliainen LK, Hayward RA. Type II (beta) errors in the hand literature: The importance of power. J Hand Surg Am. 1998, 23: 20–5.

Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials. 2009, 10: 9.

Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992, 268: 2420–5.

Farrokhyar F, Karanicolas PJ, Thoma A et al. Randomized controlled trials of surgical interventions. Ann Surg. 2010, 251: 409–16.

Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg Am. 2004, 29: 727–34; discussion 35-7.

Higgins JPT, Green S (Eds.) *Cochrane handbook for systematic reviews of interventions*, Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: http://handbook.cochrane.org.

Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ. 1999, 319: 670–4.

Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: A systematic review. Ann Surg. 2006, 244: 677–83.

Jadad AR, Moore RA, Carroll D et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials. 1996, 17: 1–12.

Karanicolas PJ, Bhandari M, Taromi B et al. Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. J Bone Joint Surg Am. 2008, 90: 1026–33.

Kim JM, Zimmerman RM, Jones CM, Muhit AA, Higgins JP, Means KR Jr. The quality of randomised controlled trials involving surgery from the hand to the elbow: A critical analysis of the literature. Bone Joint J. 2017, 99-B: 94–9.

McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: Problems and possible solutions. BMJ. 2002, 324: 1448–51.

Oxford Centre for Evidence-based Medicine. Levels of evidence. March 2009. http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. J Clin Epidemiol. 2009, 62: 1006–12.

Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther. 2008, 88: 156–75.

Post SF, Selles RW, McGrouther DA et al. Levels of evidence and quality of randomized controlled trials in hand and wrist surgery: An analysis of two major hand surgery journals. J Hand Surg Eur. 2014, 39: 900–2.

Rosales RS, Reboso-Morales L, Martin-Hidalgo Y, Diez de la Lastra-Bosch I. Level of evidence in hand surgery. BMC Res Notes. 2012, 5: 665.

Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Estevez J. Intention to treat analysis is related to methodological quality. BMJ. 2000, 320: 1007–8.

Sauerland S, Davis TR. The consolidated standards of reporting trials (CONSORT): Better presentation of surgical trials in the Journal of Hand Surgery. J Hand Surg Br. 2004, 29: 621–4.

Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Open Med. 2010, 4: e60–8.

Slim K. Limits of evidence-based surgery. World J Surg. 2005, 29: 606–9.

Somford MP, van Deurzen DF, Ostendorf M, Eygendaal D, van den Bekerom MP. Quality of research and quality of reporting in elbow surgery trials. J Shoulder Elbow Surg. 2015, 24: 1619–26.

Sugrue CM, Joyce CW, Sugrue RM, Carroll SM. Trends in the level of evidence in clinical hand surgery research. Hand (N Y). 2016, 11: 211–5.

Veiga Filho J, Castro AA, Veiga DF et al. Quality of reports of randomized clinical trials in plastic surgery. Plast Reconstr Surg. 2005, 115: 320–3.

Voineskos SH, Coroneos CJ, Ziolkowski NI et al. A systematic review of surgical randomized controlled trials: Part 2. Funding source, conflict of interest, and sample size in plastic surgery. Plast Reconstr Surg. 2016, 137: 453e–61e.

Voineskos SH, Coroneos CJ, Ziolkowski NI et al. A systematic review of surgical randomized controlled trials: Part i. Risk of bias and outcomes: Common pitfalls plastic surgeons can overcome. Plast Reconstr Surg. 2016, 137: 696–706.

Yu J, Li X, Li Y, Sun X. Quality of reporting in surgical randomized clinical trials. Br J Surg. 2017, 104: 296–303.