GENERAL REVIEW

A review of trauma and orthopaedic randomised clinical trials published in high-impact general medical journals

Luke Farrow[1,2] · William T. Gardner[2] · Andrew D. Ablett[3] · Vladislav Kutuzov[4] · Alan Johnstone[2]

Abstract

Introduction The recent past has seen a significant increase in the number of trauma and orthopaedic randomised clinical trials published in "the big five" general medical journals. The quality of this research has, however, not yet been established.
Methods We therefore set out to critically appraise the quality of available literature over a 10-year period (April 2010–April 2020) through a systematic search of these 5 high-impact general medical journals (JAMA, NEJM, BMJ, Lancet and Annals). A standardised data extraction proforma was utilised to gather information regarding: trial design, sample size calculation, results, study quality and pragmatism. Quality assessment was performed using the Cochrane Risk of Bias 2 tool and the modified Delphi list. Study pragmatism was assessed using the PRECIS-2 tool.
Results A total of 25 studies were eligible for inclusion. Over half of the included trials did not meet their sample size calculation for the primary outcome, with a similar proportion of these studies at risk of type II error for their non-significant results. There was a high degree of pragmatism according to PRECIS-2. Non-significant studies had greater pragmatism that those with statistically significant results ($p < 0.001$). Only 56% studies provided adequate justification for the minimum clinically important difference (MCID) in the population assessed. Overall, very few studies were deemed high quality/low risk of bias.
Conclusions These findings highlight that there are some important methodological concerns present within the current evidence base of RCTs published in high-impact medical journals. Potential strategies that may improve future trial design are highlighted.
Level of evidence Level 1.

Introduction

Evidence-based medicine is the established bedrock of good clinical care. Whilst historically there have been concerns over the strength of evidence base behind orthopaedic interventions [1], the recent past has seen an increase in the number of randomised clinical trials (RCT's) published in major medical journals, particularly the so-called big five

[2]. Understanding the strengths and limitations of these trials is vital to understanding their clinical applicability, as well as providing a key learning opportunity for future trial design and development.

Growth of the orthopaedic trial community has led to increasing interest in the concept of pragmatic trials, where the focus is to reflect real-world applicability of an intervention rather than providing causative explanations for trial outcomes. There have, however, been concerns raised about a risk of overgeneralisation, and associated scepticism about applicability to every circumstance [3, 4] associated with a pragmatic trial design.

The adequacy of reporting [5], design [6] and robustness [7] of clinical trauma and orthopaedic trials have also previously been called into question. Recent analysis [8] from trials published within a specific mainstream orthopaedic journal has identified general improvements in the quality

✉ Luke Farrow
luke.farrow@abdn.ac.uk

1 Institute of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK

2 Aberdeen Royal Infirmary, Aberdeen, UK

3 Edinburgh Royal Infirmary, Edinburgh, UK

4 Charing Cross Hospital, London, UK

and quality of analyses over time, but with trends towards smaller, single centre trial design.

However, several larger trials are now reported in high-impact non-orthopaedic medical journals, and thus were excluded from this previous analysis. As a result, there is little currently understood about the specifics of design, conduct and reporting related to these large-scale trauma and orthopaedic trials published in major general medical journals with a high impact factor.

We therefore set out to examine the quality of evidence produced from RCT's published within this setting. Given the high-impact and international influence of these journals it is integral that the literature produced is of sound methodological quality with low risk of bias in order to provide substantial high-quality evidence for interventions.

Materials and methods

Study selection

A systematic search of 5 major high-impact general medical journals (colloquially known as "the big five"—British Medical Journal (BMJ), Journal of American Medical Association (JAMA), New England Journal of Medicine (NEJM), the Lancet and Annals of Internal Medicine (Annals)) was performed from April 2010–April 2020 using online bibliographical archives on each journal website. These journals have a combined mean impact factor of 46.8 (https://academic-accelerator.com/Impact-Factor-IF/), far in excess of any trauma and orthopaedic speciality journal. They have previously been used to examine adequacy of trial design in other areas of healthcare and provide a gold standard reference for trial quality, given their exclusivity and publication standards [9]. Screening of full text articles was performed by one author (VK) and verified by another (LF). All articles pertaining to any area of trauma and orthopaedics that described a surgical treatment-based intervention randomised clinical trial were included. Those articles describing other surgical fields, or pertaining specifically to non-surgical interventions, were excluded.

Data collection

Data extraction was performed using a standardised proforma by three independent reviewers (VK, AA and TG). Any disagreement was mediated by a fourth individual (LF) until a communal decision was reached. Given the purpose of this study as a reflection of the available literature study authors were not contacted in the presence of missing data. All published or freely accessible data sources (for example, study protocols or trial monographs) for each study were, however, utilised. Data fields included in the analysis and extracted for each included trial are displayed in Table 1.

Statistical analysis

Descriptive analyses of overall trial characteristics were performed. N (%) was calculated for categorical variables, with median values and range presented for continuous variables given these were all non-normally distributed.

Comparison for the predicted control group event rate (identified from sample size calculation) versus the actual control group event rate was made for dichotomous outcomes. Assessment of study pragmatism was compared between significant and non-significant results utilising an unpaired 2-tailed t-test.

Assessment regarding risk of bias was performed for each trial using two measures:

1. The Cochrane Risk of Bias 2 tool [10], with each study summarised as either low, medium or high risk of bias.
2. The modified Delphi list [11], with a maximum score of 9 points. Only items assessed with a "yes" were given a score of 1 point. For the purposes of the study, scores 8–9 were considered high quality, scores 5–7 medium quality, scores 4–6 low quality and scores 1–3 very low quality.

Characterisation of individual article post-publication data was also performed. This included the study Altmetric attention score where available (www.altmetric.com/about-our-data/the-donut-and-score) and number of citations (https://scholar.google.co.uk/).

All statistical analyses were performed using R (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) and Microsoft Excel (Microsoft Corporation. (2018). Microsoft Excel, Washington, USA).

Results

Full details of the extracted information are located in supplementary tables 1–3. The summary results are displayed in Table 2. Overall, we identified 25 studies suitable for inclusion [12–36]. Of these studies, 9 pertained to trauma, 3 to elective hip surgery, 7 to elective knee surgery, 4 to spinal surgery and 2 to elective shoulder surgery. A greater proportion of trials—16/25 (64%) were identified in the latter half of the study period (2016–2020), than the early period—9/25 (36%). Most studies were lead from the UK 12/25 (48%), with 9/12 (75%) of these funded by the National Institute of Health Research. Patient recruitment was performed from a median of 9 centres (range 1–81). Blinding was present

Table 1 Data fields

Trial design	Sample size calculation	Results
Article title	Significance level	Actual sample size (defined as the number of patients with primary outcome data at the follow-up time point specified in the trial methodology)
Year of publication	Power	Number of patients in intervention and control arms
Country of origin	Predicted effect size of minimum clinically important difference (MCID)	Outcomes in intervention and control arms
Funding source	Use of MCID	Significance level of result
Intervention group	Appropriate justification for predicted effect size/MCID (e.g. Pilot RCT, modelling study or Delphi process pertaining to that specific disease or injury)	Reporting of 95% confidence intervals
Control group	Reporting of the population standard deviation if MCID used	Number of patients lost to follow up (defined as the number of patients without primary outcome assessment at the follow-up time point specified in the trial methodology)
Study design (superiority vs non-inferiority)	Required sample size calculated for assessment of the primary outcome	Early termination
Pragmatic versus explanatory (Pragmatic Explanatory Continuum Indicator Summary—PRECIS 2 tool)		Unanticipated sample size modification within trial
Blinding		Cochrane risk of bias assessment
Anatomical Region		Delphi list
Trauma vs Elective		
Number of centres involved		
International study		
Primary outcome		
Type of primary outcome (mortality vs complication vs patient reported outcome or functional score)		

in approximately half of trials 13/25 (52%); of these, 8/25 (32%) were single (assessor) blinded, and 5/25 (20%) were double blinded (patient and assessor). Regarding outcome assessment, the vast majority—22/25 (88%) utilised patient-reported outcome measures or functional scores as the primary outcome measure. Complication rate was only utilised in 3/25 (12%) trials as the primary outcome measure. No trials used mortality as the primary endpoint.

Regarding specific journals, 7 articles were published in the BMJ, 7 in the NEJM, 6 in the JAMA, 5 in the Lancet and none in the Annals. The mean Altmetric attention score© was 242 (range 22–681), and the mean number of citations per article was 230 (range 22–743).

Study pragmatism

The PRECIS-2 (Pragmatic Explanatory Continuum Indicator Summary—2) tool[37] was utilised to assess the pragmatism of included studies. Scores for individual domains in each trial are displayed in supplementary table 4. Overall, there was a high degree of pragmatism identified (mean aggregated score across all studies and domains 4.2/5). Studies with statistically significant results had a lower mean overall PRECIS-2 score compared to those with non-significant results (mean 3.71 vs 4.4, respectively; $p < 0.001$).

Sample size

All studies were set significance as $p < 0.05$. For studies with power data available, 15/23 (65.2%) utilised a power value of 80%, 7/23 (30.4%) a power value of 90%, and 1/23 (4.4%) a power value of 81.5%. Twenty-three out of 25 studies (92%) reported use of the MCID in order to perform their sample size calculation; however, only 14/25 studies (56%) had appropriate justification for use of the MCID or predicted effect size, and only 15/23 (65.2%) reported standard deviation in outcome for the target population when using MCID that is required to perform appropriate sample size calculations. Furthermore, only 14/25 studies (56%) achieved their target sample size for assessment of the primary outcome. Five out of 24 studies

Table 2 Summary of results

Study characteristics

Identified in latter half of study period (2016–2020) no. (%)	Superiority no. (%)	Pragmatic no. (%)	Any form of blinding no. (%)	Single assessor blinding no. (%)	Double assessor blinding no. (%)	Median no. centres (range)	International recruitment no. (%)	PROMs as primary outcome no. (%)	Complication rate as primary outcome no. (%)
16/25 (64)	23/25 (92)	18/25 (72)	13/25 (52)	8/25 (32)	5/25 (20)	9 (1–81)	2/25 (8)	22/25 (88)	3/25 (12)

Sample size

Power data reported no. (%)	80% power value no. (%)	90% power value no. (%)	81.5% power value no. (%)	Used MCID for power calculation no. (%)	Justification for MCID or predicated effect size no. (%)	Reported SD required for sample size calculation no. (%)	Achieved target sample size for assessment of primary outcome no. (%)	Made adjustments to sample size calculation during trial no. (%)
23/25 (92)	15/23 (65.2)	7/23 (30.4)	1/23 (4.4)	23/25 (92)	14/25 (56)	15/23 (65.2)	11/25 (44)	5/24 (20.8)

Results

Statistically significant ($p<0.05$) no. (%)	Trials with a non-significant result and underpowered no. (%)	Trials with dichotomous outcomes allowing for assessment of FI/RFI no. (%)	HEALTH study RFI (LTFU)	FAITH study RFI (LTFU)	WHIST study RFI (LTFU)	Potential for overturning of results depending on results of LTFU no. (%)
7/25 (28)	10/18 (55.6)	3/25 (12)	10 (29)	8 (383)	7 (29)	3/3 (100)

Comparison for the predicted control group event rate versus the actual control group event rate

HEALTH study	FAITH study	WHIST study
4.9% vs 8.3%	25% vs 21.8%	15% vs 6.7%

Cochrane risk of bias

Low risk no. (%)	Some risk no. (%)	High risk no. (%)	High risk due to deviations from intended interventions no. (%)	Some risk of bias due to use of PROMs without patient blinding no. (%)	High risk of bias due to use of PROMs without patient blinding no. (%)
3/25 (12)	18/25 (72)	4/25 (6)	3/4 (75)	16/25 (64)	

(20.8%) made amendments to the sample size calculation while the trial was ongoing.

Results

Seven out of 25 (28%) trials reported statistically significant results ($p < 0.05$) for the primary outcome. Seven out of 18 (38.9%) of those trials that reported non-significant results had an actual sample size for the primary outcome smaller than the predicted sample size, indicating a potential type II error. Only 3 trials reported dichotomous outcomes that allowed for assessment of the FI/RFI and comparison of the predicted control group event rate versus the actual control group event rate. Given all three reported non-significant results, the RFI was utilised. For the HEALTH study [16], the RFI was 10, with loss to follow-up (LTFU) of 29 patients. For the FAITH study [29], the RFI was 8, with LTFU of 383 patients. For the WHIST study [21], the RFI was 7, with LTFU of 29 patients. Results of all three trials could have been overturned dependent on the results of those lost to follow-up. With regard to the comparison of the predicted control group event rate versus the actual control group event rate, both the WHIST and HEALTH studies had differences > 50% (Table 2).

Risk of bias

The summary results for the Cochrane risk of bias analysis, including assessment for each included domain, are displayed in Table 3. Three out of 25 (12%) trials were adjudged to be at low risk of bias, 18/25 (72%) trials at some risk of bias, and 4/25 (6%) at high risk of bias. Three out of 4 (75%) of those studies judged at high risk of bias were due to deviations from the intended interventions. Sixteen out of 25 (64%) studies had at least some risk of bias in outcome measurement attributable to the use of PROMs without patient blinding where the outcome may have been influenced by knowledge of the intervention received. A summary bar chart of the percentage of risk by category is displayed in Supplementary Fig. 1.

The summary results for the Delphi list assessment are displayed in Table 4. Five out of 25 (20%) trials were adjudged to be high quality, and 20/25 (80%) of trials were designated as medium quality. No trials were assessed to be of either low or very low quality. The main reason for designation of lower study quality was lack of blinding to patient, care provider or assessor.

Discussion

Key findings from this analysis included that only a very small proportion of trauma and orthopaedic RCT's published in high-impact general medical journals were judged to be of high quality/low risk of bias. Many published trials did not achieve their target sample size for assessment of the primary outcome, and several did not describe appropriate techniques to justify use of the MCID for the intended intervention. We identified a high degree of study pragmatism, with a lower likelihood of statistically significant results for more pragmatic designs.

Despite these concerns, these studies were highly cited in the literature, with evidence for widespread dissemination according to Altmetric attention scores©. Knowledge of deficiencies in the design and reporting of trauma and orthopaedic RCT's can assist in the planning of future trials to improve scientific rigour and ensure widespread clinical applicability.

Patient-reported outcome measures

The vast majority of studies utilised PROMs, with the MCID as the primary method of defining the delta within the sample size calculation. The MCID is an important concept; however, there are a few potential issues that need to be addressed when considering its use in this context. The first is that there is no universal definition of how best to evaluate the MCID. The MCID produced varies by technique used and depends on the patient's baseline status, as well as study context [38]. Age is perhaps the most predominant example of this is, having previously been shown to influence baseline PROMs and response to surgical interventions for knee arthritis [39]. Use of PROMs with ceiling effects is also known to impart bias on outcomes following trauma and orthopaedic randomised clinical trials [40]. Given the identified widespread use of MCID, it is imperative future trials utilise appropriate techniques to ensure correct definition of the MCID in the population to be tested by the intervention.

It was notable that a significant proportion of trials (44%) did not achieve their target sample size for calculation of the primary outcome at the prespecified time point. This suggests that current estimates regarding participant retention are often incorrect. Overall, there was mean underestimation of eventual study sample size available for the primary outcome by approximately 5%, but this was as high as 28% in the studies by Frobell et al. [12] and Försth et al. [15], and above 10% in a number of others [14, 24, 25, 36]. Further careful consideration of factors potentially influential towards ongoing involvement or crossover is required during sample size calculation to ensure that sufficient recruitment. Guidelines for the conduct and reporting of RCT sample size calculations (DELTA 2 [41]) have previously been described and should be utilised to ensure a high probability of a study achieving its primary aim.

Table 3 Revised Cochrane risk-of-bias tool for randomised trials (RoB 2)

	Beard et al., 2018 (CSAW)	Beard et al., 2019 (TOPKAT)	Bhandari et al., 2017 (FAITH)	Bhandari et al., 2019 (HEALTH)	Clark et al., 2016 (VAPOUR)	Costa et al., 2012	Costa et al., 2014 (DRAFFT)	Costa et al., 2017 (UK FixDT)	Costa et al., 2020 (WHIST)	Firanescu et al., 2018 (VERTOS IV)	Försth et al., 2016	Frobell et al., 2010
Random sequence generation	green	green	yellow	green	green	green	green	green	green	green	green	green
Risk of bias due to deviations from the intended interventions	green	yellow	green	yellow	green	green	yellow	green	red	green	green	green
Risk of bias due to missing outcome data	green	green	yellow	green	green	green	yellow	green	green	green	green	green
Risk of bias in outcome measurement	yellow	yellow	green	yellow	green	yellow	green	yellow	green	green	yellow	yellow
Risk of bias in selection of the reported result	green	green	green	green	green	green	green	green	green	green	green	green
Overall risk of bias	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	red	green	yellow	yellow

Sample size

Identified smaller than predicted sample sizes are a concern for those trials with negative results (38.9%), where there is an associated risk of type II error. It is therefore difficult to determine whether the results for these trials were due to absence of evidence or actual evidence of absence of effect. We also identified significant differences in the predicted and actual effect size in the control group for included studies, which may additionally influence the ability to perform accurate outcome assessment. Future use of adaptive trial designs may help to eliminate some of these issues [42] and minimise research that does not achieve its desired intention [8].

Study pragmatism

Another notable finding from our results was the high degree of pragmatism (according to PRECIS-2) identified in the included trials and the fact that a greater degree of pragmatism in approach was associated with lower likelihood of a significant result. Other research has previously

Table 3 (continued)

	Frobell et al., 2013	Ghogawala et al., 2016	Griffin et al., 2014	Griffin et al., 2018	Katz et al., 2013	Paavola et al., 2018	Palmer et al., 2019	Rangan et al., 2015 (PROFHER)	Sihvonen et al., 2013	Skou et al., 2015	Van de Graaf et al., 2018 (ESCAPE)	Willett et al., 2016	Costa et al. 2018 (WOLFF)
Random sequence generation	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Yellow	Green	Green
Risk of bias due to deviations from the intended interventions	Green	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Red	Green	Green	Yellow
Risk of bias due to missing outcome data	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow
Risk of bias in outcome measurement	Yellow	Green	Yellow	Yellow	Yellow	Green	Red	Yellow	Green	Yellow	Yellow	Yellow	Yellow
Risk of bias in selection of the reported result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Overall risk of bias	Yellow	Yellow	Yellow	Yellow	Red	Yellow	Red	Yellow	Green	Red	Yellow	Yellow	Yellow

Green circle = Low risk, Yellow circle = Some risk, Red circle = High risk

highlighted how questions over the routine use of pragmatic trials may have had a role to play in the lack of translation from some trials towards change in clinical practice [43] and issues with trial recruitment [4, 44]. Surgeon and patient preference have both been shown to be influential in recruitment and retention to trauma and orthopaedic pragmatic RCTs [45]. It is vital that equipoise within the surgical community is established prior to embarking on any pragmatic trial, as recruitment bias and crossover remain a major concern. Use of the readiness assessment for pragmatic trials (RAPT) model may provide one method of determining suitability of an intervention for testing in a pragmatic trial [46]. We could not find any evidence of this tool having been used in trauma and orthopaedic research to determine the suitability of previously conducted pragmatic trials.

Table 4 Delphi list assessment

	1. Treatment allocation a) Was a method of randomisation performed?	1. Treatment allocation b) Was treatment allocation concealed?	2. Were the groups similar at baseline regarding the most important prognostic factors?	3. Were the eligibility criteria specified?	4. Were the outcome assessors blinded?	5. Was the care provider blinded?	6. Were point estimates and measures of variability presented for the primary outcome measures?	7. Did the analysis include an intention-to-treat analysis?	Overall score /9
Frobell et al. [12]	Yes	No	Yes	Yes	No	No	Yes	Yes	6
Costa et al. [26]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Sihvonen et al. [13]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	8
Frobell et al. [23]	Yes	No	Yes	Yes	No	No	Yes	Yes	6
Katz et al. [17]	Yes	No	Yes	Yes	No	No	Yes	Yes	6
Griffin et al. [23]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Costa et al. [35]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Rangan et al. [32]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Skou et al. [18]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Ghogawala et al. [14]	Yes	No	Yes	Yes	No	No	Yes	No	5
Försth et al. [15]	Yes	No	Yes	Yes	No	No	Yes	No	5
Willett et al. [34]	Yes	No	Yes	Yes	No	No	Yes	No	5
Clark et al. [20]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	8
Costa et al. [31]	Yes	Yes	Yes	Yes	No	No	Yes	Yes	7
Bhandari et al. 2017	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Paavola et al. [25]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	8
Firanescu et al. [27]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	8
Beard et al. [28]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	8
Griffin et al. [30]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Van der Graaf et al. [33]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Bhandari et al. [16]	Yes	No	Yes	Yes	No	No	Yes	Yes	6
Beard et al. [19]	Yes	No	Yes	Yes	No	No	Yes	Yes	6
Palmer et al. [24]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Costa et al. [21]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7
Costa et al. [36]	Yes	No	Yes	Yes	Yes	No	Yes	Yes	7

Alternative approaches to trial design

Another option that requires further exploration regarding utility in the domain of clinical Trauma & Orthopaedic research is a Bayesian approach to trial design. This technique has been increasingly used in the wider trial community [47, 48] and may provide significant potential benefit in the heterogeneous populations seen across the breadth of Trauma and Orthopaedics [49].

Strengths and limitations

Strengths of the study include the in-depth assessment of study quality and design for gold-standard benchmarks regarding the current state of orthopaedic research. Potential limitations to our study include that it may be possible that higher-quality evidence and lower risk of bias are seen in studies contained within other journals, but this is contradictory to what has been previously reported [8]. Post hoc power calculations were not conducted due to known methodological issues with this approach [50].

Applicability

We provide a summary of the literature with note of areas for improvement, but it should be clear that these concerns are not applicable to all included studies, and the use of many of the methods discussed such as the use of PROMs, the MCID and pragmatic trials is supported when utilised appropriately.

Conclusions

The majority of trauma and orthopaedic RCTs published in high-impact major medical journals have evidence of significant knowledge dissemination, but some notable concerns related to study quality.

We suggest the following changes may assist in future publication of low-risk trials: International co-operation in the development and funding of large-scale multi-centre randomised trials, appropriate calculation of the relevant MCID for the study hypothesis with use of widely validated PROMs, measures to improve trial retention, blinding of participants to intervention allocation when utilising PROMs, prior assessment of community equipoise (with potential increased use of explanatory trials where appropriate), and potential use of Bayesian approaches to trial design.

Caution should be used in the interpretation of highly pragmatic trials as these appear less likely to be associated with statistically significant results, although the exact nature of this relationship is unclear.

Availability of data and material (data transparency) Data are available on request.

Code availability (software application or custom code) No code was utilised.

Declarations

Conflicts of interest The authors have no relevant financial or non-financial interests to disclose.

Consent for participation/publication Not required.

Ethical approval (include appropriate approvals or waivers) Not required.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate (include appropriate statements) Not required.

Consent for publication (include appropriate statements) Not required.

References

1. Lohmander LS, Roos EM (2016) The evidence base for orthopaedics and sports medicine: scandalously poor in parts. Br J Sports Med 50:564–565

2. Matar HE, Platt SR (2019) Overview of randomised controlled trials in orthopaedic research: search for significant findings. Eur J Orthop Surg Traumatol 29:1163–1168

3. Haddad FS (2019) Which evidence will you act on? Bone Joint J 101-B:497–499

4. Robinson AHN, Johnson-Lynn SE, Humphrey JA, Haddad FS (2019) The challenges of translating the results of randomized controlled trials in orthopaedic surgery into clinical practice. Bone Joint J 101-B:121–123

5. Poolman RW, Struijs PA, Krips R, Sierevelt IN, Lutz KH, Bhandari M (2006) Does a "Level I Evidence" rating imply high quality of reporting in orthopaedic randomised controlled trials? BMC Med Res Methodol 6:44

6. Freedman KB, Back S, Bernstein J (2001) Sample size and statistical power of randomised, controlled trials in orthopaedics. J Bone Joint Surg Br 83:397–402

7. Checketts JX, Scott JT, Meyer C, Horn J, Jones J, Vassar M (2018) The robustness of trials that guide evidence-based orthopaedic surgery. J Bone Joint Surg Am 100:e85

8. Smith CS, Mollon B, Vannabouathong C, Fu JM, Sales B, Bhandari M et al (2020) An assessment of randomized controlled trial quality in the journal of bone & joint surgery: update from 2001 to 2013. J Bone Joint Surg Am 102:e116

9. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G (2016) the fragility index in multicenter randomized controlled critical care trials. Crit Care Med 44:1278–1284

10. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I et al (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366:l4898

11. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM et al (1998) The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 51:1235–1241

12. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS (2010) A randomized trial of treatment for acute anterior cruciate ligament tears. N Engl J Med 363:331–342

13. Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H et al (2013) Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 369:2515–2524

14. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN et al (2016) Laminectomy plus Fusion versus Laminectomy Alone for Lumbar Spondylolisthesis. N Engl J Med 374:1424–1434

15. Forsth P, Olafsson G, Carlsson T, Frost A, Borgstrom F, Fritzell P et al (2016) A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med 374:1413–1423

16. Investigators H, Bhandari M, Einhorn TA, Guyatt G, Schemitsch EH, Zura RD et al (2019) Total hip arthroplasty or hemiarthroplasty for hip fracture. N Engl J Med 381:2199–2208

17. Katz JN, Losina E (2013) Surgery versus physical therapy for meniscal tear and osteoarthritis. N Engl J Med 369:677–678

18. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O et al (2015) A randomized, controlled trial of total knee replacement. N Engl J Med 373:1597–1606

19. Beard DJ, Davies LJ, Cook JA, MacLennan G, Price A, Kent S et al (2019) The clinical and cost-effectiveness of total versus partial knee replacement in patients with medial compartment osteoarthritis (TOPKAT): 5-year outcomes of a randomised controlled trial. Lancet 394:746–756

20. Clark W, Bird P, Gonski P, Diamond TH, Smerdely P, McNeil HP et al (2016) Safety and efficacy of vertebroplasty for acute painful osteoporotic fractures (VAPOUR): a multicentre, randomised, double-blind, placebo-controlled trial. Lancet 388:1408–1416

21. Costa ML, Achten J, Knight R, Bruce J, Dutton SJ, Madan J et al (2020) Effect of incisional negative pressure wound therapy vs standard wound dressing on deep surgical site infection after surgery for lower limb fractures associated with major trauma: the WHIST randomized clinical trial. JAMA 323:519–526

22. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS (2013) Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. BMJ 346:f232

23. Griffin D, Parsons N, Shaw E, Kulikov Y, Hutchinson C, Thorogood M et al (2014) Operative versus non-operative treatment for closed, displaced, intra-articular fractures of the calcaneus: randomised controlled trial. BMJ 349:g4483

24. Palmer AJR, Ayyar Gupta V, Fernquest S, Rombach I, Dutton SJ, Mansour R et al (2019) Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: multicentre randomised controlled trial. BMJ 364:l185

25. Paavola M, Malmivaara A, Taimela S, Kanto K, Inkinen J, Kalske J et al (2018) Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. BMJ 362:k2860

26. Costa ML, Achten J, Parsons NR, Edlin RP, Foguet P, Prakash U et al (2012) Total hip arthroplasty versus resurfacing arthroplasty in the treatment of patients with arthritis of the hip joint: single centre, parallel group, assessor blinded, randomised controlled trial. BMJ 344:e2147

27. Firanescu CE, de Vries J, Lodder P, Venmans A, Schoemaker MC, Smeets AJ et al (2018) Vertebroplasty versus sham procedure for painful acute osteoporotic vertebral compression fractures (VERTOS IV): randomised sham controlled clinical trial. BMJ 361:k1551

28. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N et al (2018) Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet 391:329–338

29. Fixation using Alternative Implants for the Treatment of Hip fractures I, (2017) Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. Lancet 389:1519–1527

30. Griffin DR, Dickenson EJ, Wall PDH, Achana F, Donovan JL, Griffin J et al (2018) Hip arthroscopy versus best conservative care for the treatment of femoroacetabular impingement syndrome (UK FASHIoN): a multicentre randomised controlled trial. Lancet 391:2225–2235

31. Costa ML, Achten J, Griffin J, Petrou S, Pallister I, Lamb SE et al (2017) Effect of locking plate fixation vs intramedullary nail fixation on 6-month disability among adults with displaced fracture of the distal tibia: the UK FixDT randomized clinical trial. JAMA 318:1767–1776

32. Rangan A, Handoll H, Brealey S, Jefferson L, Keding A, Martin BC et al (2015) Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. JAMA 313:1037–1047

33. van de Graaf VA, Noorduyn JCA, Willigenburg NW, Butter IK, de Gast A, Mol BW et al (2018) Effect of early surgery vs physical therapy on knee function among patients with nonobstructive meniscal tears: the ESCAPE randomized clinical trial. JAMA 320:1328–1337

34. Willett K, Keene DJ, Mistry D, Nam J, Tutton E, Handley R et al (2016) Close contact casting vs surgery for initial treatment of unstable ankle fractures in older adults: a randomized clinical trial. JAMA 316:1455–1463

35. Costa ML, Achten J, Parsons NR, Rangan A, Griffin D, Tubeuf S et al (2014) Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: randomised controlled trial. BMJ 349:g4807

36. Costa ML, Achten J, Bruce J, Tutton E, Petrou S, Lamb SE et al (2018) Effect of negative pressure wound therapy vs standard wound management on 12-month disability among adults with severe open fracture of the lower limb: the WOLLF randomized clinical trial. JAMA 319:2280–2288

37. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M (2015) The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 350:h2147

38. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 7:541–546

39. Williams DP, Price AJ, Beard DJ, Hadfield SG, Arden NK, Murray DW et al (2013) The effects of age on patient-reported outcome measures in total knee replacements. Bone Joint J 95-B:38–44

40. Gagnier JJ, Johnston BC (2019) Poor quality patient reported outcome measures bias effect estimates in orthopaedic randomized studies. J Clin Epidemiol 116:36–38

41. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA et al (2018) DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. BMJ 363:k3750

42. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV et al (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 16:29

43. Pawson R (2019) Pragmatic trials and implementation science: grounds for divorce? BMC Med Res Methodol 19:176

44. Ramasamy A, Humphrey J, Robinson AHN (2019) The Ankle Injury Management (AIM) trial: the way forward or off target? Bone Joint J 101-B:1466–1468

45. Keding A, Handoll H, Brealey S, Jefferson L, Hewitt C, Corbacho B et al (2019) The impact of surgeon and patient treatment preferences in an orthopaedic trauma surgery trial. Trials 20:570

46. Baier RR, Jutkowitz E, Mitchell SL, McCreedy E, Mor V (2019) Readiness assessment for pragmatic trials (RAPT): a model to assess the readiness of an intervention for testing in a pragmatic trial. BMC Med Res Methodol 19:156

47. Ferreira D, Barthoulot M, Pottecher J, Torp KD, Diemunsch P, Meyer N (2020) Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review. Br J Anaesth 125:201–207

48. Angus DC, Derde L, Al-Beidh F, Annane D, Arabi Y, Beane A et al (2020) Effect of hydrocortisone on mortality and organ support in patients with severe COVID-19: the REMAP-CAP COVID-19 corticosteroid domain randomized clinical trial. JAMA 324:1317–1329

49. Jansen JO, Pallmann P, MacLennan G, Campbell MK, Investigators U-RT (2017) Bayesian clinical trial designs: another option for trauma trials? J Trauma Acute Care Surg 83:736–741

50. Plate JDJ, Borggreve AS, van Hillegersberg R, Peelen LM (2019) Post hoc power calculation: observing the expected. Ann Surg 269:e11