

Describing randomization in trials included in systematic reviews in orthopaedic surgery

From School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia

Correspondence should be sent to M. Tang matttang26@gmail.com

Cite this article:
Bone Jt Open 2024;5(12):1072–1080.

DOI: 10.1302/2633-1462.512.BJO-2024-0042.R1

M. Tang,¹ K. K. Lun,² A. M. Lewin,^{1,3} I. A. Harris^{1,3,4}

¹School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia

²The University of Sydney, Camperdown, Australia

³Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, Liverpool, Australia

⁴South Western Sydney Local Health District, Liverpool Hospital, Sydney, Australia

Aims

Systematic reviews of randomized controlled trials (RCTs) are the highest level of evidence used to inform patient care. However, it has been suggested that the quality of randomization in RCTs in orthopaedic surgery may be low. This study aims to describe the quality of randomization in trials included in systematic reviews in orthopaedic surgery.

Methods

Systematic reviews of RCTs testing orthopaedic procedures published in 2022 were extracted from PubMed, Embase, and the Cochrane Library. A random sample of 100 systematic reviews was selected, and all included RCTs were retrieved. To be eligible for inclusion, systematic reviews must have tested an orthopaedic procedure as the primary intervention, included at least one study identified as a RCT, been published in 2022 in English, and included human clinical trials. The Cochrane Risk of Bias-2 Tool was used to assess random sequence generation as ‘adequate’, ‘inadequate’, or ‘no information’; we then calculated the proportion of trials in each category. We also collected data to test the association between these categories and characteristics of the RCTs and systematic reviews.

Results

We included 917 unique RCTs. We found that 374 RCTs (40.8%) reported adequate sequence generation, 61 (6.7%) were inadequate, 410 (44.7%) lacked information, and 72 (7.9%) were observational studies incorrectly included as RCTs within the systematic review. Publication year, an author with statistical or epidemiological qualifications, and journal impact factor were each associated with adequate randomization. We found that 45 systematic reviews (45%) included at least one inadequately randomized RCT or an observational study incorrectly treated as a RCT.

Conclusion

There is evidence of a lack of random allocation in RCTs included in systematic reviews in orthopaedic surgery. The conduct of RCTs and systematic reviews should be improved to minimize the risk of bias from inadequate randomization in RCTs and mislabelling of non-randomized studies as RCTs.

Take home message

- The common, unrecognized inclusion of non-randomized studies in systematic reviews of randomized controlled trials means that the conclusions of orthopaedic systematic reviews may be biased and the findings unreliable.

Introduction

Systematic reviews of randomized controlled trials (RCTs) are the highest level of scientific evidence.¹ The cornerstone of RCTs is randomization, which ensures that the only difference between treatment groups is the allocated intervention,²⁻⁵ permitting the examination of causal relationships between interventions

and clinical outcomes.^{6,7} The unrecognized inclusion of non-randomized studies in systematic reviews of RCTs may undermine the quality of the evidence, bias the conclusions, and negatively affect patient care.^{5,8,9}

While there has been a consistent rise in the number of RCTs conducted in the orthopaedic literature,¹⁰ there is evidence of methodological flaws in orthopaedic RCTs, including use of inadequate randomization and failure to report method of randomization.^{8,11,12} While previous studies have assessed the quality of randomization in RCTs, there has been little consideration of the RCTs included in systematic reviews. There is also no prior assessment of the proportion of systematic reviews that include trials with inadequate randomization.

The primary aim of this study was to measure the proportion of RCTs included in systematic reviews of orthopaedic surgery that use an adequate method of randomization. The secondary aims were to determine if there is any association between trial characteristics and the use of an adequate randomization method; measure the proportion of systematic reviews that include trials with inadequate randomization; and determine if there is any association between systematic review characteristics and the inclusion of inadequately randomized trials.

Methods

Study design

We performed a meta-epidemiological study of all RCTs included in 100 systematic reviews in orthopaedic surgery. Ethics approval was not required for this study as all data are freely available in the public domain. This study was registered on PROSPERO (record ID CRD42023480758).¹³

Eligibility

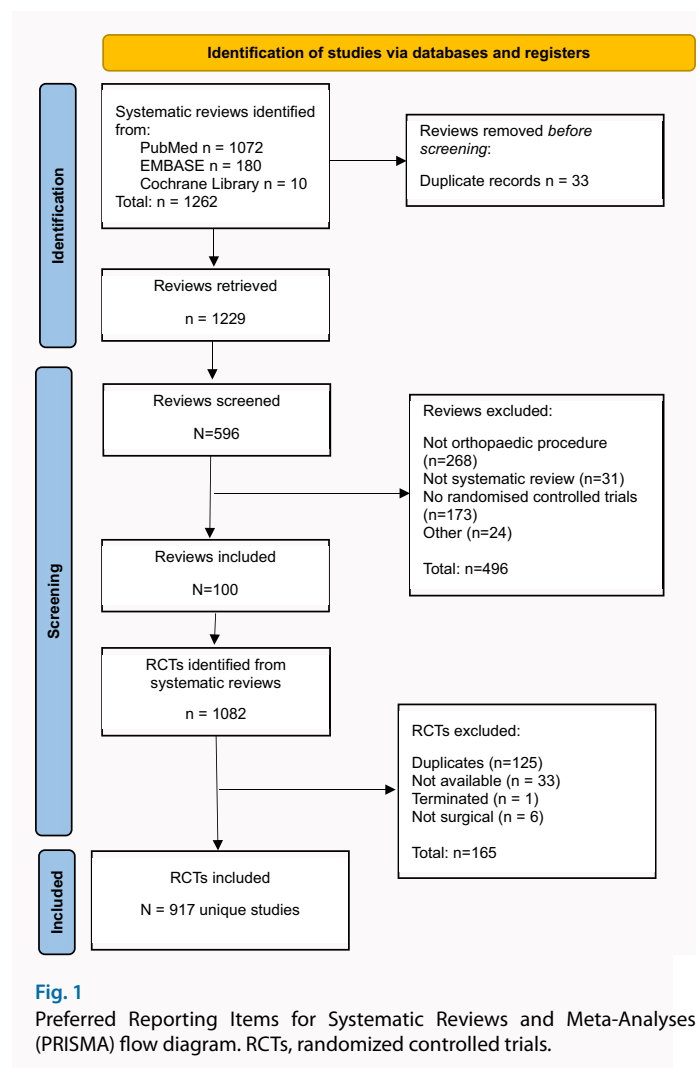
We included systematic reviews that tested an orthopaedic operative procedure as the primary intervention (defined as any procedure conducted in an operating theatre by an orthopaedic team which involved penetration of the skin); included at least one RCT; were published in 2022; and were available in English. We excluded systematic reviews of non-orthopaedic aspects of orthopaedic procedures, e.g. anaesthesia, perioperative medication, rehabilitation, or injections.

Information sources and search strategy

PubMed, Embase, and the Cochrane Library were searched on 30 April 2023. The search strategy is available in Supplementary Material 1.

Selection process

Search results were imported into bibliographic software where duplicates were removed. The resulting list of potentially eligible systematic reviews was sorted by first author surname and exported to Excel (Microsoft, USA), where we used the Power Query list.random() function, with seed set to 115, to generate a random number between 0 to 1 for each review. Reviews were then sorted in ascending order of the random number. Starting at the top of the list, two authors (MT, KKL) independently screened all studies by title and abstract, followed by a full-text review where applicable. The first 100 reviews to meet eligibility criteria were selected



to form the sample. Disagreements were resolved through discussion between the two authors at all stages and no consultation of a third author was required. A list of all studies retrieved and included is available in the supplement (Supplementary Material table). References were managed using Endnote X9 (Clarivate, USA).

Study outcomes

The primary outcome of the study was the proportion of RCTs included in systematic reviews of orthopaedic surgery that used adequate randomization to assign participants. The secondary outcomes were to determine if there was any association between trial characteristics (publication year, journal impact factor, inclusion of an author with statistical or epidemiological qualifications, sample size, country of research, intervention type, body region), and adequate randomization. We also determined the proportion of systematic reviews that included at least one trial with inadequate randomization. During data extraction, we encountered observational studies treated as RCTs, so we added this category to our study outcomes. We determined if there was any association between systematic review characteristics (use of a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist,¹⁴ registration in PROSPERO, inclusion of an author with statistical or epidemiological qualifications, systematic review

Table I. Characteristics of randomized controlled trials included in systematic reviews of surgical procedures.

Variable	Adequately randomized (n = 374)	Not adequately randomized (n = 471)*	Observational studies (n = 72)	Total (n = 917)
Year, n (%)				
Pre-2000	12 (3.2)	60 (12.7)	4 (5.6)	76 (8.3)
2000 to 2009	50 (13.4)	89 (18.9)	11 (15.3)	150 (16.4)
2010 to 2019	251 (67.1)	282 (59.9)	52 (72.2)	585 (63.8)
2020 to 2023	61 (16.3)	40 (8.5)	5 (6.9)	106 (11.6)
Author with statistical or epidemiological qualifications, n (%)				
	30 (8.0)	49 (10.4)	2 (2.8)	81 (8.8)
Impact factor				
Range	0.2 to 158.5	0.2 to 10.5	0.4 to 4.8	0.2 to 158.5
Median (IQR)	2.7 (1.9 to 3.8)	2.2 (1.5 to 3.0)	1.9 (1.2 to 2.8)	2.4 (1.6 to 3.2)
Sample size				
Range	8 to 2,948	20 to 1,000	10 to 5,390	8 to 5,390
Median (IQR)	80 (54 to 123.5)	80 (50 to 121.2)	60 (40.3 to 99.5)	80 (51 to 120)
Country of research, n (%)†				
China	51 (13.6)	54 (11.5)	13 (18.1)	118 (12.9)
USA	37 (9.9)	58 (12.3)	9 (12.5)	104 (11.3)
England	26 (7.0)	28 (5.9)	2 (2.8)	56 (6.1)
Sweden	13 (3.5)	38 (8.1)	1 (1.4)	52 (5.7)
South Korea	27 (7.2)	16 (3.4)	3 (4.2)	46 (5.0)
Germany	16 (4.3)	26 (5.5)	1 (1.4)	43 (4.7)
India	23 (6.2)	31 (6.6)	3 (4.2)	57 (6.2)
Canada	18 (4.8)	12 (2.6)	1 (1.4)	31 (3.4)
Japan	3 (0.8)	19 (4.0)	8 (11.1)	30 (3.3)
Norway	12 (3.2)	18 (3.8)	0 (0)	30 (3.3)
Other	148 (39.6)	171 (36.3)	31 (43.1)	350 (38.2)
Interventions, n (%)‡				
Arthroplasty	159 (42.3)	180 (38.2)	45 (62.5)	384 (41.9)
Fixation	110 (29.4)	173 (36.7)	12 (16.7)	295 (32.1)
Arthrodesis	14 (3.7)	32 (6.8)	4 (5.6)	50 (5.5)
Arthroscopy	30 (8.0)	23 (4.9)	2 (2.8)	55 (6.0)
Decompression	14 (3.7)	19 (4.0)	6 (8.3)	39 (4.3)
Other	47 (12.6)	44 (9.3)	3 (4.2)	94 (10.3)
Body region, n (%)§				
Hip	140 (37.4)	221 (46.9)	13 (18.1)	374 (40.8)
Knee	86 (23.0)	68 (14.4)	32 (44.4)	186 (20.3)
Shoulder	63 (16.8)	53 (11.3)	2 (2.8)	118 (12.9)
Spine	36 (9.6)	63 (13.4)	12 (16.7)	111 (12.1)
Other	49 (13.1)	66 (14.0)	13 (18.1)	128 (14.0)

*"No information" and "inadequate randomization".

†Only top ten countries presented for brevity.

‡Only top five interventions presented for brevity.

§Only top four body regions presented for brevity.

Table II. Association between adequate random sequence generation and trial characteristics.

Characteristic	Unadjusted OR (95% CI)	p-value*	Adjusted OR (95% CI)	p-value*
Publication year	1.06 (1.04 to 1.09)	< 0.001	1.04 (1.00 to 1.08)	0.033
Author with epidemiology or statistics qualification	2.22 (1.38 to 3.60)	0.001	1.78 (1.02 to 3.11)	0.043
Impact factor	1.35 (1.20 to 1.53)	< 0.001	1.18 (1.04 to 1.39)	0.029
Sample size	1.00 (1.00 to 1.00)	0.140	1.00 (1.00 to 1.00)	0.841
Country of research				
China	Reference			
USA	0.68 (0.34 to 1.18)	0.172	0.48 (0.23 to 1.01)	0.053
Canada	1.59 (0.70 to 3.70)	0.272	0.92 (0.35 to 2.46)	0.869
England	0.98 (0.51 to 1.90)	0.960	0.67 (0.28 to 1.56)	0.352
Germany	0.65 (0.31 to 1.34)	0.251	0.49 (0.21 to 1.13)	0.096
India	0.79 (0.40 to 1.52)	0.474	0.99 (0.33 to 3.03)	0.990
Japan	0.17 (0.04 to 0.53)	0.006	0.11 (0.02 to 0.39)	0.002
Norway	0.71 (0.30 to 1.60)	0.408	0.62 (0.20 to 1.86)	0.384
Other	0.92 (0.59 to 1.43)	0.698	0.69 (0.39 to 1.23)	0.213
South Korea	1.79 (0.87, 3.76)	0.118	1.42 (0.61, 3.40)	0.415
Sweden	0.36 (0.17 to 0.74)	0.007	0.24 (0.08 to 0.63)	0.005
Intervention				
Arthroplasty	1.19 (0.90 to 1.57)	0.212	1.12 (0.64 to 1.93)	0.692
Fixation	0.63 (0.48 to 0.84)	0.002	0.83 (0.47 to 1.46)	0.523
Arthrodesis	0.61 (0.33 to 1.08)	0.098	0.97 (0.34 to 2.73)	0.950
Arthroscopy	1.96 (1.16 to 3.35)	0.012	1.09 (0.46 to 2.63)	0.844
Decompression	1.06 (0.53 to 2.10)	0.860	2.32 (0.72 to 7.64)	0.158
Body region				
Hip	Reference			
Knee	2.00 (1.36 to 2.93)	< 0.001	1.82 (1.12 to 2.96)	0.016
Other	1.17 (0.76 to 1.79)	0.465	1.19 (0.60 to 2.35)	0.620
Shoulder	1.88 (1.23 to 2.87)	0.003	1.47 (0.75 to 2.90)	0.258
Spine	0.90 (0.57 to 1.42)	0.661	0.57 (0.21 to 1.52)	0.264

*Two-tailed Wald test.
OR, odds ratio.

journal impact factor, number of included studies, and country of research), and the inclusion of inadequately randomized trials or observational studies treated as RCTs.

Data extraction

One author (MT) extracted study characteristics from systematic reviews and RCTs (and observational studies treated as RCTs). Another author (AML) independently audited a 10% sample of extracted data.

Assessment of randomization

Random sequence generation was assessed using the Cochrane Risk of Bias tool for randomized trials (RoB-2).¹⁵ Based on RoB-2, we considered a study to have adequate randomization if a random component was used in

the sequence generation process, e.g. “computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots”.¹⁵ We used this definition because it requires the least subjective assessment, thus improving replicability. We classified studies as inadequately randomized if no random element was used or the sequence was predictable e.g. “alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method”.¹⁵ Studies where the only information was a statement that the study was randomized or where the approach to sequence generation is described incompletely, without confirming there was a random component, were classified

Table III. Characteristics of included systematic reviews.

Characteristic	Total (n = 100)
Author with statistical or epidemiological qualifications, n (%)	20 (20)
Journal impact factor	
Range	1 to 8.4
Median (IQR)	2.5 (2.3 to 3.5)
Number of included studies	
Range	1 to 115
Median (IQR)	6 (3 to 11)
Use of PRISMA checklist, n (%)	85 (85)
Registered in PROSPERO, n (%)	38 (38)
Country of research, n (%)*	
China	24 (24)
England	13 (13)
USA	12 (12)
Canada	6 (6)
Italy	6 (6)
Germany	5 (5)
Ireland	4 (4)
Japan	4 (4)
Other	27 (27)

*Only top eight countries presented for brevity.

as “no information” and treated as neither adequately nor inadequately randomized. For studies identified and treated as RCTs by the systematic review authors but which were observational studies, we classified those as “Not RCT”.

Other study variables

Author qualification was determined from the full-text article and drawn from their included position, title, or associated organization. Sample size was the total number of randomized participants. Journal impact factor was determined from the Journal Citation Report in the year of publication. Country of research was the country from which the participants were recruited, or the review was conducted. Intervention type(s) and body region(s) were assigned from pre-determined categories. For systematic reviews we also assessed the number of included studies, inclusion of a PRISMA checklist from the full-text article, and supplemental data. PROSPERO registration was determined if there was inclusion of a PROSPERO registration number.

Statistical analysis

We analyzed RCTs and systematic reviews separately. We described study characteristics using frequency and proportions for categorical variables and median and IQR for continuous variables. We calculated the proportion of RCTs in each randomization category. After excluding observational studies treated as RCTs, we used logistic regression to assess the association between study characteristics and adequate randomization. For categorical variables, the most frequent category was set as the reference level for analysis.

Table IV. Quality of randomization in randomized controlled trials and proportion of systematic reviews including studies by randomization quality.

Variable	Randomization quality in included RCTs (n = 917)	Systematic reviews with at least one study with randomization quality rated (n = 100)
Adequately randomized, n (%)	374 (40.8)	83 (83)
Not adequately randomized, n (%)	61 (6.7)	29 (29)
No information, n (%)	410 (44.7)	84 (84)
Not a RCT, n (%)	72 (7.9)	32 (32)

RCT, randomized controlled trial.

All variables were retained in a multivariable regression model. Results were reported as odds ratios (ORs) with 95% CIs. We calculated the proportion of systematic reviews which included at least one inadequately randomized RCT or observational study treated as a RCT. We used logistic regression to assess the association between review characteristics and the inclusion of at least one inadequately randomized RCT or observational study treated as a RCT. Results were reported as ORs with 95% CIs. P-values for logistic regression coefficients were computed using two-tailed tests, comparing the z-ratio to the standard normal distribution as implemented in the R's `glm()` function. Statistical analysis was performed using R statistical software v. 4.2.2 (R Foundation for Statistical Computing, Austria).

Results

Study selection

A total of 1,262 systematic reviews were retrieved in the search; after removal of duplicates there were 1,229 eligible systematic reviews. A total of 596 systematic reviews were screened in order to find 100 systematic reviews that met the inclusion criteria. Reasons for exclusion are shown in [Figure 1](#). Of 1,082 RCTs included in the 100 systematic reviews, 165 were excluded (125 duplicates; 33 studies were not available; 1 study was terminated and showed no results; and six were not surgical), leaving 917 unique RCTs ([Figure 1](#)).

Characteristics of included RCTs

Characteristics of the 917 included RCTs are provided in [Table I](#).

Quality of randomization in included RCTs

Of the 917 studies, 374 (40.8%) reported adequate sequence generation, 61 (6.7%) were inadequately randomized, and 410 (44.7%) included no information about the method of randomization. A further 72 studies (7.9%) were observational studies (not RCTs) treated as RCTs by systematic review authors.

Table V. Association between inclusion of inadequate randomized controlled trials (RCTs) or observational studies treated as RCT and systematic review characteristics.

Characteristic	Unadjusted OR (95% CI)	p-value*	Adjusted OR (95% CI)	p-value*
PRISMA checklist				
No inclusion	Reference			
Inclusion	0.49 (0.15 to 1.48)	0.211	0.99 (0.20 to 4.92)	0.992
PROSPERO registration				
No registration	Reference			
Registered	0.83 (0.36 to 1.86)	0.649	1.09 (0.38 to 3.19)	0.875
Author with statistical or epidemiological qualifications				
No	Reference			
Yes	2.14 (0.80 to 6.01)	0.136	4.50 (1.19 to 19.6)	0.033
Systematic review impact factor				
	1.26 (0.94 to 1.77)	0.136	1.14 (0.66 to 1.93)	0.619
Number of included studies				
	1.12 (1.05 to 1.22)	0.003	1.13 (1.05 to 1.24)	0.005
Country of research				
China	Reference			
Canada	0.14 (0.01 to 1.07)	0.097	0.05 (0.00 to 0.68)	0.046
England	0.45 (0.11 to 1.74)	0.252	0.12 (0.01 to 0.90)	0.055
Germany	2.86 (0.35 to 60.5)	0.379	0.64 (0.02 to 24.0)	0.793
Ireland	0.71 (0.07 to 6.79)	0.756	0.89 (0.08 to 9.93)	0.918
Italy	0.71 (0.11 to 4.57)	0.713	1.06 (0.14 to 8.20)	0.953
Japan	2.14 (0.23 to 46.9)	0.534	2.41 (0.22 to 57.6)	0.499
Other	0.38 (0.12 to 1.17)	0.096	0.23 (0.04 to 1.04)	0.064
USA	0.36 (0.08 to 1.46)	0.164	0.46 (0.08 to 2.39)	0.371

*Two-tailed Wald test.

OR, odds ratio; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Association between adequate randomization and RCT characteristics

The adjusted and unadjusted associations between the characteristics of the RCTs and adequate randomization are provided in [Table II](#).

Characteristics of included systematic reviews

Characteristics of the 100 systematic reviews are provided in [Table III](#).

Proportion of systematic reviews including studies with different randomization ratings

There were 29 reviews (29%) that included at least one trial with inadequate randomization, while 32 (32%) reviews included at least one observational study treated as a RCT ([Table IV](#)). The proportion of systematic reviews including at least one inadequately randomized RCT or observational study treated as a RCT was 45 (45%).

Association between inclusion of inadequately randomized trials or observational studies treated as RCTs and systematic reviews

The adjusted and unadjusted associations between systematic review characteristics and inclusion of adequately randomized

trials or observational studies treated as RCTs are provided in [Table V](#).

Discussion

In this meta-epidemiological study of RCTs included in orthopaedic surgery systematic reviews, we found only 40.8% of included RCTs reported adequate methods of random sequence generation. An increase in publication year, journal impact factor, and inclusion of an author with a statistical or epidemiological qualification were positively associated with adequate randomization of studies. We also found that 6.7% of included RCTs were inadequately randomized, and a further 7.9% were observational studies treated as RCTs by the systematic review authors. Importantly, nearly half of all systematic reviews of RCTs included at least one RCT that was not randomized.

Interpretation of results

Our findings on randomization are consistent with previous studies in orthopaedic surgery.^{8,16-18} A study in 2020 by Smith et al⁸ found that 8.4% of studies published in the *Journal of Bone & Joint Surgery* from 2001 to 2013 did not use adequate sequence generation. Our study expands on this, describing characteristics in studies from 1977 to 2023 from

213 unique journals. The high proportion of RCTs that did not include information regarding the method of randomization is consistent with an study of 304 RCTs registered on Clinicaltrials.gov between 2011 and 2012 with the keyword “surgery”, which found that 34% of trials failed to state the methods used to generate the random sequence.¹⁷ The proportion of studies reporting adequate randomization was larger than a study of Chinese language trial databases that reported that only 7% of studies identified as RCTs were truly randomized.¹⁹ This difference may be due to different criteria to define adequate randomization or differences between studies published in Chinese from those published in English. The inclusion of observational studies as RCTs in reviews may be due to inadequate assessment of RCT quality by the systematic review authors.

We found the odds of adequate randomization increased with publication year. The Consolidated Standard of Reporting Trials (CONSORT) statement was initially published in 1996,²⁰ with updates in 2001 and 2010 aiming to increase the transparency of trial reporting to allow for scrutiny by audiences.²⁰ It is likely that widespread adoption of the CONSORT statement has driven improvements and greater transparency in the reporting of RCTs.^{21–23} We also found greater odds of adequate random sequence generation with increasing impact factor. While impact factor cannot be considered a sole representation of journal or study quality,²⁴ this finding may be due to higher impact factor journals having a larger pool of submitted articles and thus having an opportunity to select the highest-quality studies for publication. It may also reflect different reviewer and editorial practices between journals. These findings are consistent with previous studies. A study of 163,129 RCTs in PubMed using a machine learning (ML) algorithm found that more recent publication years and higher journal impact factor were both associated with decreased incidence of questionable research practices.²⁵ A similar ML study of 176,620 trials published between 1966 and 2018 in PubMed found a lower risk of bias of randomization in journals with an impact factor larger than ten.²⁶

Our study found a positive association between inclusion of research-qualified authors and the use of adequate random sequence generation, consistent with a previous study that found an association between including an author with an epidemiological degree and CONSORT statement compliance.¹⁶

While 14.6% of all supposed trials were either inadequately randomized RCTs or observational studies treated as RCTs, these trials were overrepresented in the systematic reviews, with 45% of systematic reviews including at least one inadequately randomized RCT or observational study treated as an RCT. While previous studies have investigated the reporting quality of RCTs and systematic reviews separately, to our knowledge, our study is the first to assess the quality of RCTs included in systematic reviews in orthopaedic surgery.

We found that increasing number of included studies in systematic reviews was associated with greater odds of including inadequate RCTs or observational studies treated as RCTs. This may reflect a higher chance of including inadequately randomized RCTs or observational study treated as a RCT when more studies are included.

Strengths and limitations

A strength of this study was the use of an objective definition of adequate randomization based on the Cochrane RoB-2 Tool.⁵ Only the trials explicitly rated as having adequate randomization according to the tool, using specific free-text descriptions in the reporting of randomization, were considered to be adequately randomized for our study. This increases replicability, though it may underestimate the use of adequate randomization. Our study uses a broad definition of adequate randomization, including methods requiring human input (i.e. coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots); however, it is important to acknowledge that higher-quality journals may require more stringent methods. High-quality randomization involves techniques that are not susceptible to external manipulation, such as through use of computer-generated random numbers. In addition, our assessment of randomization relies on the quality and accuracy of reporting of the original studies. It is possible that included studies may have been misclassified in our study due to inadequate reporting. It is also possible that study authors reported different randomization methods to what was used in a study. However, the aim of this study was to provide a high-level overview of quality of randomization in orthopaedic surgery systematic reviews. Another strength is that our sample of RCTs, while being a selected sample that may not reflect general RCT quality, potentially represents higher impact studies due to their inclusion in systematic reviews, which inform clinical practice guidelines. Finally, our study involved two authors reviewing and extracting data.

A limitation of the study was our sample size of 100 systematic reviews, which proved underpowered to ascertain associations between systematic review characteristics and inclusion of inadequately randomized RCTs (or observational studies treated as RCTs). However, the inclusion of 917 RCTs provided adequate power to assess these associations at the RCT level. Another limitation was our method of assessing randomization, as we relied on the quality of reporting in the published trials. There is possible discordance between the methodological quality of a study and how it was reported, leading to an underestimate of adequately or inadequately randomized studies. However, as adherence to reporting standards and guidelines has improved in recent years, our results may overestimate the quality of RCTs included in systematic reviews in orthopaedic surgery because studies are more likely to use a lower-quality methodology than their reporting suggests.^{27,28}

Implications

The finding that a large proportion of systematic reviews of RCTs of orthopaedic procedures include either inadequately randomized RCTs or non-randomized observational studies suggests that the level of evidence upon which many clinical decisions rely is lower than expected. Poor critical appraisal at the systematic review level and methodological understanding at the RCT level is an example of avoidable research waste.²⁹ Importantly, including non-randomized trials may result in misleading conclusions in systematic reviews,^{5,8,9,30} which may then impact practice, given the reliance on systematic reviews in clinical practice guidelines.

It is important that consumers of medical literature are aware that an appreciable number of trials in orthopaedic

surgery lack adequate randomization; caution must be applied to the interpretation of results and the conclusions of such trials, and the large proportion of systematic reviews that include these trials.³⁰ Knowledge of the factors which are associated with adequate randomization in trials can support the appraisal of evidence. Additionally, practitioners must be aware of poor research methodology and critically assess the evidence guiding their practice.

Future research

Future studies could focus whether including inadequately randomized trials impacts systematic review outcomes and further clinical practice guidelines. Given the influence of systematic reviews on clinical practice guidelines, it is important to determine the extent to which low-quality, methodologically flawed RCTs may undermine the quality of evidence that informs guidelines.

In conclusion, this study has provided insight into the quality of RCTs found in systematic reviews in orthopaedic surgery. A large proportion of reviews include either inadequately randomized RCTs or observational studies that are incorrectly treated as RCTs. These results identify a need for better research methods in both clinical trials and reviews of such trials.

Supplementary material

Microsoft Excel file detailing the results of all searches, and table showing the search strategy.

References

1. Soucacos PN, Johnson EO, Babis G. Randomised controlled trials in orthopaedic surgery and traumatology: overview of parameters and pitfalls. *Injury*. 2008;39(6):636–642.
2. Berger VW, Bour LJ, Carter K, et al. A roadmap to using randomization in clinical trials. *BMC Med Res Methodol*. 2021;21(1):168.
3. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. *Trials*. 2015;16:405.
4. Phillips MR, Kaiser P, Thabane L, Bhandari M, Chaudhary V, Retina Evidence Trials InterNational Alliance (R.E.T.I.N.A.) Study Group. Risk of bias: why measure it, and how? *Eye (Lond)*. 2022;36(2):346–348.
5. Higgins J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken, New Jersey, USA: John Wiley & Sons, 2019.
6. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. *BJOG*. 2018;125(13):1716.
7. Lai D, Wang D, McGillivray M, Baajour S, Raja AS, He S. Assessing the quality of randomization methods in randomized control trials. *Healthcare (Amst)*. 2021;9(4):100570.
8. Smith CS, Mollon B, Vannabouathong C, et al. An assessment of randomized controlled trial quality in the *Journal of Bone & Joint Surgery*: update from 2001 to 2013. *J Bone Joint Surg Am*. 2020;102-A(20):e116.
9. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev*. 2017;6(1):131.
10. Matar HE, Platt SR. Overview of randomised controlled trials in orthopaedic research: search for significant findings. *Eur J Orthop Surg Traumatol*. 2019;29(6):1163–1168.
11. Chess LE, Gagnier J. Risk of bias of randomized controlled trials published in orthopaedic journals. *BMC Med Res Methodol*. 2013;13:76.
12. Li P, Mah D, Lim K, Sprague S, Bhandari M. Randomization and concealment in surgical trials: a comparison between orthopaedic and non-orthopaedic randomized trials. *Arch Orthop Trauma Surg*. 2005;125(1):70–72.
13. Tang M, Lun KK, Lewin AM, Harris IA. PROSPERO 2023 CRD42023480758. https://www.crd.york.ac.uk/prospéro/display_record.php?ID=CRD42023480758 (date last accessed 21 November 2024).
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:71.
15. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
16. Adie S, Harris IA, Naylor JM, Mittal R. CONSORT compliance in surgical randomized trials: are we there yet? A systematic review. *Ann Surg*. 2013;258(6):872–878.
17. Chapman SJ, Aldaffa M, Downey CL, Jayne DG. Research waste in surgical randomized controlled trials. *Br J Surg*. 2019;106(11):1464–1471.
18. Robinson NB, Fremes S, Hameed I, et al. Characteristics of randomized clinical trials in surgery from 2008 to 2020: a systematic review. *JAMA Netw Open*. 2021;4(6):e2114494.
19. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials*. 2009;10:46.
20. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
21. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol*. 2007;60(3):241–249.
22. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev*. 2012;1:60.
23. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials*. 2016;17(1):301.
24. Law R, Leung D. Journal impact factor: a valid symbol of journal quality? *Tour Econ*. 2020;26(5):734–742.
25. Damen JA, Heus P, Lamberink HJ, et al. Indicators of questionable research practices were identified in 163,129 randomized controlled trials. *J Clin Epidemiol*. 2023;154:23–32.
26. Vinkers CH, Lamberink HJ, Tjiddink JK, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLoS Biol*. 2021;19(4):e3001162.
27. Patole S. Systematic reviews, meta-analysis, and evidence-based medicine. In: Patole S, ed. *Principles and Practice of Systematic Reviews and Meta-Analysis*. Springer International Publishing, Cham, 2021: 1–10.
28. Page MJ, Altman DG, Shamseer L, et al. Reproducible research practices are underused in systematic reviews of biomedical interventions. *J Clin Epidemiol*. 2018;94:8–18.
29. Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ*. 2015;350(mar24 20):h809.
30. Ioannidis JPA. Hundreds of thousands of zombie randomised trials circulate among us. *Anaesthesia*. 2021;76(4):444–447.

Author information

M. Tang, BSc(Med) Hons, Medical Student (Honours), School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia.

K. K. Lun, BMedSc, MD, Senior Resident, The University of Sydney, Camperdown, Australia.

A. M. Lewin, MSc, Research Fellow, Epidemiologist, School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia; Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, Liverpool, Australia.

I. A. Harris, MBBS, MMed (Clin Epi), MSc (Hlth Data Sc), PhD, Professor of Orthopaedic Surgery, Honorary Professor, Institute for Musculoskeletal Health, Director, Co-Director, Surgical Outcomes Research Centre (SOuRCe), Co-Chair, Musculoskeletal

Network, ACI (Agency for Clinical Innovation), School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia; Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, Liverpool, Australia; South Western Sydney Local Health District, Liverpool Hospital, Sydney, Australia.

Author contributions

M. Tang: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.
K. K. Lun: Data curation, Investigation, Writing – review & editing.
A. M. Lewin: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.
I. A. Harris: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Funding statement

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

ICMJE COI statement

The authors have no conflicts of interest to disclose.

Data sharing

The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Ethical review statement

This study used publicly available, published data, and thus ethics approval was not required.

Open access funding

The open access fee for this article was self-funded.

© 2024 Tang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>