Assessment of 30 Years of Randomized Controlled Trials in *The American Journal of Sports Medicine:* 1990-2020

Ajay Shah,* MD, Graeme Hoit,*† MD, PhD(c), Lucy Lan,‡ BHSc, and Daniel B. Whelan,*§‖ MD, MSc
Investigation performed at the Division of Orthopaedic Surgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada

Background: Randomized controlled trials (RCTs) stand atop the evidence-based hierarchy of study designs for their ability to arrive at results with the lowest risk of bias. Even for RCTs, however, critical appraisal is essential before applying results to clinical practice.

Purpose: To analyze the quality of reporting of RCTs published in *The American Journal of Sports Medicine* (*AJSM*) from 1990 to 2020 and to identify trends over time and areas of improvement for future trials.

Study Design: Systematic review; Level of evidence, 1.

Methods: We queried the *AJSM* database for RCTs published between January 1990 and December 2020. Data pertaining to study characteristics were recorded. Quality assessments were conducted using the Detsky quality-of-reporting index and the modified Cochrane risk-of-bias (mROB) tool. Univariate and multivariable models were generated to establish factors with associations to study quality. The Fragility Index was calculated for eligible studies.

Results: A total of 277 RCTs were identified with a median sample size of 70 patients. A total of 19 RCTs were published between 1990 and 2000 (t_1); 82 RCTs between 2001 and 2010 (t_2); and 176 RCTs between 2011 and 2020 (t_3). From t_1 to t_3, significant increases were observed in the overall mean-transformed Detsky score (from 68.2% ± 9.8% to 87.4% ± 10.2%, respectively; $P < .001$) and mROB score (from 4.7 ± 1.6 to 6.9 ± 1.6, respectively; $P < .001$). Multivariable regression analysis revealed that trials with follow-up periods of <5 years clearly stated primary outcomes, and a focus on the elbow, shoulder, or knee were associated with higher mean-transformed Detsky and mROB scores. The median Fragility Index was 2 (interquartile range, 0-5) for trials with statistically significant. Studies with small sample sizes (<100 patients) were more likely to have low Fragility Index scores and less likely to have a statistically significant finding in any outcome.

Conclusion: The quantity and quality of published RCTs published in *AJSM* increased over the past 3 decades. However, single-center trials with small sample sizes were prone to fragile results.

Keywords: critical appraisal; evidence-based medicine; quality appraisal; research methodology; sports medicine

When making treatment decisions, orthopaedic surgeons must consider patient preferences and values, along with their own clinical experience and expertise, all integrated with the best available evidence. Atop the hierarchy of study designs sits the randomized controlled trial (RCT), as it is thought to minimize bias by controlling for it as well as for confounding factors.[28] Over time, there has been a shift in the orthopaedic sports community away from anecdote and opinion toward evidence-based medicine, with increasing demand that treatments are based on best evidence, ideally derived from RCTs.[27]

Previous studies have demonstrated a higher level of evidence within sports medicine literature compared with other orthopaedic and surgical subspecialties,[33] with a greater proportion of randomized and prospective study designs.[4,5,9,10] However, quantity does not necessarily equal quality, and the strength of conclusions drawn from this literature may be compromised by conflicting evidence from small, underpowered trials or those of poor methodological quality.[14,26,34] Accordingly, critical appraisal of the literature is an essential step before making inferences from study results and applying that to clinical practice.

The purpose of the present study was to identify and examine the quality of all RCTs published in *The American Journal of Sports Medicine* (*AJSM*) from 1990 to 2020 and to identify trends in study quality over time and areas of improvement for future clinical trials. The Fragility Index—a measurement of the robustness of statistically significant findings—and its associated variables were another important outcome.[30] It was hypothesized that the

quality of RCTs in *AJSM* would have increased over the past 3 decades and that the Fragility Index value would be superior to that of other published RCTs in orthopaedic sports medicine.

METHODS

Study Selection and Data Extraction

A search was conducted on the *AJSM* website (http://www.ajsm.org) for RCTs published between January 1990 and December 2020. All other study types (cohort studies, case-control studies, case series, case reports, meta-analyses, and reviews) were excluded. Two investigators (A.S., L.L.) independently reviewed eligible trial abstracts to identify trials with patients randomly allocated to interventions. The abstract screening was then followed by a full-text review. Discrepancies between reviewers were resolved by consensus discussion, involving independent review by the senior authors (D.B.W., G.H.) when an agreement could not be reached.

The following variables were extracted from each included RCT: first author's profession; study type; cited statistical support or support by an epidemiology department; location of the trial; whether it was multicentered; financial support; body region; category of intervention; prior trial registration (protocols cross-referenced with ClinicalTrials.gov for outcomes); allocation concealment; and blinding of outcome assessors and statistically significant ($P < .05$) findings.

Quality Assessment

The quality assessments for each study were conducted independently by 2 research associates (A.S., L.L.), with discrepancies resolved by consensus agreement after discussion or independent review by the senior authors. Trials were reviewed using the Detsky quality-of-reporting index and the modified Cochrane risk-of-bias (mROB) tool, which were considered the 2 primary outcome measures.[17] The Detsky score evaluates the quality of reporting based on 14 questions covering 5 categories, each worth 4 points for a total possible score of 20 (Supplemental Table S1, available separately).[11] The score was then converted into a percentage (mean-transformed Detsky score). Studies scoring >75% on the transformed score were considered high quality.

The mROB assessment evaluates the methodological quality of the study based on the following 10 categories: (1) randomization; (2) allocation concealment; (3) orthopaedic surgeon or treatment provider blinding; (4) assessor blinding; (5) patient blinding; (6) patient follow-up; (7) selective outcome reporting; (8) objectivity of outcomes; (9) adequate sample size; (10) and orthopaedic surgeon experience with treatment. The maximum score on this scale is 10 points, indicating a low risk of bias. Trials scoring ≥8 of 10 points on the mROB assessment were considered high quality.

Fragility Index

Studies with a statistically significant finding in any reported dichotomous outcome were selected for the Fragility Index calculation. The Fragility Index for each outcome was calculated according to the method described by Walsh et al[30] using 2×2 contingency tables. The P value for each outcome was first recalculated using a 2-sided Fisher exact test. We then added events to the group with a smaller number of events while subtracting nonevents from the same group to keep the total number of participants constant. Events were added iteratively until the calculated P value became $> .05$. The smallest number of additional events required to obtain $P > .05$ was the Fragility Index for that outcome.

Statistical Analysis

The kappa statistic (κ) was used to calculate the level of agreement between reviewers for the inclusion of studies. An a priori κ criterion of >0.65 was selected to indicate adequate agreement.[8] The intraclass correlation coefficient (ICC) with a 95% CI was used to calculate interrater agreement for the mROB assessment and the Detsky score. Descriptive statistics were calculated, with categorical variables presented as proportions and continuous data presented as means with standard error of the mean (SEM).

All statistical tests were 2-tailed, and significance was set at $P < .05$. The primary analysis examined the effect of independent variables on the dependent variables (mean-transformed Detsky score and mROB). Analysis of variance (ANOVA) with a Bonferroni correction was used to account for multiple comparisons, and independent Student t tests were used to compare the differences in the mean-transformed Detsky scores and mROB scores. Variables significantly associated with study quality in the univariate analyses for either quality assessment tool were

ǁAddress correspondence to Daniel B. Whelan, MD, MSc, Division of Orthopaedic Surgery, Department of Surgery, University of Toronto, 149 College Street, Room 508-a, Toronto, Ontario, M5T 1P5, Canada (email: db_whelan@hotmail.com).
*Division of Orthopaedic Surgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada.
†Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.
‡Michael G. DeGroote School of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.
§Division of Orthopaedic Surgery, Saint Michael's Hospital, Unity Health, Toronto, Ontario, Canada.

TABLE 1
Characteristics of the Included RCTs (N = 277)[a]

Variable	No. of Studies	Variable	No. of Studies
First author profession		Follow-up time	
Surgeon	192 (69.3)	<4 wk	18 (6.5)
Professor/researcher	23 (8.3)	1 to <12 mo	69 (24.9)
PT/kinesiologist	30 (10.8)	12 to <24 mo	65 (23.4)
MD (eg, sports, PMR)	21 (7.6)	24 to <36 mo	70 (25.2)
Trainee/other	11 (4)	36 to <60 mo	13 (4.7)
First author gender		≥5 y	42 (15.2)
Male	226 (82.6)	Number of sites	
Female	46 (16.6)	Single	239 (86.3)
Unknown	5 (1.8)	Multiple/cluster	38 (13.7)
Type of intervention		Financial support[c]	
Drug	79 (28.5)	None	105 (37.9)
Surgical	120 (43.3)	Conflict of interest	62 (22.4)
Nonsurgical[b]	78 (28.2)	Grant	91 (32.9)
Placebo controlled		Industry funded	71 (25.6)
Yes	98 (35.4)	Statistical support	88 (31.8)
No	179 (64.6)	Trial registered	185 (66.8)
PRP-related study		Protocol published	21 (7.6)
Yes	34 (12.3)	Primary outcome clearly stated	166 (59.9)
No	242 (87.4)	Follow-up of previously published trial	41 (14.8)
Area of body		Significant findings	
Shoulder	48 (17.3)	Of primary outcome	72 (26.0)
Elbow	9 (3.2)	Of secondary outcome	137 (49.5)
Hip/thigh	10 (3.6)	Of any outcome	166 (59.9)
Knee/leg	144 (52.0)		
Foot/ankle	40 (14.4)		
Multiple/injury prevention	26 (9.4)		
Trial location			
North America	80 (28.9)		
South America	6 (2.2)		
Africa	1 (0.4)		
Europe	124 (44.8)		
Asia	33 (11.9)		
Australia/Oceania	26 (9.4)		
Multiple	7 (2.5)		

[a]Data are presented as n (%). Conflict of interest indicates ≥1 author reporting a financial conflict of interest in the author disclosures. Statistical support indicates the support of an epidemiologist or a statistician in the acknowledgment or among the listed authors. MD, medical doctor; PMR, physical medicine and rehabilitation; PRP, platelet-rich plasma; PT, physical therapist; RCT, randomized controlled trial.
[b]Nonsurgical treatments included rehabilitation studies, injury prevention, and laboratory- or imaging-based studies.
[c]Categories are not mutually exclusive.

included in a multivariable linear regression model, with results reported as beta coefficients with 95% CIs.

Studies were grouped into 3 time periods, each spanning 1 decade: t_1 (1990-2000); t_2 (2001-2010); and t_3 (2011-2020). The chi-square test and ANOVA were used to determine whether there were significant differences between the trials within each decade for the previously stated categorical and continuous independent variables, respectively. Linear regression was used to assess for significant changes in the transformed Detsky scores and mROB scores over time. Similarly, the association between the Fragility Index with sample size, funding, trial registration, number of centers, and Detsky and mROB scores was evaluated with the Mann-Whitney U test or the Kruskal Wallis test for categorical variables and the Pearson correlation coefficient (r) for continuous variables. The correlations were grouped as follows: $r < 0.20$ = no correlation; $0.20 < r < 0.40$ = weak correlation; $0.40 < r < 0.60$ = moderate correlation; and $r > 0.60$ = strong correlation. All analyses were performed using SAS Version 9.4 (SAS Institute Inc).

RESULTS

Study Identification and Characteristics

A total of 7143 citations were published in *AJSM* between January 1990 and December 2020. After the exclusion of 6866 nonrandomized trials, 277 RCTs (3.9%) were included in our analysis (Table 1 and Supplemental Table S2). The
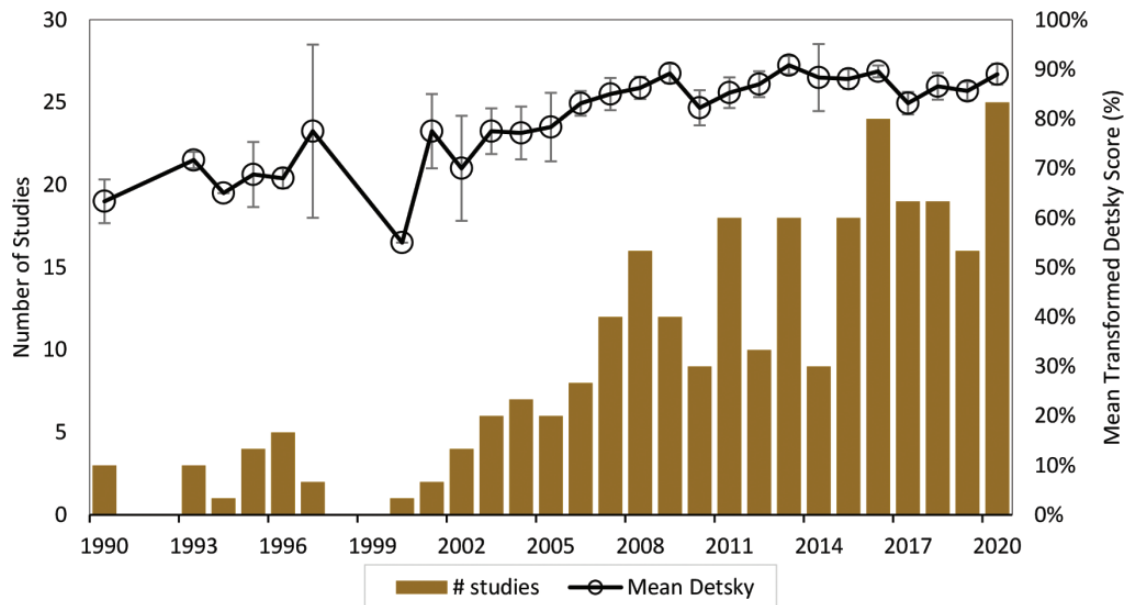
Figure 1. Number of randomized controlled trials published in *AJSM* and the mean Detsky score from 1990 to 2020. Error bars represent SEM. The Pearson correlation for the number of studies versus the year of publication, $r = 0.89$; and for the mean-transformed Detsky score versus year of publication, $r = 0.83$. *AJSM, The American Journal of Sports Medicine;* ICC, Pearson correlation coefficient.

TABLE 2
Characteristics of Trials Across Decades of Publication[a]

	t_1 (1990-2000)	t_2 (2001-2010)	t_3 (2011-2020)	P^b
Publications	19 (6.9)	82 (29.6)	176 (63.5)	
Sample size	259.7 ± 449.7	134.4 ± 216.3	129.3 ± 309.6	.195
Sample size, median [IQR]	83 [30-156]	71 [50-100]	70 [40-110]	.101
Significant findings	8 (42.1)	33 (40.2)	69 (39.2)	.960
Multicenter trials	2 (10.5)	16 (19.5)	27 (15.3)	.542
Received no funding	9 (47.4)	37 (45.1)	59 (33.5)	.141
Statistical support	8 (42.1)	25 (30.5)	55 (31.3)	.601
Study type				.224
Surgical	5 (26.3)	38 (46.3)	77 (43.8)	
Nonsurgical	8 (42.1)	29 (35.4)	42 (23.9)	
Drug	6 (31.6)	15 (18.3)	57 (32.4)	
Detsky score	68.2 ± 9.8	82.7 ± 11.6	87.4 ± 10.2	• t_1 vs t_2: MD = 14.6; ***P < .001*** • t_1 vs t_3: MD = 19.3; ***P < .001*** • t_2 vs t_3: MD = 4.7; ***P < .001***
mROB	4.7 ± 1.6	6.4 ± 1.7	6.9 ± 1.6	• t_1 vs t_2: MD = 1.67; ***P < .001*** • t_1 vs t_3: MD = 2.13; ***P < .001*** • t_2 vs t_3: MD = 0.47; ***P < .001***

[a]Data are presented as n (%) or mean ± SEM unless otherwise indicated. Bold *P* values indicate statistically significant differences between decades ($P < .05$). MD, mean difference; mROB, modified Cochrane risk-of-bias.

[b]3×2 chi-square tests were used for categorical variables and 1-way analysis of variance was used for continuous variables, followed by unpaired *t* test pairwise comparisons for variables with $P < .05$.

agreement between the reviewers regarding the eligibility of the studies was almost perfect ($\kappa = 0.99$).

The 277 RCTs published in *AJSM* between 1990 and 2020 demonstrated an increasing trend in the number of trials published over time (Figure 1). The annual number of studies published and the year of publication were strongly correlated ($r = 0.89$). A total of 19 RCTs were published between 1990 and 2000 (t_1), 82 RCTs between 2001 and 2010 (t_2), and 176 RCTs between 2011 and 2020 (t_3) (Table 2).

The mean sample size of included trials was 139.7 ± 18 patients (range, 10-3611 patients). The median sample size was 70 patients; 201 studies (72.6%) had <100 patients. Regression analysis showed a trend to decreased sample sizes from 1990 to 2020 ($\beta = -3.8$ [95% CI, 1.4 to -9.0]; $P = .15$). Increasing sample size was associated with a greater likelihood of a statistically significant result in any outcome (mean difference, 92.5 patients; $P = .011$).

An a priori sample size calculation was completed in 203 (73.3%) of the included trials. Of trials that showed an a priori sample size calculation, 137 (67.5%) enrolled a sufficient number of patients to achieve statistical power and 75 (36.9%) reported maintaining the required sample size at the follow-up. Of the 172 trials that had authors who reported financial support or conflicts of interest, 71 (41.3%) received funding or grants from industry.

Statistically significant results in any study outcome were reported in 166 trials (59.9%). Of these 166 trials, there was a significant finding in the primary outcome of 72 trials (43.4%). The correlation between Detsky and mROB scores was moderate ($r = 0.67$). The Science Citation Index weakly correlated with the Detsky score ($r = -0.14$) and the mROB score ($r = -0.14$). All other individual study variables are reported in Supplemental Table S3.

Assessment of the Detsky Index Quality Score

The ICC for interrater agreement on the Detsky score was 0.82 (95% CI, 0.64-1), indicating very high agreement (Supplemental Table S4). The mean-transformed Detsky score was 84.7% ± 0.7% (Figure 1). One trial (0.4%) scored <50%, 65 trials (23.5%) scored between 50% and 75%, and 211 trials (76.2%) scored >75%.

Univariate analyses demonstrated significant associations between the Detsky score and the type of intervention, a clearly stated primary outcome, a priori trial registration, the area of body studied, length of follow-up, type of financial support, and use of platelet-rich plasma (PRP) (Table 3). Multivariable linear regression analysis subsequently demonstrated significant independent associations between improved Detsky scores and follow-up durations of <5 years; trials on the shoulder, elbow, knee, or foot/ankle (reference: multiple/injury prevention); a priori trial registration; and a clearly stated primary outcome (Table 4).

Detsky scores significantly increased over time between 1990 and 2020 ($\beta = 3.5$ [95% CI, 2.5-4.5]; $P < .001$). The overall mean-transformed Detsky score increased significantly from t_1 (68.2% ± 9.8%) to t_2 (82.7% ± 11.6), and again from t_2 to t_3 (87.4% ± 10.2%) ($P < .001$ for both) (see Table 2). The Detsky score was strongly correlated with the year of publication ($r = 0.83$). The mean sample size, proportion of multicenter collaborations, number of industry-funded studies, and significant findings did not change over time (see Table 2).

Risk-of-Bias Assessment

The overall interrater agreement for the mROB score was 0.88 (95% CI, 0.72-1), corresponding to a very high

agreement (Supplemental Table S4). The mean mROB assessment score was 6.6 ± 0.1 points (Figure 2). The domains of "treatment-administrator blinding" (30/277) and "loss to follow-up >5%" (86/277) had the lowest scores, indicating a prevalent risk of study bias in these categories (Supplemental Table S5).

Univariate analysis showed a significant association with mROB scores and the type of trial, placebo-controlled comparison group, clearly stated primary outcome, a priori trial registration, number of study centers, area of body studied, length of follow-up, type of financial support, use of PRP, and those reporting results of a previous trial ($P < .05$) (see Table 3). Multivariate regression analysis showed that trials investigating the shoulder, elbow, or knee (reference: multiple/injury prevention), with follow-ups of <4 weeks, 1 to 12 months, 12 to 24 months, and 24 to 36 months (reference: >5 years), or a clearly stated primary outcome were associated with higher mROB scores (see Table 4).

The mROB scores significantly increased over time between 1990 and 2020 ($\beta = 0.07$ [95% CI, 0.04-0.10]; $P < .001$). The mean mROB score significantly increased from t_1 (4.7 ± 1.6) to t_2 (6.4 ± 1.7), and again from t_2 to t_3 (6.9 ±1.6) ($P < .001$ for both) (Table 2). The mROB score was moderately correlated with the year of publication ($r = 0.76$).

Fragility Index

The median Fragility Index was 2 (interquartile range, 0-5) for the 44 included studies, with significant findings in dichotomous outcomes (Supplemental Figure S1 and Table 5). Using the 2-sided Fisher exact test, 13 studies became nonsignificant when the P value was calculated, and therefore had a Fragility Index of 0. Increasing the Fragility Index value (indicating less fragility) was associated with a sample size of ≥100 patients ($P = .002$), a clearly stated primary outcome ($P = .010$), and a statistically significant finding in the primary outcome ($P = .020$) (see Table 4). The number of patients lost to follow-up was greater than the Fragility Index score in 75% (33/44) of studies. The Fragility Index was moderately correlated with the sample size ($r = 0.68$). The Fragility Index was not correlated with the transformed Detsky score ($r = 0.23$) or the mROB score ($r = 0.16$).

DISCUSSION

In examining all RCTs published in *AJSM* over 30 years, it was demonstrated that the mean methodological quality of RCTs in *AJSM* is relatively high and has increased over time. Multivariable analysis revealed that trials with follow-up periods of <5 years, a clearly stated primary outcome, and a focus on either elbow, shoulder, or knee were associated with higher mean-transformed Detsky and mROB scores. The median Fragility Index of studies with statistically significant findings was 2, and the number of patients lost to follow-up was greater than the Fragility Index in 75% of studies.

TABLE 3
Univariate Analysis of Characteristics Associated With Quality Scores[a]

Variable	Detsky Score, %	P[b]	mROB Score	P[b]	Variable	Detsky Score, %	P[b]	mROB Score	P[b]
Area of body		**.046**		**.010**	Financial support		.391		.331
Shoulder	88 ± 1.5		7.2 ± 0.2		Yes	84 ± 1.1		6.5 ± 0.2	
Elbow	90.6 ± 3.6		7.8 ± 0.3		No	85.2 ± 0.9		6.7 ± 0.1	
Hip/thigh	88.5 ± 3.6		7.2 ± 0.6		Industry funded		.355		.153
Knee/leg	83.3 ± 1		6.4 ± 0.1		Yes	85.8 ± 1.3		6.8 ± 0.2	
Foot/ankle	85.1 ± 1.6		6.4 ± 0.3		No	84.3 ± 0.8		6.5 ± 0.1	
Multiple/injury prevention	82.5 ± 2.6		6 ± 0.4		First author profession		.185		.174
Type of intervention		**.037**		**<.001**	Surgeon	83.7 ± 0.9		6.5 ± 0.1	
Drug	86.5 ± 1.3		7.5 ± 0.2		Professor/researcher	87.4 ± 2.2		6.8 ± 0.4	
Nonsurgical	86.1 ± 1.3		6.5 ± 0.2		PT/kinesiologist	87.8 ± 1.8		6.5 ± 0.2	
Surgical	82.7 ± 1.1		6.1 ± 0.2		MD (eg, sports, PMR)	87.6 ± 1.9		7.4 ± 0.4	
Follow-up time		**<.001**		**<.001**	Trainee/other	83.6 ± 4.0		6.2 ± 0.8	
<4 wk	82.5 ± 3.1		7.4 ± 0.5		First author gender		.532		.625
1 to <12 mo	87.5 ± 1.3		7 ± 0.2		Male	84.6 ± 0.8		6.5 ± 0.1	
12 to <24 mo	86.8 ± 1.2		7 ± 0.2		Female	85.9 ± 1.6		6.8 ± 0.3	
24 to <36 mo	84.8 ± 1.4		6.4 ± 0.2		Unknown	80 ± 4.7		6.8 ± 0.4	
36 to <60 mo	85 ± 3.1		6.3 ± 0.4		Location of trial		.209		.519
>5 y	78.3 ± 2		5.5 ± 0.2		North America	83.4 ± 1.5		6.6 ± 0.2	
PRP-related study		**<.001**		**<.001**	South America	94.2 ± 2.4		6.8 ± 0.2	
Yes	90.1 ± 1.4		7.6 ± 0.2		Africa	90 ± 0		7 ± 0	
No	84 ± 0.8		6.4 ± 0.1		Europe	83.9 ± 1.1		6.4 ± 0.2	
Trial registered		**<.001**		**<.001**	Asia	87.6 ± 1.6		6.9 ± 0.3	
Yes	89.8 ± 0.7		7.2 ± 0.1		Australia/Oceania	86.0 ± 1.9		6.6 ± 0.3	
No	82.2 ± 1.3		6.3 ± 0.2		Multiple	87.9 ± 3.4		7.7 ± 0.4	
Primary outcome clearly stated		**<.001**		**<.001**	Statistical support		.205		.452
Yes	87.9 ± 0.8		7 ± 0.1		Yes	85.6 ± 1.4		6.7 ± 0.2	
No	80 ± 1.2		6 ± 0.2		No	84.3 ± 0.8		6.5 ± 0.1	
Follow-up of previously published trial		**<.001**		**<.001**	Protocol published		.067		.059
Yes	77.8 ± 2.1		5.7 ± 0.2		Yes	88.8 ± 2.4		7.1 ± 0.3	
No	85.9 ± 0.7		6.7 ± 0.1		No	82.7 ± 0.7		6.4 ± 0.1	
Authors disclosed COI		**.022**		**.012**	Significant findings				
Yes	87.7 ± 1.2		7.1 ± 0.2		Of primary outcome		.110		.244
No	83.9 ± 0.8		6.4 ± 1.8		Yes	89.2 ± 1.2		7.2 ± 0.2	
Grant funding		**.002**		.292	No	86.9 ± 1		6.8 ± 0.2	
Yes	87.8 ± 1		6.7 ± 0.2		Of secondary outcome		.701		.909
No	83.2 ± 0.9		6.5 ± 0.1		Yes	85 ± 1.1		6.6 ± 0.2	
Placebo controlled		.782		**.039**	No	84.4 ± 1		6.6 ± 0.1	
Yes	85.5 ± 1.2		6.9 ± 0.2		Of any outcome		.058		.389
No	84.3 ± 0.9		6.4 ± 0.1		Yes	85 ± 1		6.6 ± 0.1	
Number of sites		.826		.093	No	80.1 ± 1.1		6.9 ± 0.2	
Single	86.2 ± 0.7		6 ± 0.4						
Multiple/cluster	84.5 ± 1.9		6.6 ± 0.1						

[a]Scores are reported as mean ± SEM. Bold P values indicate variables with statistically significant differences within subgroups ($P < .05$); these variables were included in the multivariable analysis (Table 4). COI, conflict of interest; MD, medical doctor; mROB, modified Cochrane risk-of-bias; PMR, physical medicine and rehabilitation; PRP, platelet-rich plasma; PT, physical therapist.
[b]Unpaired t tests for categories with 2 variables and 1-way analysis of variance for categories with >2 variables.

The present findings reflect similar results from a recent review of all surgical RCTs published in a high-impact general orthopaedic journal[29] from 1988 to 2013, which also noted a decrease in sample sizes over time despite increasing numbers of RCTs and improved study quality. The trend has also been observed in other surgical subspecialties.[1,7,35] A previous appraisal of the quality of all studies published in *AJSM* was conducted in 2016 by Brophy et al.[5] They identified an increase in the number of RCTs published and the level of evidence from the 1991-1993 and 2001-2003 periods to the 2011-2013 period. This study was limited by only sampling 3-year periods and generalizing several qualitative parameters as a proxy for methodological quality. At that time, the authors called for a more comprehensive study to assess parameters of quality across a wider breadth of published studies utilizing standardized and validated methodological quality instruments,[5] as performed in the present study.

TABLE 4
Multivariable Analysis of Characteristics Associated With Quality Scores[a]

Variable	Detsky Score		mROB Score	
	β (95% CI)	*P*	β (95% CI)	*P*
Area of body				
Shoulder	10.9 (5 to 16.9)	**<.001**	1.4 (0.5 to 2.3)	**.002**
Elbow	10.3 (1.7 to 19)	**.018**	1.5 (0.2 to 2.9)	**.02**
Hip/thigh	7.4 (−1.2 to 16)	.09	0.7 (−0.6 to 2.1)	.28
Knee/leg	8.2 (2.8 to 13.6)	**.003**	1 (0.2 to 1.8)	**.02**
Foot and ankle	6.6 (0.9 to 12.2)	**.022**	0.6 (−0.3 to 1.4)	.20
Multi/injury prevention	Reference		Reference	
Type of intervention				
Drug	−3 (−7.6 to 1.4)	.18	0.6 (−0.1 to 1.3)	.08
Nonsurgical	2.4 (−1.5 to 6.4)	.23	0.3 (−0.3 to 1.0)	.32
Surgical	Reference		Reference	
Follow-up time				
<4 wk	3.3 (−3.6 to 10.3)	.34	1.8 (0.8 to 2.7)	**<.001**
1 to <12 mo	9.2 (4.7 to 13.7)	**<.001**	1.5 (0.8 to 2.1)	**<.001**
12 to <24 mo	8.4 (4 to 12.8)	**<.001**	1.5 (0.9 to 2.1)	**<.001**
24 to <36 mo	6.4 (1.7 to 11.2)	**<.001**	1 (0.3 to 1.6)	**<.001**
36 to <60 mo	6.6 (−1.3 to 14.6)	.089	0.8 (−0.2 to 1.8)	.109
≥5 y	Reference		Reference	
PRP				
Yes	2.9 (−2.5 to 8.2)	.29	−0.04 (−0.8 to 0.8)	.91
No	Reference		Reference	
Trial registered				
Yes	4.2 (1.2 to 7.2)	**.007**	0.4 (−0.1 to 0.8)	.12
No	Reference		Reference	
Primary outcome clearly stated				
Yes	5.9 (2.9 to 8.8)	**<.001**	0.8 (0.3 to 1.2)	**<.001**
No	Reference		Reference	
Follow-up of previously published trial				
Yes	−0.9 (−4.6 to 2.8)	.62	0.08 (−0.5 to 0.6)	.78
No	Reference		Reference	
Authors disclosed COI				
Yes	2.5 (−0.9 to 5.8)	.15	0.2 (−0.2 to 0.7)	.34
No	Reference		Reference	
Grant funding				
Yes	2.4 (−0.5 to 5.3)	.11	—	
No	Reference		—	
Placebo controlled				
Yes	—		0.1 (−0.6 to 0.3)	.55
No	—		Reference	

[a]Dashes indicate variables not included in the analysis. Bold *P* values indicate statistical significance (*P* < .05). COI, conflict of interest; mROB, modified Cochrane risk-of-bias; multi, multiple; PRP, platelet-rich plasma.

Both the Detsky and mROB quality metrics showed relatively high study quality of published RCTs from 1990 to 2020. Identification of prevalent strengths and weaknesses within trial quality can help guide clinicians, researchers, and reviewers in performing and publishing high-quality research within sports medicine going forward. For example, we found that clearly stating a primary outcome was associated with higher quality on all metrics. This alludes to the authors' understanding of the research process and a structured, scientific approach to writing and reporting the trial. Based on this result, those aiming to answer orthopaedic sports medicine questions through a randomized trial should ensure that a primary outcome is identified

before the initiation of the research and that it is communicated in their paper.

During the data analysis, it was noted that the Detsky and mROB tools have several potential shortcomings in the context of assessing surgical trials. For example, the mROB tool places significant emphasis on blinding. However, a trial with a surgical versus nonsurgical intervention, in which neither the orthopaedic surgeon nor the patient can be blinded, is penalized by 3 points (30% of the total score). Additionally, no quality score incorporates a length of follow-up as a measure of strength despite the importance of long-term comparisons for surgical interventions. There is penalty for loss to follow-up of >5%, which
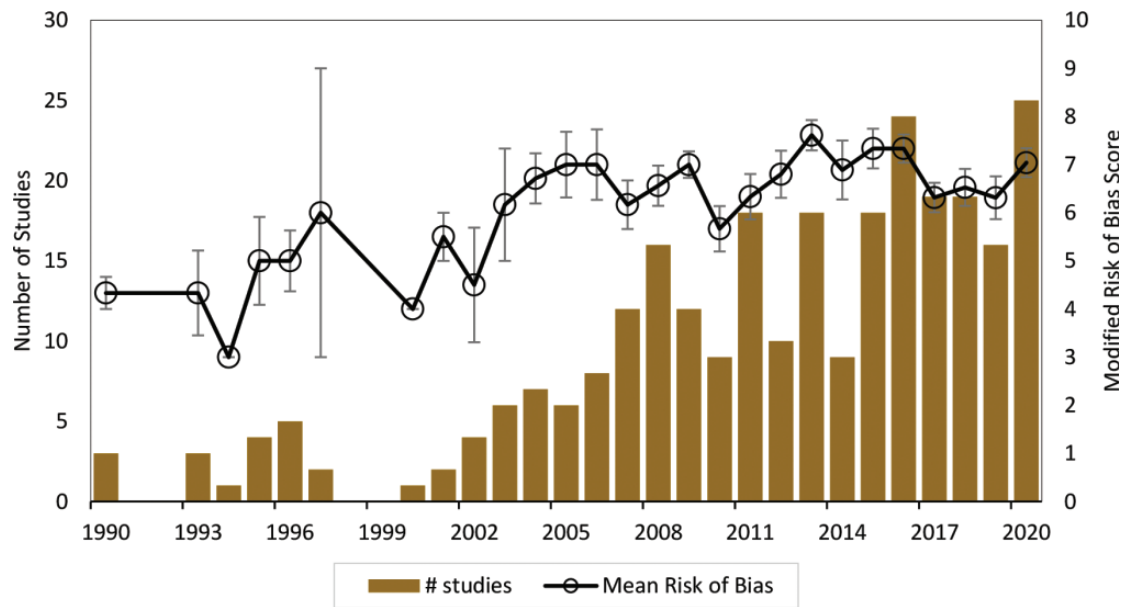
Figure 2. Number of randomized controlled trials published in *AJSM* and the mean mROB score, 1990 to 2020. The Pearson correlation coefficient for the number of studies versus the year of publication, $r = 0.89$; for the mROB score versus the year of publication, $r = 0.76$. *AJSM, The American Journal of Sports Medicine*; mROB, modified Cochrane risk-of-bias.

TABLE 5
Fragility Index Values and Study Characteristics[a]

Studies With Significant Findings in a Categorical Variable	Fragility Index	P[b]
All trials (N = 44)	2 [0-5]	
Outcome[c]		**.020**
Primary (n = 6)	7.5 [4-21.5]	
Secondary (n = 9)	0 [0-3.5]	
Other (n = 29)	2 [0-4]	
Sample size		**.002**
<100 (n = 20)	0.5 [0-2]	
≥100 (n = 24)	4 [2-10.5]	
A priori power calculation and sufficient patient recruitment		.24
Yes (n = 31)	2 [1-5]	
No (n = 13)	0 [0-5.5]	
Industry funding		.423
Yes (n = 11)	2 [2-4]	
No/unclear (n = 33)	2 [0-5.5]	
Number of centers		.076
Single (n = 40)	1 [0-4]	
Multiple/cluster (n = 14)	4 [2-7.5]	
Trial registered in database		.103
Yes (n = 16)	2 [1.25-11.5]	
No (n = 28)	1.5 [0-4.75]	
Primary outcome clearly stated		**.010**
Yes (n = 27)	2.5 [2-9.75]	
No (n = 17)	0 [0-1.75]	

[a]Data are presented as median [interquartile range]. Bold P values indicate statistically significant differences within subgroups ($P < .05$).
[b]Kruskal Wallis tests for variables of >2 categories and Mann-Whitney U tests for variables of 2 categories.
[c]Trials with significant findings in any outcome were included in the Fragility Index calculation for that outcome.

disproportionately affects trials with a longer follow-up due to their increased propensity to lose more patients. This is seen in our finding that trials with follow-ups of <3 years

had higher-quality scores. A lack of correlation between Detsky and mROB scores with other proxies for study quality, such as the Fragility Index, Citation Index, and sample

size/multicenter collaboration, was observed. One weakness of both tools is that they combine assessments of methodological quality with the quality of reporting into a composite score. It is important to distinguish between them—a trial that is poorly designed with notable bias but is well reported can receive a high-quality score, and vice versa.[25] Unfortunately, all well-known methodological quality questionnaires for RCTs have some flaws, primarily because of the clinical settings in which they were developed.[7,15,16]

Given the shortcomings of the quality assessment scores utilized to determine a high-quality grade for the RCTs we analyzed, other metrics may shed light on the confidence with which we can draw inferences from the results of these studies. The Fragility Index assessment highlights possible shortcomings of studies with small sample sizes and their robustness. For example, 13 of 44 studies reporting statistically significant results had a Fragility Index of 0, meaning that when the analysis was performed using a more conservative Fisher exact test, they were shown to be nonsignificant. Studies with a sample size of $\geq$100 patients had a median Fragility Index of 0.5, meaning that only 1 patient changing to a nonevent would alter the study's conclusions. It is interesting to note that, despite larger sample sizes being associated with a greater likelihood of a statistically significant difference in study outcomes, the mean RCT sample size in *AJSM* has shown a trend to decrease (β = –3.8 [95% CI, 1.4 to –9.0]; P = .15). The median Fragility Index of 2 is comparable with other RCTs in orthopaedic sports medicine and spinal surgery but lags behind orthopaedic trauma (Fragility Index = 5) and far behind internal medicine subspecialty trials published in high-impact factor journals (eg, *New England Journal of Medicine*, *The Lancet*, *Journal of the American Medical Association*, *BMJ*, and *Annals of Internal Medicine*) (Fragility Index = 13).[12,13,21-23]

Within the time frame we examined, small sample sizes (<50 patients; n = 75 studies) and a high proportion of single-center trials (86.3%) were observed, and there was a nonsignificant trend toward smaller mean sample sizes over time (see Table 2). Our analysis demonstrated increased fragility of the results from trials with <100 patients. Additionally, most trials (63%) failed to meet their a priori sample size calculations at the final follow-up (Supplemental Table C), and the number of patients lost to follow-up exceeded the Fragility Index in 75% of studies with significant findings. Taken together, these metrics indicate a risk of type I error in many trials that reported significant findings. Conversely, small trials are also at risk for type II error by failing to demonstrate a true difference in outcomes because of lack of power. Both errors are problematic in that they may affect the distribution of health-research resources and funding[19] and erode confidence in the efficacy of surgical procedures.[3] An opportunity exists to encourage multicenter collaboration within the orthopaedic community to produce higher-quality research in this regard. At present, orthopaedic surgery and sports medicine have lagged behind other medical disciplines in the percentage of collaborative, multicenter trials.[5,6,31] Although conducting larger, well-conducted trials may be time-consuming and expensive, the effort will increase the likelihood of producing meaningful and truthful results, with increased collaboration among institutions and appropriate planning.[18-20,32]

Limitations

Limitations of the present study include that the review did not consider trials published in other journals, limiting the generalizability of the results about the trends in the orthopaedic sports medicine literature to the global scientific community. However, *AJSM* has one of the highest impact factors among orthopaedic sports medicine journals and is likely to represent higher-quality orthopaedic trials. The quality of reporting of the included trials may have hindered the evaluation of the true methodological quality. Previous research has shown that few clinical trials adequately report on a number of statistical features, including the identification of primary or secondary analyses and providing or reporting sample size calculations.[24] Although certain criteria of the quality scores addressed this, further steps could be taken in the future to more comprehensively assess the adequacy of statistical reporting.[2]

CONCLUSION

The quantity and quality of published RCTs published in *AJSM* increased over the past 3 decades. Although these improvements are encouraging, single-center trials with small sample sizes (<100 patients) are still common (72.6% of studies) and produce fragile results. To limit bias and demonstrate the efficacy of orthopaedic treatments moving forward, there is a need to continue to conduct high-quality trials of appropriate sample size and rigorous design. This effort will undoubtedly demand an enhanced spirit of collaboration among the orthopaedic community.

Supplemental material for this article is available at https://journals.sagepub.com/doi/full/10.1177/23259671231161293#supplementary-materials.

REFERENCES

1. 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(2):199-207. doi:10.1097/SLA.0b013e31829c7795
2. Berger V, Alperson S. A general framework for the evaluation of clinical trial quality. *Rev Recent Clin Trials*. 2009;4(2):79-88. doi:10.2174/157488709788186021
3. Blom AW, Donovan RL, Beswick AD, Whitehouse MR, Kunutsor SK. Common elective orthopaedic procedures and their clinical effectiveness: umbrella review of level 1 evidence. *BMJ*. 2021;374(1):1511. doi:10.1136/BMJ.N1511
4. Brophy RH, Gardner MJ, Saleem O, Marx RG. An assessment of the methodological quality of research published in the *American Journal of Sports Medicine*. *Am J Sports Med*. 2005;33(12):1812-1815. doi:10.1177/0363546505278304

5. Brophy RH, Kluck D, Marx RG. Update on the methodological quality of research published in *The American Journal of Sports Medicine*. *Am J Sports Med*. 2016;44(5):1343-1348. doi:10.1177/0363546515591264

6. Brophy RH, Smith MV, Latterman C, et al. Multi-investigator collaboration in orthopaedic surgery research compared to other medical fields. *J Orthop Res*. 2012;30(10):1523-1528. doi:10.1002/jor.22125

7. Chess LE, Gagnier J. Risk of bias of randomized controlled trials published in orthopaedic journals. *BMC Med Res Methodol*. 2013;13(1).76. doi:10.1186/1471-2288-13-76

8. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46. doi:10.1177/001316446002000104

9. Cunningham BP, Harmsen S, Kweon C, et al. Have levels of evidence improved the quality of orthopaedic research? *Clin Orthop Relat Res*. 2013;471(11):3679-3686. doi:10.1007/s11999-013-3159-4

10. Cvetanovich GL, Fillingham YA, Harris JD, Erickson BJ, Verma NN, Bach BR. Publication and level of evidence trends in *The American Journal of Sports Medicine* from 1996 to 2011. *Am J Sports Med*. 2015;43(1):220-225. doi:10.1177/0363546514528790

11. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol*. 1992;45(3):255-265. doi:10.1016/0895-4356(92)90085-2

12. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J*. 2015;15(10):2188-2197. doi:10.1016/j.spinee.2015.06.004

13. Forrester LA, McCormick KL, Bonsignore-Opp L, et al. Statistical fragility of surgical clinical trials in orthopaedic trauma. *JAAOS Glob Res Rev*. 2021;5(11):e20.00197. doi:10.5435/JAAOSGLOBAL-D-20-00197

14. Grant HM, Tjoumakaris FP, Maltenfort MG, Freedman KB. Levels of evidence in the clinical sports medicine literature: are we getting better over time? *Am J Sports Med*. 2014;42(7):1738-1742. doi:10.1177/0363546514530863

15. 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(4):727-734. doi:10.1016/j.jhsa.2004.04.003

16. Harris JD, Erickson BJ, Abrams GD, et al. Methodologic quality of knee articular cartilage studies. *Arthroscopy*. 2013;29(7):1243-1252. e5. doi:10.1016/j.arthro.2013.02.023

17. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343(7829):D5928. doi:10.1136/bmj.d5928

18. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *J Am Med Assoc*. 2005;294(2):218-228. doi:10.1001/jama.294.2.218

19. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124

20. Katz JN, Wright JG, Losina E.Clinical trials in orthopaedics research. Part II. Prioritization for randomized controlled clinical trials. *J Bone Joint Surg Am*. 2011;93(7):e30. doi:10.2106/JBJS.J.01039

21. Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery: a systematic survey. *Am J Sports Med*. 2017;45(9):2164-2170. doi:10.1177/0363546516674469

22. Khan MS, Ochani RK, Shaikh A, et al. Fragility index in cardiovascular randomized controlled trials. *Circ Cardiovasc Qual Outcomes*. 2019;12(12):e005755. doi:10.1161/CIRCOUTCOMES.119.005755

23. Khormaee S, Choe J, Ruzbarsky JJ, et al. The fragility of statistically significant results in pediatric orthopaedic randomized controlled trials as quantified by the fragility index: a systematic review. *J Pediatr Orthop*. 2018;38(8):e418-e423. doi:10.1097/BPO.0000000000001201

24. Madden K, Arseneau E, Evaniew N, Smith CS, Thabane L. Reporting of planned statistical methods in published surgical randomised trial protocols: a protocol for a methodological systematic review. *BMJ Open*. 2016;6(6):e011188. doi:10.1136/bmjopen-2016-011188

25. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16(1):62-73. doi:10.1016/0197-2456(94)00031-W

26. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg Am*. 2005;87(12):2632-2638. doi:10.2106/JBJS.E.00370

27. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *Br Med J*. 1996;312(7023):71-72. doi:10.1136/bmj.312.7023.71

28. Saleh KJ, Bozic KJ, Graham DB, et al. Quality in orthopaedic surgery—an international perspective: AOA critical issues. *J Bone Joint Surg Am*. 2013;95(1):e3. doi:10.2106/JBJS.L.00093

29. 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(20):e116. doi:10.2106/JBJS.18.00653

30. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol*. 2014;67(6):622-628. doi:10.1016/j.jclinepi.2013.10.019

31. Wright JG, Gebhardt MC. Multicenter clinical trials in orthopaedics. *J Bone Joint Surg*. 2005;87(1):214-217. doi:10.2106/JBJS.D.02555

32. Wright JG, Katz JN, Losina E. Clinical trials in orthopaedics research. Part I. Cultural and practical barriers to randomized trials in orthopaedics. *J Bone Joint Surg Am*. 2011;93(5):e15. doi:10.2106/JBJS.J.00229

33. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85(1):1-2. doi:10.2106/00004623-200301000-00001

34. Zaidi R, Abbassian A, Cro S, et al. Levels of evidence in foot and ankle surgery literature: progress from 2000 to 2010? *J Bone Joint Surg*. 2012;94(15):e112. doi:10.2106/JBJS.K.01453

35. Zhang J, Chen X, Zhu Q, Cui J, Cao L, Su J. Methodological reporting quality of randomized controlled trials: a survey of seven core journals of orthopaedics from Mainland China over 5 years following the CONSORT statement. *Orthop Traumatol Surg Res*. 2016;102(7):933-938. doi:10.1016/j.otsr.2016.05.018