The devil is in the details: Reporting and transparent research practices in sports medicine and orthopedic clinical trials

Robert Schulz\(^1,2\), Georg Langen\(^3\), Robert Prill\(^4\), Michael Cassel\(^2\), Tracey Weissgerber\(^1\)

\(^1\) Berlin Institute of Health at Charité Universitätsmedizin, QUEST Center for Transforming Biomedical Research, Berlin, Germany

\(^2\) University of Potsdam, Department of Sport and Health Sciences, Potsdam, Germany

\(^3\) Institute for Applied Training Science, Leipzig, Germany

\(^4\) Brandenburg Medical School Theodor Fontane, Department of Orthopedics and Traumatology, Brandenburg, Germany

**Corresponding author:**

Tracey Weissgerber, PhD

Berlin Institute of Health at Charité (BIH)

BIH QUEST Center for Transforming Biomedical Research

Anna-Louisa-Karsch-Str. 2

10178 Berlin

[tracey.weissgerber@charite.de](mailto:tracey.weissgerber@charite.de)

**Word count:** 3694
Abstract

Introduction: While transparent reporting of clinical trials is essential to assess the risk of bias and translate research findings into clinical practice, earlier studies have shown that deficiencies are common. This study examined current clinical trial reporting and transparent research practices in sports medicine and orthopedics.

Methods: The sample included clinical trials published in the top 25% of sports medicine and orthopedics journals over eight months. Two independent reviewers assessed pre-registration, open data and criteria related to scientific rigor, the study sample, and data analysis.

Results: The sample included 163 clinical trials from 27 journals. While the majority of trials mentioned rigor criteria, essential details were often missing. Sixty percent (confidence interval [CI] 53-68%) of trials reported sample size calculations, but only 32% (CI 25-39%) justified the expected effect size. Few trials indicated the blinding status of all main stakeholders (4%; CI 1-7%). Only 18% (CI 12-24%) included information on randomization type, method, and concealed allocation. Most trials reported participants’ sex/gender (95%; CI 92-98%) and information on inclusion and exclusion criteria (78%; CI 72-84%). Only 20% (CI 14-26%) of trials were pre-registered. No trials deposited data in open repositories.

Conclusions: These results will aid the sports medicine and orthopedics community in developing tailored interventions to improve reporting. While authors typically mention blinding, randomization and other factors, essential details are often missing. Greater acceptance of open science practices, like pre-registration and open data, is needed. These practices have been widely encouraged, we discuss systemic interventions that may improve clinical trial reporting.

Registration: https://doi.org/10.17605/OSF.IO/9648H
Introduction

The overarching goal of medical research is to improve healthcare for patients, which requires the biomedical community to translate study outcomes into clinical practice (1). Clinical trials are central to this process, as properly conducted trials reduce the risk of bias and increase the likelihood that results about new treatments will be trustworthy, reproducible and generalizable (2,3). Clinical trials must be properly designed, conducted, and reported (4) to facilitate translation. Poorly designed and conducted trials may not be trustworthy or reproducible. This undermines public trust in biomedical research and raises concerns about whether the trial costs and patient risks were justified (5,6). Poor reporting makes it difficult to distinguish between trials with and without a high risk of bias.

To improve clinical trial reporting, the Consolidated Standards of Reporting Trials (CONSORT) guidelines (7,8) have been recommended by the ICMJE and widely disseminated by the EQUATOR network (9,10). While reporting has improved over time, major deficiencies that can impair translation are still common (11,12). Details needed to assess the risk of bias were missing from many published trials. More than half of all trials failed to address allocation concealment, and almost one third of studies did not address blinding of participants and personnel (12). Similarly, among randomized controlled trials published in the top five orthopedics journals, 60% failed to address the blinding status of the participants and 58% did not specify the number of participants included in the final analysis (13).

Orthopedics and sports medicine researchers have joined efforts to improve study design and reporting. Newly formed societies (14,15) and editorial series (16) focus on improving research quality in sports medicine and orthopedics. These efforts are urgently needed, as only 1% of studies in high-impact orthopedic journals reported all ten criteria needed for risk of bias assessment (13). In 42% of papers, risk of bias could not be assessed due to incomplete reporting (13). Incomplete reporting of exercise interventions (17) makes it impossible to
implement interventions in clinical practice or to assess the appropriateness of the control intervention (18).

In sports medicine related fields, meta-researchers suggested that scientists may be using questionable research practices (Table 1) after observing overinflated effect sizes (19) and an unreasonably high number of papers that support the study hypothesis (20). Comprehensive reporting may prevent these biases or make them easier to detect. However, earlier studies have shown that reporting deficiencies are still common in orthopedics (13) and general medical journals (12,21).

Therefore, this meta-research study examined reporting among clinical trials published in the top 25% of sports medicine and orthopedics journals. Our objective was to assess the prevalence of reporting for selected criteria, including pre-registration, open data and reporting of randomization, blinding, sample size calculations, data analysis and the flow of participants through the study. Meta-research data on clinical trial design, conduct and reporting will help researchers in sports medicine to implement targeted measures to improve trial design and reporting.
### Table 1 Terminology and concepts

<table>
<thead>
<tr>
<th>Concept</th>
<th>Questionable research practices are defined as “Design, analytic, or reporting practices that have been questioned because of the potential for the practice to be employed with the purpose of presenting biased evidence in favor of an assertion” (22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective reporting/ cherry picking</td>
<td>The decision about whether to publish a study or parts of a study is based on the direction or statistical significance of the results (23,24). Pre-registration and Registered Reports may prevent selective reporting (25,26), which is also known as cherry picking.</td>
</tr>
<tr>
<td>Publication bias</td>
<td>The decision about whether to publish research findings depends on the strength and direction of the findings (27). The odds of publication are nearly four times higher among clinical trials with positive findings, compared to trials with negative or null findings (28).</td>
</tr>
<tr>
<td>Outcome reporting bias</td>
<td>Only particular outcome variables are included in publications and decisions about which variables to include are based on the statistical significance or direction of the results (23). Outcomes that are statistically significant have higher odds of being fully reported than non-significant outcomes (29,30).</td>
</tr>
</tbody>
</table>
| Attrition bias                  | Attrition refers to reductions in the number of participants throughout the study due to withdrawals, dropouts, or protocol deviations. Attrition bias occurs when there are systematic differences between people who leave the study and those who continue (31).  
  For example, a trial shows no differences between two treatments. In one group, however, half the participants dropped out because they underwent surgery due to worsening symptoms. |
| Null hypothesis statistical testing (NHST) | NHST is originally based on theories of Fischer and Neyman-Pearson. The null hypothesis is rejected or accepted depending on the position of an observed value in a test distribution. While NHST is standard practice in many fields, the International Committee of Medical Journal Editors warns against the sole reliance on NHST due to several shortcomings of this approach (32). |
| p-Hacking                       | Describes the process of analyzing the data in multiple ways until statistically significant results are found. |
| HARKing                         | HARKing, or hypothesizing after results are known, is defined as presenting a post hoc hypothesis as if it were an a priori hypothesis (33). |
Methods

Protocol Pre-registration

The study was pre-registered on the Open Science Framework (RRID:SCR_003238) at [https://doi.org/10.17605/OSF.IO/9648H](https://doi.org/10.17605/OSF.IO/9648H). Additional details regarding sample selection and screening, data abstraction, and sample size calculation can be found in the supplemental materials.

Sample selection and screening

We systematically examined clinical trials published in the top 25% of orthopedics and sports medicine journals over eight months. This sampling strategy provides a broad overview of practices in the field while including high-impact journals, which have the potential to drive change. Journals in the orthopedics and sports medicine category were selected based on the Scimago Journal Rank indicator (34) (supplementary methods). The top 25% of journals (n=65) were entered into the PubMed search with article type (clinical trial) and publication date (2019/12:2020/08) filters. The search was run on September 16, 2020. All articles (n=175 from 27 journals) were uploaded into Rayyan (RRID:SCR_017584; 35) for screening. Two reviewers (RS, GL) screened titles and abstracts to exclude articles that were obviously not clinical trials, as defined by the ICMJE. The ICMJE defines a clinical trial as any research project that “prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the relationship between a health-related intervention and a health outcome”(9). Two independent reviewers (RS, GL, RP) then performed full-text screening. All papers meeting the ICMJE clinical trial definition were included, whereas articles that did not meet the definition were excluded. Studies looking at both health-related and non-health-related outcomes were included but data abstraction focused on health-related outcomes only. Disagreements were resolved by consensus.
Data abstraction

Two independent assessors (RS, GL, RP) reviewed each article and its supplemental files to evaluate the reporting of pre-specified criteria and extracted data using preformatted Excel spreadsheets. Table 2 presents the main criteria that were abstracted and a reason for their selection. The transparency and rigor criteria are based on CONSORT criteria for methods and results reporting (7,8). We also abstracted additional open science criteria, focusing on open access and open data (36,37). The abstraction protocol was deposited on the Open Science Framework (RRID:SCR_003238) at https://osf.io/q8b46/.

Protocol Deviations

For trials with exercise interventions, we assessed the frequency, intensity, and volume of exercise for experimental and control interventions. The protocol was modified if the control intervention did not involve exercise. Control interventions were rated as fully reported if the frequency, the content, and the duration was described. Control groups that received no intervention (e.g. wait-and-see) were rated as fully reported if the activity status or number of other treatments were monitored.

Trial registration statement assessments were amended to determine whether trials were registered prospectively or retrospectively. Two abstractors (RS, MP) assessed each trial registration. Trials were considered pre-registered if their registration was completed before the first participant was enrolled. Otherwise, the trial was classified as retrospectively registered. If the primary outcome was changed after the study began, the trial was classified as retrospectively registered.

Statistical Analysis

This exploratory study assessed the prevalence of reporting for selected criteria in sports medicine and orthopedics clinical trials. Results are presented as the percentage of trials reporting each outcome measure, with a 95% confidence interval.
Odds ratios and their 95% confidence intervals were calculated to examine the relationship between the completeness of reporting and pre-registration, the use of flow charts, or the presence of sample size calculations and the completeness of reporting. Odds ratios were interpreted as unclear if the confidence interval included 1. These analyses were not pre-registered.

**Sample Size Calculation**

This exploratory study does not require formal sample size calculations. However, we adhered to conventional sample size recommendations for exploratory designs and performed a precision-based sample size calculation to obtain rough estimates of relevant sample sizes (supplemental methods). Depending on different assumptions, a required sample size of 124 to 203 trials was estimated.
Table 2 Criteria for reporting and transparent research practices. The table shows specific questions used to assess each outcome criteria and provides a brief justification for why each criteria was selected.

<table>
<thead>
<tr>
<th>Category</th>
<th>Assessment</th>
<th>Rationale and Context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample Size calculation</strong></td>
<td>Was an a priori sample size calculation performed?</td>
<td>- Low power is associated with high rates of spurious findings and overinflated effect sizes (38), and there is evidence for low median statistical power in rehabilitation research [40].</td>
</tr>
<tr>
<td></td>
<td>What type of sample size calculation was performed?</td>
<td>- A priori sample size calculations help to prevent underpowered trials, however, they are regularly performed inadequately. Common problems include failing to justify the expected treatment effect and not stating all values required for calculation (39). The majority of sample size calculations in rehabilitation trials are missing expected effect sizes (40).</td>
</tr>
<tr>
<td></td>
<td>Did the authors provide a justification for the expected effect size?</td>
<td>- Low power is associated with high rates of spurious findings and overinflated effect sizes (38), and there is evidence for low median statistical power in rehabilitation research [40].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- A priori sample size calculations help to prevent underpowered trials, however, they are regularly performed inadequately. Common problems include failing to justify the expected treatment effect and not stating all values required for calculation (39). The majority of sample size calculations in rehabilitation trials are missing expected effect sizes (40).</td>
</tr>
<tr>
<td><strong>Randomization &amp; concealed allocation</strong></td>
<td>Did the authors address whether randomization was used? If so, were the randomization type and method mentioned? Were the following details of the allocation concealment procedure addressed? - Who generated the randomization sequence? - Who enrolled participants? - Who assigned participants to groups?</td>
<td>- Inadequate randomization and allocation concealment procedures introduce selection bias and are associated with increased odds of significant but spurious results (41) and overestimated treatment effects (42).</td>
</tr>
<tr>
<td><strong>Blinding</strong></td>
<td>Did the article include a statement on blinding?</td>
<td>- Blinding prevents ascertainment bias in clinical trials. A lack of blinding is associated with overinflated effect sizes (43). Terms like double-blind are ambiguous, interpreted differently, and don’t provide reliable information on blinding of specific stakeholder groups (44). These terms should be abandoned in favor of reporting the blinding status of all relevant stakeholders (8).</td>
</tr>
<tr>
<td></td>
<td>Was the blinding status of each of the major stakeholders mentioned (participants, healthcare providers, outcome assessors, data analysts)?</td>
<td>- Blinding prevents ascertainment bias in clinical trials. A lack of blinding is associated with overinflated effect sizes (43). Terms like double-blind are ambiguous, interpreted differently, and don’t provide reliable information on blinding of specific stakeholder groups (44). These terms should be abandoned in favor of reporting the blinding status of all relevant stakeholders (8).</td>
</tr>
<tr>
<td></td>
<td>Was each stakeholder group blinded?</td>
<td>- Blinding prevents ascertainment bias in clinical trials. A lack of blinding is associated with overinflated effect sizes (43). Terms like double-blind are ambiguous, interpreted differently, and don’t provide reliable information on blinding of specific stakeholder groups (44). These terms should be abandoned in favor of reporting the blinding status of all relevant stakeholders (8).</td>
</tr>
<tr>
<td><strong>Flow of participants</strong></td>
<td>Were the inclusion and exclusion criteria clearly stated? Did the authors define how many participants were excluded at each phase of the study and list reasons for exclusion? Did the authors present this information in a flow chart?</td>
<td>- Detailed inclusion and exclusion criteria help the reader to assess generalizability. - Knowing when and why participants dropped out or were removed from the study is essential to estimate attrition bias.</td>
</tr>
<tr>
<td><strong>Data analysis</strong></td>
<td>Was a study hypothesis presented and a primary outcome specified?</td>
<td>- Specifying the study hypothesis and the primary outcome prospectively safeguards against selective reporting. Discrepancies between the registration and the study report may indicate outcome switching, which favors statistically</td>
</tr>
<tr>
<td></td>
<td>Was the hypothesis supported or rejected?</td>
<td>- Specifying the study hypothesis and the primary outcome prospectively safeguards against selective reporting. Discrepancies between the registration and the study report may indicate outcome switching, which favors statistically</td>
</tr>
</tbody>
</table>
If NHST was performed, were exact p-values, degrees of freedom, and the test statistics presented?
Were standardized effect sizes and their precision reported?

- Reporting the test statistic and degrees of freedom allows readers to identify misreported p-values. In 13% of psychology studies, meta-researchers detected mismatches between p-values and the reported test statistic and degrees of freedom that would affect statistical conclusions (47).
- Analyses should take the magnitude, confidence, and likelihood of an effect into account, instead of focusing on whether effects are statistically significant. Effect sizes show the magnitude of effects within a study, while standardized effect sizes allow for comparisons across studies (48).

Data visualization

- Were bar graphs used to visualize continuous data?
- Using bar graphs to visualize continuous data is problematic because many different data distributions can lead to the same bar graph. The actual data may suggest different conclusions from the summary statistics alone (49,50).

Intervention reporting

- What type of intervention was performed (e.g. exercise, physical therapy, surgery)?
- For exercise interventions:
  - Was monitoring of adherence to the intervention addressed?
  - Were essential details needed to replicate the experimental and control interventions (e.g. frequency, intensity, volume, and type of exercise) provided?
- When clinical trials do not report details needed to implement the intervention, findings cannot be translated into clinical practice. The minority of exercise studies provided enough information to allow others to replicate interventions (51). The high prevalence of insufficient reporting led to the establishment of new intervention reporting guidelines (52,53).
- Adherence can affect intervention efficacy. Intervention effects can be up to three times larger in fully adherent participants compared to partly adherent participants (54).

Transparency criteria

- Was the study registered or pre-registered?
- Was a data availability statement included? Were the data publicly available?
- Was the study openly accessible?
- Half of researchers admit to selectively reporting results and presenting post-hoc analyses as if they had been pre-specified (22). Pre-registration protects against this. Pre-registration (since 2005) and data availability statements (since 2018) are mandatory for clinical trials (55).
- Open access papers generate more media coverage and citations (56).
- Open data facilitates collaboration and benefits society (56). In 2017, 21% of 316 biomedical journals (57) and 28% of funders (58) required open data.
Results

175 articles were screened, and 168 articles were reviewed from 27 sports medicine and orthopedics journals (Figure S1, Table S1). Eleven articles were excluded because they did not meet the ICMJE clinical trial criteria. One extended conference abstract was excluded because it was not a full-length research article. Analyses included the remaining 163 papers.
Rigor and Sample Criteria

**Figure 1** Reporting prevalence for rigor and sample criteria. This plot displays the percentage of trials that addressed each criteria. For information on the actual randomization or blinding status, please refer to the text. The different colored data points are for better visual differentiation of each subcategory.

**Sample size calculations**: In trials not reporting a priori sample size calculation (Figure 1), 2% (CI 0-5%; n=4) reported that no sample size calculation was performed because the study was an exploratory pilot study. Among trials reporting sample size calculations (n=98), 53% (CI 43-
63%; n=52) included a justification for the expected effect size. The remaining trials either presented no justification (39%; CI 23-42; n=32) or used arbitrary effect size thresholds (14%; CI 7-21%; n=14). Almost all sample size calculations were based on statistical power (93%; CI 88-98%; n=96). Two sample size calculations were based on precision (2%; CI 0-5%). No calculations were based on Bayes methods.

**Randomization and allocation concealment**: In trials not addressing randomization (Figure 1), two trials (1%; CI 0-3%) were not randomized, and five trials did not mention randomization (3%; CI 0-6%).

Eight percent (CI 4-12%; n=13) of trials provided complete information on the allocation concealment procedure (defined as reporting who generated the randomization sequence, and who enrolled participants and assigned them to interventions). Some of this information was available 23% (CI 16-29%; n=37) of trials, and 69% (CI 62-76%; n=113) did not report any information. Few studies reported at least some information on all three factors needed to assess randomization and allocation concealment (randomization type, method, and allocation concealment; 18%; CI 12-24%; n=30).

**Blinding**: Two-thirds of trials addressed blinding (Figure 2). Among trials that addressed blinding (Figure 1), 81% (CI 73-88%; n=84) used blinding, while 19% (CI 12-27%; n=20) were not blinded. Only 4% (CI 1-7%; n=7) of all trials addressed the blinding status of all four stakeholder groups (Figure 2). Trials were most likely to address the blinding status of the outcome assessors and the participants. The blinding status of data analysts is typically unreported.
Figure 2: The blinding status across the main different stakeholder groups across all clinical trials (n=163)
Sample-related Criteria

Approximately three-quarters of trials reported the inclusion and exclusion criteria and provided complete information on the number of participants at enrollment, after enrollment, and included in data analysis (Figure 1). Fewer trials used a flow chart to illustrate the number of included and excluded participants at each stage. Among trials that did not report the reasons for all exclusions after enrollment (Figure 1), 17% (CI 11-22%; n=24/90) reported the reasons for some exclusions and 33% (CI 26-41%; n=41/90) did not report any information.

In trials that stated participants’ sex or gender (Figure 1), a median of 51% (interquartile range (IQR) 27-71%) of participants were women in the group with the highest proportion of women, vs. 49% (IQR 22-66%) in the group with the lowest proportion of women.

Intervention Criteria

The most frequent intervention type was exercise (44%; CI 37-52%; n=72), followed by surgery (26%; 19-32%; n=42). Diet (6%; CI 2-9%; n=9), physical therapy (5%; CI 2-8%; n=8), pharmacological interventions (4%; CI 0-2%; n=7) and manual therapy (1%; CI 0-2%; n=1) were uncommon. Fifteen percent (CI 9-20% n=24) of studies used other interventions.

We next examined reporting of details needed to assess or implement exercise interventions. Sixty-two percent (CI 50-73%; n=42) of trials with exercise interventions monitored adherence or compliance, one trial (1%; CI 0-4%) reported that adherence was not monitored, and 37% (CI 25-48%; n=25) of trials did not mention intervention adherence or compliance. All trials reported at least some information about the experimental exercise intervention, and most trials provided complete information (Table 2) (83%; CI 75-92%; n=60). Fewer trials reported complete information for the control interventions (63%; CI 51-74%; n=45). Five trials did not provide any information about the control intervention (7%; CI 1-13%).
Data analysis and Transparency Criteria

![Figure 3](image)

**Figure 3** Reporting prevalence for data analysis and transparency criteria. This plot displays the percentage of trials that addressed each criteria. Abbreviations: NHST, null hypothesis statistical testing; ES, effect size

**Hypotheses and outcome measures**: Nearly half of the articles specified a primary outcome and almost two-thirds of articles presented a hypothesis (Figure 3). Among clinical trials that reported a hypothesis (Figure 3), 61% (CI 53-68%; n=62) supported the main hypothesis, while 39% of trials (CI 32-47; n=40) did not support the main hypothesis.

**Statistical Reporting**: Almost all studies used NHST (Figure 3). While most trials reported exact p-values, few reported test statistics and degrees of freedom. Approximately half of the trials
reported standardized effect sizes but only 21% included the precision of the effect size estimates. One study reported Bayesian statistics (1%; CI 0-2%).

**Data visualization:** Bar graphs were used to display continuous data in 21% (CI 15-21%; n=34) of trials. These graphs should be replaced with more informative graphics (e.g. dot plots, box plots or violin plots) that show the data distribution(49,50).

**Transparency**

Most of the studies with registration statements (Figure 3) were registered in ClinicalTrials.gov (n=52), followed by the Australian New Zealand Clinical Trials Registry (n=9), International Standard Randomized Controlled Trial Number Register (n=4), and other regional clinical trials registries (n=9). Less than half of the registered trials, and 20% of all trials, were pre-registered. The remaining trials with registration statements were registered retrospectively (58%; CI 48-69%; n=49/84). This included six prospectively registered trials where the primary outcome was changed after data collection started. Two studies with registration statements did not provide sufficient information to determine whether the study was registered prospectively or retrospectively (2%; CI 0-6%; n=2/84).

Data availability statements were uncommon (Figure 3). No trial with a data availability statement deposited data publically in an open repository. Twenty-one percent of trials with data availability statements (15-27%; n=4) noted that data were not publicly available, whereas 74% (67-80%; n=15) stated that data were available upon request. One study (5%; CI 2-9%) reported that all data were available in the main text and its supplements, however, raw data was not available in either location.
Exploratory analyses

Pre-registration and reporting: Compared to unregistered or retrospectively registered studies, pre-registered studies were more likely to report complete information for randomization (type and method) and allocation concealment (OR 4.3; CI 1.9-10.0), whether all stakeholders were blinded (OR 8.6; CI 1.6-46.5), a priori sample size calculations (OR 2.5; CI 1.1-5.8), justifications for expected effect sizes used in power calculations (OR 2.5; CI 1.1-5.8), and specifying the primary outcome measure (OR 3.3; CI 1.5-7.1). The odds of reporting (OR 1.0; CI 0.48-2.1) or rejecting (OR 1.0; CI 0.42-2.6) the study hypothesis were not clearly different between unregistered and pre-registered studies.

Sample size calculations and reporting: The odds of rejecting the main hypothesis in trials with a priori sample calculations were unclear (OR 1.3; CI 0.6-2.8). Trials that provided justifications for the expected effect size were more likely to reject the study hypothesis (OR 2.5; CI 1.2-5.2).

Flow charts and reporting: The odds of reporting all reasons for dropouts (OR 4.6; CI 2.3-9.3) and explicitly reporting the number of participants in each group that were included in the data analysis (OR 163.3; CI 21.4-1248.5) were higher among studies that used flow charts to track participant flow, compared to those that did not.
Discussion

Sports medicine and orthopedics researchers have recently emphasized rigorous study design and reporting to make research easier to understand, interpret, and translate into clinical practice (16). Calls for more transparent reporting in orthopedics and sports medicine (19,26,59) followed older studies suggesting that poor clinical trial reporting limits readers ability to assess study quality and risk of bias (13,60,61). Our study shows that while most studies include a general statement about rigor criteria, like blinding or randomization, these statements lack essential details needed to assess the risk of bias. The majority of trials report criteria related to the study sample, such as the sex of participants, inclusion and exclusion criteria, or the number of participants finally included in the analysis. Only 20% of studies were pre-registered. No study shared data in open repositories.

Opportunities to improve reporting

These results highlight two main opportunities to improve transparency and reproducibility in sports medicine and orthopedics clinical trials; improving reporting for essential details of the main CONSORT elements and increasing uptake of open science practices.

First, our results indicate that most authors are aware that they need to address factors like blinding, randomization and sample size calculations; however, few provide the essential details required to evaluate the trial and interpret the results. Almost all trials addressed blinding, for example, but only 4% reported the blinding status of all main stakeholders. Educational efforts should emphasize the difference between informative and uninformative reporting (see example in Figure 4). CONSORT writing templates may also help (60). Target criteria should include the blinding status of all main stakeholders, randomization type and method, how and by whom concealed allocation was performed, and effect size justifications in sample size calculations.
How to conduct & report sample size calculations

<table>
<thead>
<tr>
<th>What is the problem?</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% of studies use Null Hypothesis Significance Testing (NHST)</td>
</tr>
<tr>
<td>Only 60% provide a sample size calculation</td>
</tr>
<tr>
<td>32% justify the expected effect size</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Why is this a problem?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• When using NHST, sample size calculations are performed to inform the reader about the expected risk of false findings.</td>
</tr>
<tr>
<td>• If the authors make inaccurate a-priori assumptions about the expected effect size, the reported risk of false findings will be incorrect.</td>
</tr>
<tr>
<td>• Authors should justify the expected effect size. This allows the reader to assess whether the calculation is realistic and evaluate how much confidence can be placed in the results</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What is not sufficient?</th>
</tr>
</thead>
<tbody>
<tr>
<td>“A sample size of 21 subjects per group was needed to detect an effect size of 0.5 with 80% power and an alpha of 0.05.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What should be reported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The primary outcome measures for which the calculation is performed</td>
</tr>
<tr>
<td>The desired power (Type II error rate) and accepted Type I error rate (α)</td>
</tr>
<tr>
<td>The size of the expected effect</td>
</tr>
<tr>
<td>The justification for the expected effect size with reference (e.g. meta-analysis, single study, smallest meaningful effect)</td>
</tr>
</tbody>
</table>

A good example

We performed an a priori sample size calculation for our primary outcome 0, comparing intervention and placebo.

We set α at 0.05 and the statistical power at 0.8. The estimated effect size of Cohen’s d = 0.5 is based on a RCT from (X et al.) which found a mean difference of X₁ (SD=X₁) in outcome 0 between the intervention and placebo groups.

Therefore, the required number of participants per group is 21.

Figure 4 A priori sample size calculations are essential for generating meaningful results with clinical trials
Second, interventions are needed to increase pre-registration and data sharing. Although the ICJME has required clinical trial pre-registration since 2005 (61), only one-fifth of trials were pre-registered. Pre-registered studies had higher odds of reporting several rigor criteria, potentially suggesting that authors who preregister may be more aware of reporting guidelines. Our results are consistent with previous findings (62) that trial registrations were among the least reported CONSORT items in sports medicine. Sports medicine researchers have already noted that pre-registration and registered reports can prevent questionable research practice (26) (Table 1) or make them easier to detect (63).

Data were not shared in public repositories, suggesting that this topic requires special attention. The benefits of data sharing for authors include more citations (64,65), and increased opportunities to collaborate with researchers who want to perform secondary analyses (66). Recent materials have addressed many common concerns about sharing patient data, including data privacy and confidentiality (67–69). Regulations vary by country and institution. Some institutions have designated support staff for data sharing. Researchers should contact their institutions’ data privacy, statistics, or ethics offices to identify local experts. Seventy-four percent of trials with data availability statements noted that data were available on request. This is problematic, as such data are often unavailable and the odds of obtaining data decline precipitously with time since publication (70).

Interestingly, our exploratory analysis revealed that the odds of rejecting the study hypothesis were 2.5 (CI 1.2-5.2) times higher in trials that provided a justification for the expected effect size in sample size calculations. This might indicate overinflated effect sizes, as trials that based their sample size calculation on effect sizes published in
earlier studies more often failed to find a similar sized effect. Inflated effect sizes were also observed in the psychological science reproducibility project, where replicated effects were generally smaller than those in the initial studies (71).

Authors should also be encouraged to report the data analysis transparently. Our study shows that reporting of test statistics and degrees of freedom yields much potential for improvement, as well as reporting of standardized effect sizes and their precision. Focusing on the magnitude and precision of differences, instead making decisions based on p-value thresholds, reduces the likelihood of spurious findings (72,73). Twenty-five to 38% of medical articles (74), and up to 50% in psychology papers (47), contain p-values that don’t match the reported test-statistic and degrees of freedom. These inaccurate p-values may alter study conclusions in 13% of psychology papers (47). Our study shows that these assessments are impossible in sports medicine and orthopedics clinical trials, as test statistics and degrees of freedom are rarely reported.

Reporting of criteria related to the study sample and to exercise interventions highlighted some positive points. Whereas Costello et al. (75) observed that less than 40% of sports and exercise study participants were females, indicating sex bias, our study, on average, shows an even distribution of sex/gender. Similarly the number of participants included in the analysis was reported in 75% of trials in the present study, compared to 42% of randomized controlled trials in orthopedic journals (13). The introduction of flow charts to display the participant flow in CONSORT 2010 may improve reporting for sample related criteria, as trials which included flow charts were more likely to report the number of participants included in the analysis and reasons for all exclusions. While the majority of studies reported key details of exercise interventions, reporting was less
comprehensive for the control intervention and for intervention adherence or compliance.

**Options for systemic interventions to improve reporting**

Ongoing reporting deficiencies in clinical trials highlight the need for systemic interventions to improve reporting. The 2010 CONSORT guideline has been endorsed by more than 50% of the core medical journals and the ICMJE (76). Persistent reporting deficiencies (12,21) indicate that endorsement without enforcement is insufficient (77,78), and engaging individuals, journals, funders, and institutions is necessary to improve reporting (79,80).

One option to improve reporting is for journals to enforce existing guidelines and policies. All journals in our sample were peer reviewed; yet there were major essential details were often missing from published trials. This suggests that peer review alone is insufficient. Alternatives include rigorous manual review by trained “trial reporting” assessors, automated screening or a combined approach. A journal program that trained early career researchers to check for common data visualization errors was well accepted by authors and increased compliance with data presentation guidelines (81). Implementing similar programs, using paid staff, could improve CONSORT compliance. Alternatively, automated screening tools may efficiently flag missing information for peer reviewers (82,83). Peer review systems at several journals include an automated tool that checks statistical reporting and guideline adherence (84). Tools are available to screen for risk of bias (RobotReviewer;RRID:SCR_021064 (85)), and CONSORT methodology criteria (CONSORT-TM;RRID:SCR_021051 (86)). The CONSORT tool performs well for frequently reported criteria, but needs more training data for less often reported criteria (86). New tools may be need to be created to assess details like the specifics of allocation concealment, blinding of specific stakeholders or justifications of expected effect sizes. As 52% of clinical trials in our sample were published in only five journals, systemic efforts to improve reporting on journal level can make a notable difference on clinical trial reporting in the field.
A second option is automated screening of sports medicine and orthopedics preprints. Preprints, which are posted on public servers such as medRxiv and sportRxiv prior to peer review, allow authors to receive feedback and improve their manuscripts before journal submission. Large-scale automated screening of bioRxiv and medRxiv preprints for rigor and transparency criteria is feasible and could raise awareness about factors affecting transparency and reproducibility (87). Automated screening has limitations – the tools make mistakes and cannot always determine whether a particular item is relevant to a given study. Automated screening may complement peer review, but is not a replacement. The value of this approach will also depend on the proportion of trials that are posted as preprints.

Dashboards may offer a third option for improving reporting. Dashboards allow researchers to monitor changes over time and may incentivize transparent practices. Examples include dashboards on open science (88), and trial results reporting (89). In sports medicine and orthopedics, clinical trial dashboards could track transparent research practices for journals, society publishers, or all publications, and should include commonly missed items identified in this study. Researchers may need to develop new automated tools to track some criteria.

The scientific community has long relied on educational resources to improve reporting. On-demand resources include the CONSORT guideline use webinar by Altman (90), and open webinars on pre-registration, sample size justification and other topics offered by the Society for Transparency, Openness, and Replication in Kinesiology (91). Creating a single platform with field-relevant resources; then collaborating with large journals, publishers, and societies, may help to disseminate materials to the global orthopedics and sports medicine community.

Limitations

Our CONSORT-based evaluation criteria for intervention reporting were not optimized for non-exercise or wait-and-see control interventions. While the assessments required by guidelines for intervention reporting (52,53) were beyond the scope of this study, previous studies assessed
intervention reporting in detail (17,51,54,92). Larger, confirmatory studies are needed to examine relationships between different variables, as odds ratios calculated in the present study were exploratory post-hoc calculations. We examined the top 25% sports medicine and orthopedics journals; hence our findings may not be generalizable to journals that are not indexed by PubMed, lower tier journals, or non-English journals.

Conclusions

Transparent reporting of clinical trials is essential to assess the risk of bias and translate research findings into clinical practice. Despite some improvements over time, older studies and studies in other fields show persisting deficiencies in clinical trial reporting. The present study in recent sports medicine and orthopedics clinical trials shows that authors often report general information on rigor criteria but few provide the essential details to assess risk of bias required by existing guidelines. Examples include the blinding status of all main stakeholders, information on the concealed assignment, or the justification of expected effect sizes in sample size calculations. Further, transparent research practices like pre-registration or data sharing are rarely used in sports medicine and orthopedics.

As reporting guidelines for clinical trial reporting are long established and well accepted across medical fields, the persisting lack of detailed reporting suggests that further interventions and different approaches are needed to improve clinical trial reporting further. We present different options for future interventions might investigate rigorous peer-reviewer training, automated screening of submitted manuscripts and preprints, and field-specific dashboards to monitor reporting and transparent research practices to increase awareness and track improvements over time. Our results show which aspects of clinical trial reporting have the greatest need for improvement. Researchers can use this data to tailor future interventions to improve reporting to the needs of the sports medicine and orthopedics community.
Data availability Statement

All data are available on the OSF and may be accessed under the Creative Commons Attribution 4.0 International License at the following link: https://osf.io/q8b46/

Acknowledgements

We would like to acknowledge Mia Pattillo for her valuable contributions to extracting the registration status of trials.

Competing Interests

All authors declare no competing interests.
References


84. BMC. Advancing peer review at BMC [Internet]. 2021 [updated 2021 Apr 8; cited 2021 Apr 8]. Available from: https://www.biomedcentral.com/about/advancing-peer-review


89. EU Trials Tracker. EU Trials Tracker — Who's not sharing clinical trial results? [Internet]: Evidence-Based Medicine Data Lab; University of Oxford. 2021 [updated 2021 Apr 8; cited 2021 Apr 8]. Available from: https://eu.trialstracker.net/


