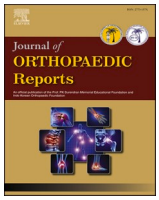
JOREP 4 (2025) 100573

Contents lists available at ScienceDirect

Journal of Orthopaedic Reports

journal homepage: www.journals.elsevier.com/journal-of-orthopaedic-reports



Biostatistics in orthopedic surgery: Common data pitfalls and how to avoid them

A B S T R A C T

Background: Biostatistical expertise in orthopaedic research is critical as much of the research typically involves multiple research sites, complex datasets and collection strategies, and multifaceted hypotheses. However, in much of the orthopaedic literature, studies use inconsistent data collection methods without proper data auditing or validation, improper statistical methods, and overstate p-values and hypothesis testing results.

Study Objectives & Rationale: In this short communication, we highlight biostatistical considerations in research design, common pitfalls in data collection and analysis, and strategies to address these challenges.

Conclusion: Improper biostatistical methods result in biased or misleading conclusions which may adversely affect clinical practice and patient outcomes.

1. Introduction

Biostatistics is a cornerstone of orthopedic research, offering tools to design rigorous studies, analyze complex data, and generate reliable findings. In orthopedic research, where studies often involve complex datasets and multifaceted hypotheses, biostatistics ensures that conclusions are reliable, reproducible, and clinically meaningful applications of statistical methods. These are for evidence-based research, influencing both clinical decision-making and patient outcomes.

Despite its importance, common statistical pitfalls can compromise research outcomes. Challenges such as inconsistent data collection, improper statistical methods, and overreliance on p-values can lead to biased or misleading conclusions. Additionally, running multiple comparisons can inflate the likelihood of false-positive results, further compromising the validity of research. For instance, a previous study found that only 6%–15% of articles in orthopedic journals used appropriate correction methods, resulting in an estimated 54% risk of at least one false significant result in unadjusted studies.[1,2] Addressing these pitfalls with careful planning and attention to detail is critical for ensuring the validity and reliability of orthopedic studies.

This manuscript highlights statistical considerations in research design, common pitfalls in data collection and analysis, and strategies to address these challenges. By integrating statistical practices into orthopedic research, investigators can enhance the quality and impact of their studies, fostering advancements in both science and patient care.

2. Methods

2.1. Statistical considerations in research design

A well-structured research design lays the foundation for successful studies. Proper planning minimizes biases, ensures data quality, and aligns outcomes with study objectives.[2] This section outlines key considerations for experimental design, statistical design, and research feasibility.

2.2. Experimental design

The first step in designing a study is formulating a clear, testable hypothesis. For example, a study evaluating knee osteochondral allograft transplantation (OCAT) may hypothesize, "VAS pain scores will improve by at least 20% at six months post-surgery compared to baseline." Measurements must align with these hypotheses, capturing primary outcomes such as pain reduction and secondary outcomes like joint mobility.

Randomization is another critical element. Randomizing participants into groups reduces selection bias and ensures comparability. Researchers can use simple randomization, stratified randomization (to balance variables like age or severity), or cluster randomization (group-level assignment). The choice depends on the study's objectives, resources, and constraints.

Data collection must follow a standardized protocol specifying what, when, and how data are collected to ensure consistency. Using electronic data capture systems with built-in validation rules minimizes errors. Training data collectors and conducting regular audits ensures protocol adherence and reduces variability.

2.3. Statistical design

The statistical methods must align with the research question and, subsequently, the data collected. Chi-square tests, t-tests, and ANOVA are suitable for straightforward comparisons, while non-inferiority tests are useful for assessing whether a new treatment is not worse than a standard. Advanced models like mixed-effects analyses are essential for handling repeated measures or hierarchical data.

Sample size and power calculations are crucial to determine the

https://doi.org/10.1016/j.jorep.2025.100573
Received 11 December 2024; Accepted 24 January 2025
Available online 4 February 2025

study's ability to detect meaningful differences. Small sample sizes risk Type II errors (false negatives), while overly large sizes waste resources on trivial effects. Effect size estimates from pilot studies or prior literature help optimize sample size calculations, accounting for expected attrition. It is improper to use study results to determine power for the same study, as this approach leads to circular reasoning and can produce misleading conclusions.[3]

2.4. Research feasibility

Feasibility assessments ensure a study's goals are achievable within available resources. This includes recruitment potential, funding, and access to facilities. Selected outcome measures must be practical to collect, reliable, and sensitive to the expected changes. Feasibility can be validated through pilot testing. Regular evaluations during the study allow for adjustments in response to challenges to maintain alignment with objectives.

2.5. Common data pitfalls in data collection and analysis

Data collection and analysis are particularly prone to errors that undermine the validity and reliability of results, even in well-designed studies.

2.6. Pitfalls in data collection

One of the most significant challenges in data collection is inconsistency (Fig. 1). This often arises from unclear data collection protocols and ambiguous coding conventions, such as using different systems to denote the same variable (e.g., "Yes/No" versus "1/0") or inconsistent handling of missing values (e.g., leaving fields blank versus using codes like "-9999"). Such inconsistencies can complicate analysis and lead to misclassification.[4,5]

Researchers risk discrepancies and errors during data entry and analysis without clear documentation of variable names, definitions, formats, and ranges. Multiple, uncoordinated datasets without a centralized master file add another layer of complexity. Researchers must establish standardized data entry rules and adopt clear conventions for coding variables. A robust data dictionary should be developed at the study's outset and shared. Centralizing data storage in a master dataset and using version-control, ensures consistency and reduces redundancy. Adequate training for data collectors is essential, with periodic quality checks to identify and correct errors.

2.7. Pitfalls in data analysis

Errors often stem from the misuse of statistical methods or misinterpretation of results, such as the overreliance on p-values as the sole measure of significance.[6,7] P-values indicate whether an effect exists but do not convey its magnitude or clinical significance, which is often influenced by sample size. Independent of sample size, effect sizes quantify standardized observed differences, ensuring that results are both statistically and clinically meaningful (Fig. 2). Researchers should complement p-values with effect sizes and confidence intervals, which provide a more nuanced understanding of the magnitude and precision of results.

Failing to adjust for multiple comparisons is defined as testing several hypotheses simultaneously without correcting for inflated Type I error rates, which increases the likelihood of false-positive results. Techniques such as Bonferroni corrections or false discovery rate control help account for multiple comparisons.

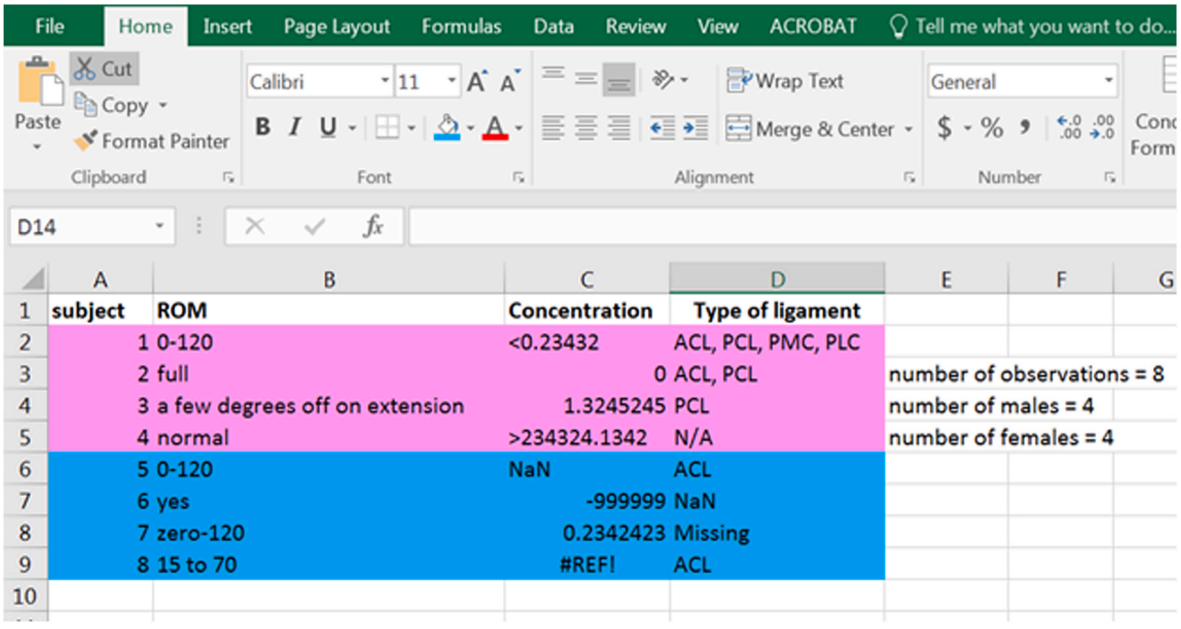
Similarly, the inappropriate application of statistical models can bias conclusions. For instance, repeated measures data require models such as mixed models or generalized estimating equations to account for within-subject correlations. Consulting with a biostatistician can prevent errors and enhance the credibility of the study's findings.

3. Discussion

Common pitfalls in research design, data collection, and analysis often undermine the validity of findings in orthopaedic research. Addressing these challenges requires thoughtful planning, standardized protocols, and adherence to ethical and methodological best practices.

4. Conclusion

Collaboration with biostatisticians during study design is critical to avoid common errors and ensure methodological rigor.



	A	B	C	D	E	F	G
1	subject	ROM	Concentration	Type of ligament			
2	1	0-120	<0.23432	ACL, PCL, PMC, PLC			
3	2	full	0	ACL, PCL	number of observations = 8		
4	3	a few degrees off on extension	1.3245245	PCL	number of males = 4		
5	4	normal	>234324.1342	N/A	number of females = 4		
6	5	0-120	NaN	ACL			
7	6	yes	-999999	NaN			
8	7	zero-120	0.2342423	Missing			
9	8	15 to 70	#REF!	ACL			
10							

Fig. 1. An example of collected data from unclear data collection protocols and inconsistent coding conventions. Inconsistent data collection protocols may result in non-standardized ROM descriptions, irregular concentration formatting, missing/invalid values, and inconsistent ligament type entries.

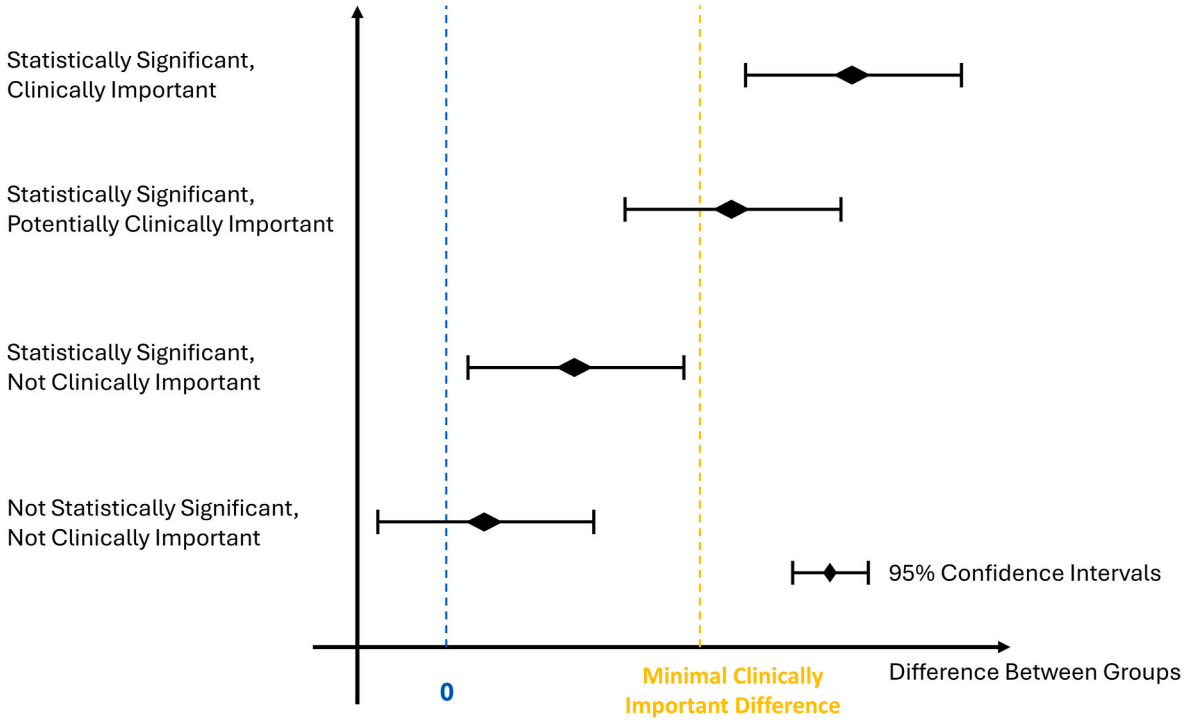


Fig. 2. Difference between statistical and clinical significance. Confidence intervals illustrating statistical significance (blue line at 0) and clinical importance (orange line at minimal clinically important difference), showing four scenarios of statistical and clinical significance.

CRediT authorship contribution statement

Emily Leary: Conceptualization, Methodology, Writing – review & editing, Project administration, Visualization. **Jinpu Li:** Writing – original draft, Visualization.

Data statement

No data were used in this work.

Guardian Patients consent

No patient data or patients were recruited for this study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

Emily Leary received consulting fees from EVLVE Analytics and Consulting LLC, payment or honoraria from JISAKOS, participated on a Data Safety Monitoring Board or Advisory Board for DSMB R01 McCrae PI (R01NR017168), leadership or fiduciary role in JISAKOS Editorial Board, American Statistical Association, Statistical Consulting Section and The Journal of Knee Surgery. The rest of the authors have no interests to declare.

Acknowledgement

The authors thank the faculty, residents, and fellows of the University of Missouri Department of Orthopaedic Surgery during 2017 for their interest and comments during a Grand Rounds presentation drawn from this work.

References

1. Walenkamp MM, Roes KC, Bhandari M, Goslings JC, Schep NW. Multiple testing in orthopedic literature: a common problem? *BMC Res Notes*. 2013;6(1):374. https://doi.org/10.1186/1756-0500-6-374.
2. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Jt Surg*. 2004;86(3):607–620.
3. Althouse AD. Post hoc power: not empowering, just misleading. *J Surg Res*. 2021;259:A3–A6. https://doi.org/10.1016/j.jss.2019.10.049.
4. Lee B, Ebrahimi M, Ektas N, et al. Implementation and quality assessment of a clinical orthopaedic registry in a public hospital department. *BMC Health Serv Res*. 2020;20(1):393. https://doi.org/10.1186/s12913-020-05203-8.
5. Kyte D, Ives J, Draper H, Keeley T, Calvert M. Inconsistencies in quality of life data collection in clinical trials: a potential source of bias? Interviews with research nurses and trialists. In: *Lam WWT*. vol. 8. PLoS ONE; 2013, e76625. https://doi.org/10.1371/journal.pone.0076625, 10.
6. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305–307. https://doi.org/10.1038/d41586-019-00857-9.
7. Sullivan GM, Feinn R. Using effect size—or why the *P* value is not enough. *J Grad Med Educ*. 2012;4(3):279–282. https://doi.org/10.4300/JGME-D-12-00156.1.

Emily Leary[a,b,*], Jinpu Li[a]

[a] *Thompson Laboratory for Regenerative Orthopaedics, Department of Orthopaedic Surgery, School of Medicine, University of Missouri, 1100 Virginia Ave, Columbia, MO, 65212, USA*

[b] *Department of Orthopaedic Surgery, University of Missouri, 1100 Virginia Ave, Columbia, MO, 65212, USA*

[*] Corresponding author. 1100 Virginia Ave, Columbia, MO, 65212 USA.
E-mail address: learye@umsystem.edu (E. Leary).