POSITION PAPER

Nr 2025;15 (2):121-129

The orthopedic surgeon's guide to statistics: interpreting research with confidence

Anna-Maria Nau[1], Nerissa Naidoo[2], Thomas Nau[2,3]

[1] *The Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, U.S.A.*
[2] *Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, U.A.E.*
[3] *King's College Hospital London, Dubai, U.A.E.*

CORRESPONDING AUTHOR:
Thomas Nau
King's College Hospital London Dubai
Marabea Street East
Dubai, U.A.E.
E-mail: thnau@hotmail.com

DOI:
10.32098/mltj.02.2025.01

LEVEL OF EVIDENCE: N/A

SUMMARY
This study provides an accessible guide to fundamental statistical principles for orthopedic surgeons, essential for interpreting research and applying evidence-based practices. Using a hypothetical study comparing robotic-assisted total knee replacement (RTKR) to conventional knee replacement (CKR), we demonstrate key statistical concepts like variability, hypothesis testing, and confidence intervals. We explain common errors, such as Type 1 and Type 2 errors, and strategies to address them. The paper covers statistical tests like t-tests, ANOVA, and regression, and examines how study quality and bias impact findings. This guide helps surgeons critically evaluate research to enhance clinical decision-making and contribute to orthopedic advancements.

KEY WORDS
Critical reading; evidence-based medicine; orthopedic; research; statistics.

INTRODUCTION

The rapid expansion of orthopedic research necessitates a solid understanding of statistical principles to ensure accurate interpretation of study findings. Orthopedic surgeons frequently encounter scientific literature in their clinical and academic pursuits, yet many lack formal statistical training, which may lead to misinterpretation of results and suboptimal decision-making in patient care. This study aims to bridge this knowledge gap by providing a structured and accessible guide to fundamental statistical concepts tailored specifically to the needs of orthopedic surgeons.

A key challenge in orthopedic research is differentiating high-quality evidence from studies that may be methodologically flawed. With the increasing availability of studies freely accessible on the internet, surgeons often encounter research of varying quality, making critical appraisal skills even more essential. In the era of evidence-based medicine, statistical literacy is crucial in assessing study validity, identifying biases, and critically appraising findings. Misinterpretation of statistical outcomes can lead to incorrect conclusions, potentially influencing clinical practice inappropriately. Therefore, enhancing orthopedic surgeons' ability to evaluate research critically will improve evidence-based decision-making and, ultimately, patient outcomes (1-3).

To illustrate the application of essential statistical concepts, a hypothetical research scenario is introduced, comparing robotic-assisted total knee replacement (RTKR) to conventional knee replacement (CKR). This study explains key statistical methods such as hypothesis testing, confidence intervals, and common statistical errors, while also emphasizing proper study design and data interpretation.

Furthermore, this investigation highlights the importance of study quality assessment, including randomization, blinding, and bias reduction strategies. The goal is to provide orthopedic surgeons with a practical framework for evaluating research, enhancing their ability to interpret findings accurately and apply them effectively in clinical practice. By improving statistical literacy, orthopedic

professionals will be better equipped to engage with scientific literature, contribute to research, and ensure optimal patient care based on sound evidence.

SCENARIO AND STUDY SETUP

A hypothetical orthopedic research team designed a study to evaluate the potential benefits of a novel robotic-assisted surgical technique in total knee replacement (RTKR). Conventional total knee replacement (CTKR) generally provides good results; however, approximately 20% of patients report ongoing pain and limited function following the procedure (4). The novel robotic-assisted technique is hypothesized to improve patient outcomes by enhancing surgical precision and optimizing implant positioning.

To assess clinical effectiveness, the study employs the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) as the primary outcome measure. The WOMAC score is widely used in orthopedic research to evaluate pain, stiffness, and function in patients undergoing total knee replacement. The scoring system ranges from 0 to 96 points, where 0 represents the best possible outcome and 96 the worst. Its validity and reliability in assessing post-TKR function make it a suitable measure for this investigation (5).

To ensure methodological rigor, key aspects of study design, such as randomization and blinding, were incorporated (6-11). Patients meeting the eligibility criteria for total knee replacement were randomly assigned to either the RTKR or CTKR group to minimize selection bias. Additionally, assessors evaluating the WOMAC scores remained blinded to the surgical approach used, reducing the risk of detection bias.

SAMPLE MEAN AND VARIABILITY (12-16)

The orthopedic researchers performed RTKR on a group of patients, and after one year, the WOMAC score was measured in the sample. The mean score was calculated, and the researchers assumed that this sample mean score can be generalized to all patients who undergo RTKR. However, if another sample was collected or other patients were included, the sample means would have likely varied somewhat. This raised the question: How confident can the researchers be that the mean WOMAC score of the study sample applies to all patients undergoing RTKR?

To answer this, it is important to understand the concepts of standard deviation and standard error (**figure 1**).

- Standard Deviation (SD) measures variability within the population. Different patients will naturally have different WOMAC scores, and SD represents the average distance of any value from the mean.
- Standard Error (SE) measures the variability of the sample means. The larger the sample size, the smaller the standard error, as the sample means become closer together. The standard error is commonly used to calculate a confidence interval (CI), which typically represents a 95% confidence level. A 95% CI means that 95 out of 100 samples would contain the population mean within this interval (17, 18).

In this case, if the sample mean WOMAC score for RTKR was 50 with a 95%CI of [45-55], the researchers were 95% confident that the true population mean lies within this range.

TESTING A HYPOTHESIS (19-25)

The research team sought to compare the outcomes (WOMAC scores) between RTKR and CTKR to determine if the robotic technique provided a measurable benefit. To achieve this, they formulated the following research question: "Does RTKR result in better mean WOMAC scores compared to CTKR?".

In statistical analysis, hypothesis testing typically begins with a null hypothesis (H_0), which assumes that no difference exists between the mean WOMAC scores of RTKR and CTKR. This is opposed by the research hypothesis (H_1), which suggests that RTKR leads to improved WOMAC scores.

Patients were divided into two groups, each receiving one of the surgical techniques. The researchers then analyzed the two sample means to assess whether the observed difference was statistically significant. The next section will explore the statistical methods used to determine significance, ensuring that results are not due to random variation alone.

EXPLANATION OF TYPE 1 AND TYPE 2 ERRORS (26-31)

After collecting data from both groups (RTKR and CTKR), the research team conducted hypothesis testing to determine whether significant differences in WOMAC scores existed between the two surgical techniques.

Formulating the hypotheses

1. Null Hypothesis (H_0): There is no difference in mean WOMAC scores between RTKR and CTKR.
   - H_0: "The mean WOMAC score for RTKR is equal to that of CTKR".

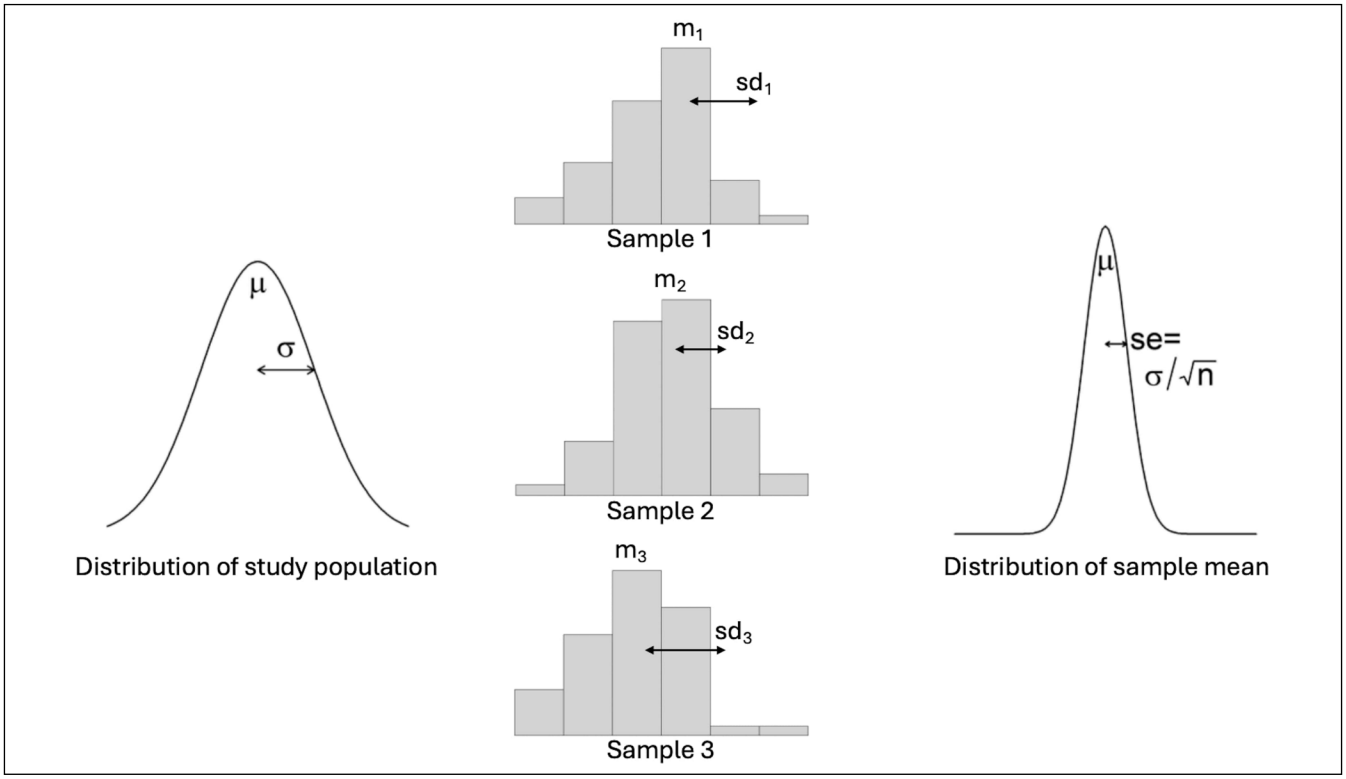


Figure 1. Starting with a population distribution defined by a mean (μ) and a standard deviation (σ) as illustrated on the left, individual samples can be drawn from this population. The center panel shows three such samples, each characterized by its own mean (m) and standard deviation (sd), serving as estimates for μ and σ. On the right, the distribution of sample means is displayed, representing the mean values of all possible samples. This distribution has the same mean as the population (μ) but exhibits variability described by the standard error (se), which is influenced by the population standard deviation (σ) and the sample size (n).

2. Alternative Hypothesis (H_1): There is a difference in mean WOMAC scores between RTKR and CTKR.
   - H_1: "The mean WOMAC score for RTKR is different (better or worse) than that of CTKR".

The objective of hypothesis testing is to collect sufficient evidence to either reject the null hypothesis (favoring the alternative hypothesis) or fail to reject the null hypothesis (indicating insufficient evidence to support a difference).

Type 1 error (false positive)

A Type 1 error occurs when the null hypothesis is incorrectly rejected, meaning that the study concludes there is a difference between RTKR and CTKR when no such difference exists in the population.

- The probability of making a Type 1 error is denoted by alpha (α), typically set at 0.05 (5%). This means there is a 5% chance that the study will wrongly conclude that RTKR is better (or worse) than CTKR, even if both techniques performed similar.

Clinical Consequences of a Type 1 error

- Unnecessary changes in surgical practices.
- Increased costs for the healthcare system and patients due to adopting a technique that does not provide real benefits.
- Ethical concerns if patients are exposed to interventions without proven superiority.

Type 2 error (false negative)

A Type 2 error occurs when the null hypothesis is incorrectly accepted, meaning the study concludes there is no difference when, in fact, RTKR does offer superior outcomes.

- The probability of making a Type 2 error is denoted by beta (β), and it is related to study power (1 - β). A study

with low power (*e.g.*, small sample size) is more prone to Type 2 errors.

Clinical Consequences of a Type 2 error

- Missed opportunity to improve patient outcomes through superior surgical techniques.
- Delay in the adoption of beneficial new technologies.
- Reduced innovation and advancement in orthopedic surgery due to failure to recognize true benefits.

Reducing Type 1 and Type 2 errors

- Alpha (α) and Type 1 errors: to reduce the risk of a Type 1 error, the alpha level could be set lower than 0.05 (*e.g.*, at 0.01). However, this would increase the risk of a Type 2 error, as it becomes harder to find statistically significant results.
- Beta (β), power, and Type 2 errors: to reduce the risk of a Type 2 error, the study must have sufficient power. Power depends on several factors, including:
  1. Sample size: larger sample sizes increase power by reducing the standard error, making it easier to detect real differences between groups.
  2. Effect size: studies with larger effect sizes (*i.e.*, bigger differences between the groups being compared) are more likely to find significant results. Smaller effect sizes require more participants to detect.
  3. Variability of the data: less variability within the data (*i.e.*, smaller standard deviations) makes it easier to detect significant differences.

By ensuring appropriate study design and adequate sample sizes, researchers can minimize the risks of both Type 1 and Type 2 errors, improving the reliability of study conclusions.

COMMON STATISTICAL TESTS IN ORTHOPEDIC RESEARCH (32-35)

Understanding which statistical test to use is crucial for accurately analyzing data and interpreting results. In orthopedic research, several key statistical tests are commonly applied to compare outcomes and determine the significance of observed differences. Below, the most relevant tests for comparing clinical outcomes are discussed, with examples that orthopedic surgeons might encounter in their practice.

T-test (independent samples t-test)

The t-test is used to compare the means of two independent groups to determine whether a statistically significant difference exists between them.

- Example: comparing the mean WOMAC scores between patients who underwent robotic-assisted total knee replacement (RTKR) and those who underwent conventional total knee replacement (CTKR). The t-test assesses whether the difference in mean scores between the two groups is statistically significant.
- Hypothesis:
  - Null hypothesis (H_0): there is no difference in the mean WOMAC score between RTKR and CTKR.
  - Alternative hypothesis (H_1): there is a difference in the mean WOMAC score between RTKR and CTKR.
- Application: If the t-test yields a p-value less than 0.05, the null hypothesis is rejected, indicating that the difference in scores between the two techniques is statistically significant.

ANOVA (analysis of variance)

ANOVA is used to compare means across more than two groups or to examine the influence of additional variables on an outcome.

- One-Way ANOVA: used when comparing the means of more than two groups based on one independent variable.
  - Example: assessing whether age influences outcomes of total knee replacement (TKR) by categorizing patients into three groups: under 50 years old, 50-65 years old, and over 65 years old.
  - Hypothesis:
    - Null hypothesis (H_0): there is no difference in mean WOMAC scores between the age groups.
    - Alternative hypothesis (H_1): at least one age group has a different mean WOMAC score.
  - Application: a significant ANOVA result ($p < 0.05$) suggests that at least one group differs, prompting post-hoc tests (*e.g.*, Tukey's test) to determine which groups differ.
- Two-Way ANOVA: examines the effect of two independent variables on the outcome and their interaction.
  - Example: assessing the influence of both age and walking activity level (low, moderate, high) on WOMAC scores.
  - Application: this test determines whether age and walking activity independently affect outcomes or if their combination influences recovery after knee replacement.

Chi-Square test

The chi-square test assesses relationships between categorical variables and is particularly useful for analyzing complication rates, success/failure rates, or other binary outcomes.

- Example: comparing complication rates between RTKR and CTKR patients (*e.g.*, whether each patient experienced a complication: yes/no).
- Hypothesis:
  - Null hypothesis (H_0): there is no difference in complication rates between RTKR and CTKR.
  - Alternative hypothesis (H_1): there is a difference in complication rates between RTKR and CTKR.
- Application: a significant chi-square result ($p < 0.05$) suggests that complication rates differ between the surgical techniques.

Pearson correlation

Pearson correlation measures the strength and direction of the relationship between two continuous variables.

- Example: evaluating the relationship between BMI and WOMAC scores following knee replacement surgery.
- Application:
  - An r-value close to +1 indicates a strong positive relationship (higher BMI is associated with worse WOMAC scores).
  - An r-value close to -1 indicates a strong negative relationship.
  - An r-value near 0 suggests no linear relationship.

Linear regression

Linear regression models the relationship between one or more independent variables (predictors) and a continuous dependent variable (outcome).

- Example: predicting postoperative WOMAC scores based on patient age, BMI, and walking activity level.
- Application:
  - The model provides coefficients for each predictor, showing how the WOMAC score changes with each unit increase in variables such as age or BMI.

CHOOSING THE RIGHT TEST (36-39)

Selecting an appropriate statistical test depends on the type of data being analyzed and the hypothesis being tested. **Table I** provides a summary of statistical tests, their purposes, applicable data types, and example use cases.

Table I. Summary of the common statistical tests, their purposes, data types, and examples.

Statistical Test	Purpose	Type of Data	Example
T-Test	Comparing the means of two independent groups.	Continuous	Comparing mean WOMAC scores between RTKR and CTKR patients.
ANOVA	ANOVA is used to compare the means across multiple groups (one-way ANOVA) or to investigate the interaction between two or more independent variables (two-way ANOVA).	Continuous	One-way ANOVA: comparing mean WOMAC scores across three age groups. Two-way ANOVA: investigate how age and walking activity affect WOMAC scores.
Chi-square test	Analyzing relationships between categorical variables.	Categorical	Testing whether the proportion of patients experiencing complications (yes/no) differs between RTKR and CTKR groups.
Pearson correlation	Measuring the strength and direction of the relationship between two continuous variables.	Continuous	Evaluating the correlation between BMI and WOMAC scores to determine if higher BMI is associated with poorer outcomes.
Linear regression	Predicting outcomes based one or more independent variables.	Continuous (dependent variable); mixed (independent variables)	Predicting WOMAC scores based on age, BMI, and surgical type (RTKR or CTKR).

- Example: if the goal is to evaluate the effects of both age and walking activity level (low, moderate, high) on WOMAC scores, a two-way ANOVA is appropriate. This test examines the main effects of each independent variable (age and walking activity) and assesses whether their interaction effect influences the outcome. By doing so, it determines whether age and activity levels independently affect WOMAC scores or if their combination leads to significant differences in outcomes.

Choosing the correct test ensures the validity of statistical conclusions and strengthens the reliability of research findings in orthopedic surgery.

ASSESSING STUDY QUALITY AND BIAS IN ORTHOPEDIC RESEARCH (40-45)

The validity of study conclusions in orthopedic research depends heavily on study design and the control of potential sources of bias. Understanding these concepts is essential for critically evaluating research articles and applying their findings in clinical practice. Below, key factors affecting study quality and common forms of bias are discussed.

Randomization and blinding

Randomization is the process of assigning patients to different treatment groups by chance, ensuring comparability between groups at the study's outset. This approach minimizes selection bias, where pre-existing differences (*e.g.*, age, comorbidities) could influence outcomes.

- Example: in a study comparing RTKR to CTKR, randomization ensures that baseline factors like BMI and activity level are evenly distributed, reducing their potential impact on the outcome.

Blinding conceals treatment allocation from participants, healthcare providers, or outcome assessors to reduce performance bias (differential treatment) and detection bias (biased assessment of results).

- Example: if surgeons know which patients receive RTKR, they may unintentionally modify postoperative care. Blinding ensures unbiased assessment of recovery outcomes.

Sample size and statistical power

A well-designed study requires an adequate sample size to detect true differences between groups. Insufficient sample sizes increase the risk of a Type 2 error (failing to detect a real difference).

- Example: a study comparing RTKR and CTKR with only 20 patients per group may lack power to detect a meaningful difference in WOMAC scores. Increasing the sample size to 100 patients per group improves the likelihood of identifying clinically significant differences.

A power analysis during study design helps determine the required sample size based on expected effect sizes, significance levels, and study power (typically 80% or higher).

Confounding variables

Confounding variables influence both the treatment and outcome, potentially leading to false conclusions about causality.

- Example: if a study comparing RTKR and CTKR includes older patients primarily in one group, age may confound the results. Adjusting for confounders through statistical methods (*e.g.*, ANCOVA, stratified analysis) helps isolate the treatment effect.

Selection bias

Selection bias arises when patient characteristics systematically differ between treatment groups, affecting study validity.

- Example: if patients with severe arthritis preferentially receive RTKR, worse outcomes in this group may be due to disease severity rather than surgical technique. Randomization reduces this bias, while observational studies require statistical adjustments (*e.g.*, propensity score matching).

Attrition bias and loss to follow-up

Attrition bias occurs when dropout rates differ systematically between groups, leading to biased results.

- Example: if more CTKR patients with poor outcomes drop out, the study may overestimate CTKR's effectiveness. Reporting loss to follow-up and using intention-to-treat (ITT) analysis mitigate this bias.

Reporting bias

Reporting bias occurs when only certain outcomes or time points are selectively reported, often favoring positive results.

- Example: if a study on RTKR reports significant improvements at 6 months but omits 1-year data where no difference exists, readers may overestimate RTKR's benefits. To prevent this, researchers should pre-register study protocols and report all planned outcomes.

By recognizing and addressing these biases, orthopedic researchers and clinicians can critically appraise study findings, ensuring that clinical decisions are based on robust and reliable evidence.

PRACTICAL TOOLS FOR ASSESSING STUDY QUALITY (46-48)

When evaluating orthopedic research, surgeons can utilize standardized tools to assess study quality and identify potential sources of bias. One widely recognized framework is the CONSORT (Consolidated Standards of Reporting Trials) checklist, which provides a structured approach for evaluating the transparency and methodological rigor of randomized controlled trials (RCTs) (6).

The original CONSORT checklist consists of 25 items covering various aspects of trial design, implementation, and reporting. However, for practical application in orthopedic research, we have tailored our discussion to highlight the most relevant aspects for clinicians. These selected elements emphasize key considerations that orthopedic surgeons can use to assess the validity and applicability of studies in their daily practice.

Critical reading checklist for orthopedic surgeons

When critically evaluating a study, orthopedic surgeons should consider the following key questions:

- Was the study randomized and blinded? Ensuring randomization and blinding minimizes selection and performance bias, increasing the reliability of the results.
- Was the sample size adequate to detect clinically meaningful differences? Studies with insufficient sample sizes may lack statistical power, increasing the risk of Type 2 errors.
- Were confounding variables considered and adjusted for? Proper statistical adjustments (*e.g.*, ANCOVA, propensity score matching) help isolate the true effect of the intervention.
- Was there significant loss to follow-up, and was an intention-to-treat (ITT) analysis performed? High attrition rates can introduce bias; ITT analysis ensures that all randomized participants are accounted for.
- Were all relevant outcomes reported, including both significant and non-significant results? Selective reporting of only favorable results may lead to an overestimation of treatment efficacy.

By systematically applying these principles, orthopedic surgeons can enhance their ability to critically assess research findings, ensuring that clinical decisions are based on robust and reliable evidence. This structured approach fosters evidence-based practice, leading to improved patient care and better integration of new research findings into orthopedic surgery (49, 50).

CONCLUSIONS

In today's rapidly evolving field of orthopedic surgery, the ability to critically assess and interpret scientific studies is crucial for making informed clinical decisions. A comprehensive understanding of basic statistical principles, recognition of potential biases, and the ability to evaluate study quality enable orthopedic surgeons to differentiate high-quality evidence from flawed research.

By mastering these fundamental concepts, clinicians can ensure that their practices are guided by robust evidence, ultimately leading to improved patient outcomes and more efficient use of healthcare resources. Whether comparing innovative surgical techniques such as robotic-assisted knee replacement or analyzing long-term outcomes of conventional procedures, proficiency in statistical methods –including t-tests, ANOVA, and chi-square tests – facilitates accurate and meaningful data interpretation.

Beyond statistical methods, an awareness of biases – such as selection bias, performance bias, and attrition bias – is essential in preventing misleading conclusions. Surgeons must also critically evaluate study methodologies, with a focus on randomization, sample size adequacy, and proper handling of confounding variables, to enhance the validity of research findings.

In conclusion, continuous learning and application of these statistical principles empower orthopedic surgeons to remain at the forefront of evidence-based medicine. As the volume of available literature continues to expand, these analytical skills will remain indispensable in advancing orthopedic surgery, improving individual patient care, and contributing to the broader field of orthopedic research.

FUNDINGS

None.

DATA AVAILABILITY

N/A.

CONTRIBUTIONS

AN: description of statistical techniques, writing statistical part, final corrections. NN: literature search, writing medical part. TN: setting up the idea and structure, writing medical part, final corrections.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

REFERENCES

1. Tez M, Yildiz B. How Reliable Are Medical Textbooks. J Grad Med Educ. 2017;9(4):550-3. doi: 10.4300/JGME-D-17-00209.1.
2. Battineni G, Baldoni S, Chintalapudi N, et al. Factors affecting the quality and reliability of online health information. Digit Health. 2020;6:2055207620948996. doi: 10.1177/2055207620948996.
3. Ioannidis J. Why most published research findings are false. PLoS Med. 2005;2(8):e124. doi: 10.1371/journal.pmed.0020124.
4. Nakano N, Shoman H, Olavarria F, Matsumoto T, Kuroda R, Khanduja V. Why are patients dissatisfied following a total knee replacement? A systematic review. Int Orthop. 2020;44(11):1971-2007. doi: 10.1007/s00264-020-04607-9.
5. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient-relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40.
6. Schulz K, Altman D, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. BMJ. 2010;340:c332. doi: 10.1136/bmj.c332.
7. Zhang TD, Scheinberg M, Hoffman A, Mehta L. Evolution in the use of statistical testing in the field of orthopedics from 1993 to 2023. Cureus. 2024;16(2):e20234. doi: 10.7759/cureus.72709.
8. Ranstam J. Problems in orthopedic research. Acta Orthop Scand. 2002;73(2):123-9. doi: 10.1080/00016470216327.
9. Leary E, Li J. Biostatistics in orthopedic surgery: common data pitfalls and how to avoid them. J Orthop Rep. 2025;10(1):22-30.
10. Indrayan A. Statistical fallacies in orthopedic research. Indian J Orthop. 2007;41(1):37-43. doi: 10.4103/0019-5413.30524.
11. Petrie A. Statistics in orthopedic papers. J Bone Joint Surg Br. 2006;88(9):1121-36. doi: 10.1302/0301-620X.88B9.17896.
12. Parsons NR, Price CL, Hiskens R, Achten J. An evaluation of the quality of statistical design and analysis of published medical research. BMC Med Res Methodol. 2012;12:60. doi: 10.1186/1471-2288-12-60.
13. Walenkamp MMJ, Roes KCB, Bhandari M. Multiple testing in orthopedic literature: a common problem? BMC Res Notes. 2013;6:374. doi: 10.1186/1756-0500-6-374. J Arthroplasty. 2022;37(10):1945-50. doi: 10.1016/j.arth.2022.05.026.
14. Zaniletti I, Devick KL, Larson DR, Lewallen DG. P-values and power in orthopedic research: myths and reality.
15. Park MS, Kim SJ, Chung CY, Choi IH, Lee SH, Lee KM. Statistical consideration for bilateral cases in orthopedic research. J Bone Joint Surg Am. 2010;92(8):1732-7. doi: 10.2106/JBJS.I.00724.
16. Wright JG, Feinstein AR. Improving the reliability of orthopedic measurements. J Bone Joint Surg Br. 1992;74(2):287-91. doi: 10.1302/0301-620X.74B2.1544971.
17. Kallogjeri D, Francis DO, Piccirillo JF. Use of Confidence Intervals to Improve the Quality of Research Results Reporting and Interpretation. JAMA Otolaryngol Head Neck Surg. 2019;145(6):491-3. doi: 10.1001/jamaoto.2019.0793.
18. Sim J, Reid N. Statistical inference by confidence intervals: issues of interpretation and utilization. Phys Ther. 1999;79(2):186-95.
19. Sedgwick P. Confidence intervals and statistical significance: rules of thumb. BMJ. 2012;345:e4960. doi: 10.1136/bmj.h1113.
20. Smith SD. Statistical tools in the quest for truth: hypothesis testing, confidence intervals, and clinical studies. Ophthalmology. 2008;115(3):423-4. doi: 10.1016/j.ophtha.2007.12.016.
21. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation over hypothesis testing. BMJ. 1986;292(6522):746-50. doi: 10.1136/bmj.292.6522.746.
22. Wilcox RR, Serang S. Hypothesis testing, p values, confidence intervals, and effect sizes. Educ Psychol Meas. 2017;77(4):673-89. doi: 10.1177/0013164416667983.
23. Perdices M. Null hypothesis significance testing and Bayesian methods in modern research. Brain Impairment. 2018;19(2):109-19.
24. Schober P, Vetter TR. Confidence intervals in clinical research. Anesth Analg. 2020; 130(5):1303. doi: 10.1213/ANE.000000000
25. Fethney J. Statistical and clinical significance, and how to use confidence intervals. Aust Crit Care. 2010;23(2):93-7. doi: 10.1016/j.aucc.2010.03.001.
26. Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research - a review of common pitfalls. Swiss Med Wkly. 2007;137(3-4):44-9. doi: 10.4414/smw.2007.11587.
27. Almuqrin MA. Next-generation statistical methodology: advances in health science research. Alexandria Eng J. 2024;63(1):88-105.
28. Ramirez L, Liuzzi J, Yoo C. Big data analysis using modern statistical and machine learning methods in medicine. Int Neurourol J. 2014;18(2):50-7. doi: 10.5213/inj.2014.18.2.50.
29. Yi D, Ma D, Li G, Zhou L, Xiao Q, Zhang Y, Liu X. Statistical use in clinical studies: is there evidence of a methodological shift? PLoS One. 2015;10(10):e0140159. doi: 10.1371/journal.pone.0140159.
30. Marinell G, Strasak AM, Zaman Q. The use of statistics in medical research: a comparison of The New England Journal of Medicine and Nature Medicine. Am Stat. 2007;61(3):230-240.
31. Simon R. A decade of progress in statistical methodology for clinical trials. 1991 Dec;10(12):1789-817. doi: 10.1002/sim.4780101203.
32. Edwards LJ. Modern statistical techniques for the analysis of longitudinal data in biomedical research. Pediatr Pulmonol. 2000;30(4):330-40. doi: 10.1002/1099-0496(200010)30:4<330::AID-PPUL10>3.0.CO;2-D.
33. Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980s. Stat Med. 1991;10(12):1887-99. doi: 10.1002/sim.4780101205.
34. Hannigan A, Lynch CD. Statistical methodology in oral and dental research: pitfalls and recommendations. J Dent. 2013;41(5):367-75. doi: 10.1016/j.jdent.2013.02.005.
35. Scales CD, Norris RD, Peterson BL. Clinical research and statistical methods in the urology literature. J Urol. 2005;174(5):2027-

31. doi: 10.1097/01.ju.0000173640.91654.b5.
36. Horton NJ, Switzer SS. Statistical methods in the Journal. N Engl J Med. 2005;353(18):1927-30. doi: 10.1056/NEJM200511033531823.
37. Campbell G. Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. Stat Med. 1994;13(5-6):499-508. doi: 10.1002/sim.4780130513.
38. Rajula HSR, Verlato G, Manchia M, Antonucci N. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas). 2020;56(9):455. doi: 10.3390/medicina56090455.
39. Simon R. A decade of progress in statistical methodology for clinical trials. Stat Med. 1991;10(12):1899-920. doi: 10.1002/sim.4780101203.
40. Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C. Bias in clinical research. Kidney Int. 2008;73(2):148-53. doi: 10.1038/sj.ki.5002648
41. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010;126(2):619-25. doi: 10.1097/PRS.0b013e3181de24bc.
42. Greenhalgh T, Ryan S, Snow R, Rees S, Salisbury H. Six 'biases' against patients and carers in evidence-based medicine. BMC Med. 2015;13:200. doi: 10.1186/s12916-015-0437-x.
43. McAlister FA, Graham I, Karr GW. Evidence-based medicine and the practicing clinician. J Gen Intern Med. 1999;14(4):236-42. doi: 10.1046/j.1525-1497.1999.00323.x.
44. Attia A. Bias in RCTs: confounders, selection bias and allocation concealment. Utoronto Scholaris. 2005.
45. Shyam A. Bias and the evidence 'biased' medicine. J Orthop Case Rep. 2015 Jul-Sep;5(3):1-2. doi: 10.13107/jocr.2250-0685.291.
46. Lambert J. How to assess bias in clinical studies? Clin Orthop Relat Res. 2011 Jun;469(6):1794-6. doi: 10.1007/s11999-010-1538-7.
47. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017;390(10092):415-423. doi: 10.1016/S0140-6736(16)31592-6.
48. Littell JH. Evidence-based or biased? The quality of published reviews of evidence-based practices. Child Youth Serv Rev. 2008;30(10):1219-25.
49. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. J Eval Clin Pract. 2014;20(6):908-14. doi: 10.1111/jep.12147.
50. Sica GT. Bias in research studies. Radiology. 2006;238(3):780-9. doi: 10.1148/radiol.2383041109.