DOI: https://doi.org/10.4038/tsljos.v9i1.25

Demystifying Statistics in Surgical Research: Navigating the Path to Effective Treatment Evaluation

Chulasiri M.P.P.U[1],
Atapattu A.H.A.K.S[2]

1 Anti Malaria Campaign, Ministry of Health

2 Postgraduate Institute of Medicine, University of Colombo.

Abstract

Many surgeons often find statistics to be unfamiliar and challenging. In the context of contemporary medicine, understanding statistical concepts is crucial for professional development. This article provides a basic overview of statistics and terminology commonly used in Orthopaedics and Trauma research. Orthopaedic surgery research often involves comparing two groups: one that has undergone a specific procedure and another that hasn't or has had a different procedure. The central question is whether the differences in outcomes are due to the surgical intervention or random chance. For instance, in a study assessing a new knee replacement surgery, there's a need to balance Type 1 and Type 2 errors, which relate to incorrectly concluding effectiveness. A larger sample size increases statistical power but also comes with practical and cost constraints. Choosing the appropriate study design depends on research questions, ethical considerations, resources, and the nature of the intervention. Common designs include Randomized Controlled Trials (RCTs), Prospective Cohort Studies, Retrospective Cohort Studies, Case-Control Studies, Cross-Sectional Studies, Systematic reviews and Meta-Analyses. Understanding the nature of data (qualitative and quantitative) and data distribution, especially Gaussian distribution, is essential in orthopaedic or any type of research. Different types of data require distinct statistical approaches, ensuring the selection of the most suitable analytical techniques. Proper selection of a statistical test starts with formulating a null hypothesis. The p-value is crucial in assessing the strength of evidence against the null hypothesis. A low p-value (typically < 0.05) indicates a significant difference, while a high p-value suggests results could be due to random chance. Various statistical tests, like Chi-Square, Student's T Test, ANOVA, and regression analysis, can be used depending on the research question and data type. Each test is applied to analyze specific aspects of the data, such as relationships between categorical variables, differences in means, or the influence of various factors. In summary, understanding statistics is fundamental in assessing the efficacy of orthopaedic treatments. Researchers must manage errors, choose the right study design, understand the nature of data, and select appropriate statistical tests. This statistical literacy is essential for evidence-based decision-making in orthopaedic practice.

Keywords: Orthopaedic surgery research, Statistical analysis, Data distribution, Hypothesis testing, Effectiveness assessment

Introduction

Many surgeons find statistics to be an unfamiliar and challenging field. In the context of contemporary medicine, where publications and research play a crucial role in professional development, it is of utmost importance to have a solid understanding of statistical concepts (1). The ability to effectively examine a study and make informed decisions regarding its validity, as well as the capacity to present one's own work in a compelling and persuasive manner, are both facilitated by this skill. This article offers a rudimentary

- Dr. Pubudu Chulasiri, MBBS (Colombo), MSc, MD (Community Medicine), Consultant Community Physician, Program Manager in Trainings and Vector control, Anti Malaria Campaign
- Dr. Kasuni Atapattu, MBBS(Colombo), MSc Registrar in Community Medicine, Postgraduate Institute of Medicine, University of Colombo

Corresponding author:
Chulasiri M.P.P.U
Email:
pubudu.chulasiri@gmail.com

Sri Lankan Journal of Orthopeadic
2023 : 08 : 47-53

overview of the statistics and the terminology used for most studies done in Orthopaedics and Trauma.

In the realm of orthopaedic surgery research, a fundamental study typically involves the examination of two distinct groups: one group of patients who have undergone a specific orthopaedic surgical procedure, and another group of patients who have not undergone any surgical intervention or undergone a different surgical procedure or a variation of the specific procedure under consideration. The research team meticulously documents the outcomes for both groups. The central inquiry revolves around determining whether there exists a notable dissimilarity in outcomes—this is the anticipated outcome for the researcher. The key question at hand is, what is the probability that the disparity in outcomes can be attributed solely to the orthopaedic surgical intervention and is not merely a result of random chance?

The Gold standard of studies is a RCT. Imagine a group of researchers-in this case orthopaedic surgeons- aim to assess the effectiveness of a novel knee replacement surgery. Researchers divide their participants into two groups:

> **Intervention Group:** This group consists of individuals who have received the new knee replacement surgery.
>
> **Control Group:** This group comprises individuals with similar knee conditions, but they have had the conventional knee replacement surgery.

First the researchers need to determine that both groups are comparable in every aspect other than what knee replacement they will receive; a similar age range, pre-operative functional range, pre-operative pain score, co-morbidities etc....Otherwise it could be interpreted that it was something else that was responsible for the outcome and not the intervention itself. These factors that may have an extraneous influence on the outcome are called confounders and need to be accounted for.

This is what leads to the term *Controlled.* Thereafter the scientists need to make sure that every patient has an equal chance of either receiving the intervention -in this case the new knee replacement -or not. In other words, there is no *bias* in who receives and who does not receive the intervention. Because inadvertently one might be influenced by who gets the new knee or not. This process of randomly allocating the treatment leads to the term *Randomized.* Which leads to the term RCT. The platinum standard of a RCT is a triple blinded RCT where the subjects, researchers and those who analyse the research data are all kept blinded as to which is which. However, as one can see even if one keeps the patients and the data analysts blinded it is a near impossibility to keep the surgeon blinded as to what he/she is doing!

In the above example both groups are closely monitored over a specific period, and their knee function and pain levels are recorded. The hypothesis here is that the new knee replacement surgery will lead to better outcomes in terms of improved knee function and reduced pain compared to the control group.

In the orthopaedic study assessing the effectiveness of a new knee replacement surgery, researchers face the challenge of distinguishing whether observed differences in outcomes between the intervention and control groups are real or the result of random chance. To overcome these challenges, researches needs to identify following errors that are possible (2).

1. *Type 1 Error (False Positive):*

If the researchers are too strict in their judgment, they might wrongly think the surgery is effective when there's no real difference between the new knee replacement surgery versus the conventional surgery. It's like getting too excited for something that's not actually working.

2. *Type 2 Error (False Negative):*

On the other hand, if they are too lenient in their judgment, they could miss the true benefits of the surgery. They might conclude there's no real difference when there actually is. This is like missing out on something good.

A larger sample size provides increased statistical power, making it more likely to detect real effects while reducing the risk of both Type 1 and Type 2 errors. However, increasing the sample size can come with practical and cost constraints, so it's essential to strike a balance between achieving adequate statistical power and managing resources effectively in research studies (3).

Balancing these errors is crucial for the orthopaedic study. Researchers must choose an appropriate level of judgment and strive to maximize the study's statistical power. By doing so, they ensure that any significant differences in outcomes are reliably detected while minimizing the chances of both false positive and false negative conclusions. In the above example this careful balance allows for a more accurate assessment of the surgical intervention's true impact on knee function and pain levels.

It is imperative to understand that sample size, specificity of the research question and the power of the study are intimately related to one another. For example, if the researchers want to be 95% confident - a very powerful study- and consider that a 5% improvement of the functional outcome is the significant difference, then the sample size will have to be much larger than if they expected a 30% improvement in the functional outcome as being significant. On the other hand, if they wanted only to be 75% confident about their conclusions, then the samples once again can be smaller. The lower the difference that is expected to be demonstrated as significant; the sample size will increase with the power of the study.

A proper study design and adequate sample size is mandatory to keep this balance.

The choice of the best study design to assess the effectiveness of an orthopaedic surgical intervention depends on various factors, including the research question, ethical considerations, available resources, and the nature of the intervention (4-6). Sample size calculators or statistical software to calculate the required sample size, considering alpha, power, and effect size can be utilized. However, these are best discussed with a statistician before commencing on your research project.

In orthopaedic research, various study designs are employed to assess the effectiveness of surgical interventions. The gold standard is the Randomized Controlled Trial (RCT), where patients are randomly assigned to receive the 'surgery of interest '(intervention group) or not (control group), allowing for a rigorous evaluation. This is unfortunately not possible at all times.(5, 7, 8). Prospective Cohort Studies involve following a group of patients over time, providing insights into long-term outcomes of different modes of intervention. This is generally the most common and accepted study method in surgical specialties inclusive of orthopaedics. Retrospective Cohort Studies are the next best type of studies; they analyze existing medical records to compare patient outcomes, offering real-world effectiveness insights. Case-Control Studies are useful for rare conditions or when prospective studies aren't feasible, comparing historical data between surgery and non-surgery groups (4). Cross-Sectional Studies measure outcomes at a specific point in time, offering a snapshot of patient results, while Meta-Analyses and Systematic Reviews synthesize existing research to provide a comprehensive overview of evidence and are considered as the highest level of evidence based medicine.

The choice of study design depends on the specific research question, the availability of resources, and the ethical considerations involved in conducting orthopaedic research. In many cases, a well-designed RCT or a prospective cohort study offers the most robust evidence for assessing the effectiveness of surgical interventions, but other designs can also provide valuable insights, especially when RCTs are not feasible or ethical (4).

In any research study, a fundamental principle is to thoroughly understand and accurately describe the nature of the data being collected. This crucial step forms the foundation for determining the appropriate statistical analysis methods. Different types of data, such as categorical, continuous, or ordinal, require distinct statistical approaches. For example, nominal or categorical data might necessitate chi-squared tests, while continuous data may be analyzed using t-tests or regression models. The critical aspect here is that the choice of statistical analysis hinges on a comprehensive comprehension of the data's characteristics, distribution, and measurement scales. An accurate description of the data ensures that researchers select the most suitable analytical techniques, ultimately leading to more valid and reliable research findings. It underscores the importance of data quality and integrity in the research process (5-7, 9).

Types of Data

Raw data refers to a collection of observations that have been obtained from a sample taken from the larger population of interest. Each discrete data point is referred to as a variable. These can be divided into two broad categories- Qualitative data/ Categorical data and Quantitative/numerical data (5)

1. ***Qualitative data- Often known as categorical data, does not include numbers but rather categories.***

Nominal data (Categorical): Researchers may use a nominal scale to categorize patients' preferences for anesthesia type during the surgery (forefoot reconstructive surgery). Patients might have three options: General Anesthesia, Spinal Anesthesia, and an ankle block. Using a nominal scale, patients can select their preferred type without any inherent order or ranking. This allows for the collection of qualitative data to understand the distribution of preferences among patients.

Ordinal Scale: To assess the level of satisfaction with the total knee replacement therapy, an ordinal scale can be employed. Patients may be asked to rate their satisfaction on a scale of "Very Dissatisfied," "Dissatisfied," "Neutral," "Satisfied," and "Very Satisfied." The ordinal scale captures qualitative data that provides insight into patients' perceptions of their overall experience with the treatment. This data can help healthcare providers identify areas for improvement and patient satisfaction trends.

2. ***Quantitative data***

Quantitative Data: Quantitative data in the context of total knee replacement therapy involves numerical measurements and objective assessments. For example, researchers may collect data on:

Range of Motion (ROM): Measurement of the degree to which a patient can flex and extend their knee joint post-surgery.

Pain Scores: Numerical ratings from patients assessing their pain levels, often on a scale from 0 to 10, where 0 indicates no pain and 10 is the worst possible pain.

Physical Function Scores: Objective assessments of a patient's ability to perform specific tasks, such as walking, climbing stairs, or getting in and out of a chair.

Quantitative data provides precise, measurable information. In our example- allowing researchers and healthcare providers to quantify the degree of improvement or any potential complications following total knee replacement.

While focusing on the type of data, the distribution of the data is important and crucial in statistical analysis. Data distribution, particularly the Gaussian distribution, plays a pivotal role in analyzing patient outcomes.

Imagine a study that aims to assess the effectiveness of a novel surgical procedure for patients with hip fractures. Researchers collect data on the time it takes for patients to regain mobility- fully weight-bearing- after the surgery. This dataset might exhibit a Gaussian distribution, meaning that the majority of patients' recovery times cluster around a central value, while fewer patients experience significantly faster or slower recoveries.

In this scenario, the Gaussian distribution provides insights into the typical patient's recovery time, represented by the mean. Researchers can calculate the mean recovery time to understand the central tendency, which is vital for establishing an average benchmark for recovery expectations (6). In a Gaussian distribution the mean (average), mode (commonest) and the median (middle) are the same value.

Additionally, the dispersion of data, often quantified through the standard deviation, helps to understand how much the recovery times vary from the mean. A smaller standard deviation indicates that recovery times are more tightly clustered around the mean, suggesting a more predictable outcome for patients. A larger standard deviation, on the other hand, implies greater variability, potentially highlighting factors affecting patient recovery that require further investigation (5, 6, 9).

It is best to discuss with a statistician as to how you should best represent your data.

Selecting a proper statistical test

Null hypothesis:

In the realm of statistical analysis for clinical studies, selecting a proper statistical test is critical to draw meaningful conclusions about the effectiveness of interventions. The foundation of this process often begins with the formulation of a null hypothesis (H0). In the context of a knee replacement surgery, a null hypothesis might be stated as follows: "There is no significant statistical difference in post-operative pain scores between patients who received the traditional knee replacement procedure and those who underwent a minimally invasive knee replacement". This null hypothesis serves as the starting point for statistical analysis and is typically tested against an alternative hypothesis (H1), which suggests that there is a significant statistical difference between the two surgical approaches. The chosen statistical test, would then be applied to the data to determine whether the null hypothesis can be rejected, indicating a meaningful difference, or retained, signifying no significant distinction in post-operative pain scores between the two groups. Accurate hypothesis testing is fundamental in providing evidence for guiding clinical decision-making.

P value

In the context of selecting a proper statistical test, understanding the concept of the p-value is essential when assessing the effectiveness of treatments like total knee replacement therapy. Once a null hypothesis (H0) is formulated, such as "There is no statistically significant difference in post-operative pain scores between patients who received traditional knee replacement and those who underwent minimally invasive knee replacement," the p-value represents a key component in hypothesis testing. The p-value quantifies the strength of evidence against which the null hypothesis is tested. In our example, a low p-value suggests (less than 0.05) that the observed difference in post-operative pain scores between the two surgical approaches is unlikely to have occurred by random chance alone, indicating that the null hypothesis may be rejected. Conversely, a high p-value (Usual cut off level is 0.05) implies that the observed difference could have occurred due to random variability, leading to the retention of the null hypothesis. Proper selection of a statistical test and a clear understanding of the p-value are pivotal in providing statistical evidence for the efficacy of clinical decision-making, and contributing to the advancements in surgical (cliical) practice.

When and what type of statistical test can be used (Figure 1)?

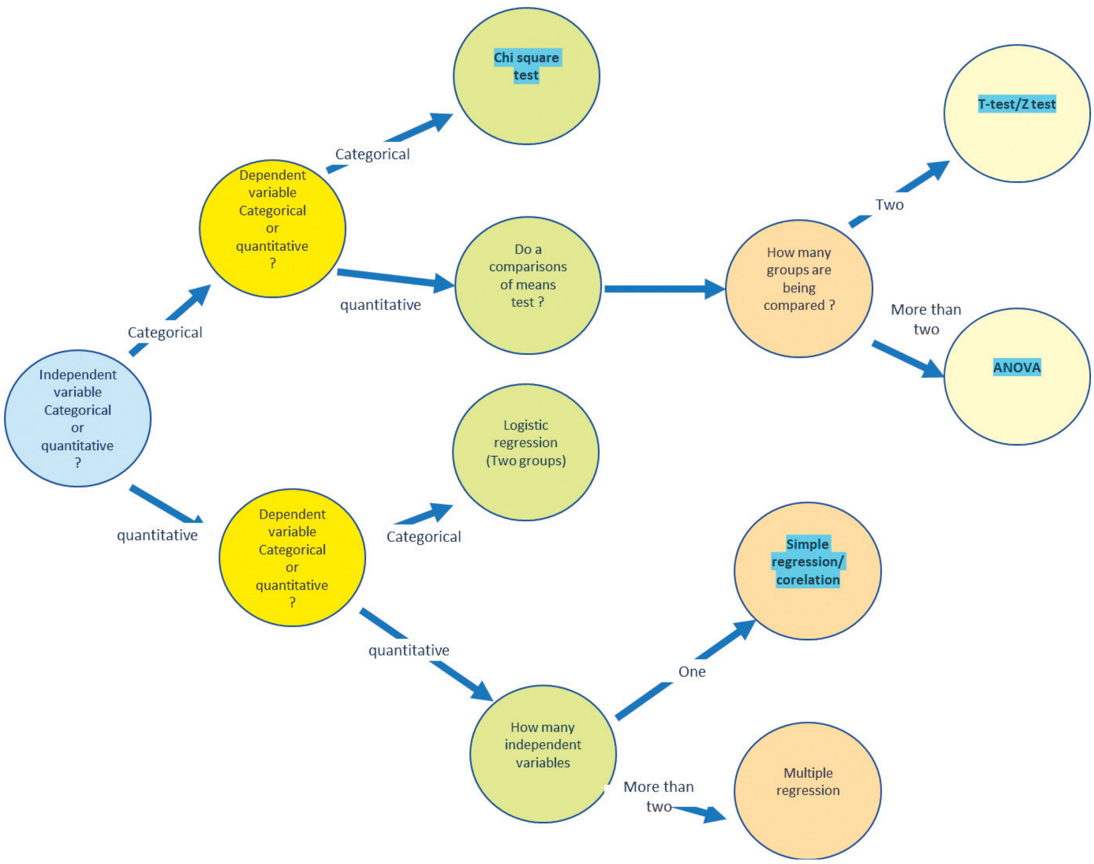

Figure 1- Schematic diagram on how to select most suitable statistical test (4)

- **Chi Square test -** The chi-square test is a valuable tool when investigating the relationship between categorical variables. For example, it can help determine whether there is a significant association between the type of hip implant used (e.g., metal or ceramic) and post-operative complications (such as infections or dislocations). Data is collected from a sample of patients, organized into a contingency table, and hypotheses are formulated. When the chi-square test is applied, it calculates a statistic by comparing observed and expected values. By comparing this statistic to a critical value and assessing the p-value, researchers can interpret the results. If the test shows a significant association, it implies that the choice of implant material significantly influences post-operative complications, providing crucial guidance for orthopaedic surgeons in optimizing hip implant selection for various patient profiles (4, 5, 8, 10).
- **Student T Test -** The Student's t-test serves as a valuable statistical tool in the assessment surgery outcomes. In this scenario, it's employed to discern whether there is a significant distinction in post-operative recovery times between two groups of patients—one with the metal implants and the other with ceramic implants. By collecting data from 100 patients, 50 in each group, and noting the time required for them to regain full mobility post-surgery, the t-test aims to evaluate differences. The null hypothesis posits that there's no substantial variation in recovery time, while the alternative hypothesis suggests otherwise. Upon calculation, a t-statistic is obtained (e.g., 2.5), and its comparison to a critical value from a t-distribution table, along with a p-value assessment, determines the outcome. If the t-statistic surpasses the critical value and the p-value is less than 0.05, the null hypothesis is rejected. This implies a statistically significant difference in post-operative recovery time between patients with metal and ceramic implants, informing orthopaedic surgeons about the expected recovery durations and aiding in more informed implant selection for patients (4-6).
- **ANOVA,** also known as Analysis of Variance, is a statistical method employed to compare the means of three or more groups that are mutually exclusive. Analysis of Variance (ANOVA) is a powerful tool when evaluating the impact of different surgical techniques. In this particular scenario under consideration, ANOVA is applied to investigate whether there is a significant difference in post-operative mobility improvement among patients who have undergone total knee replacement using three distinct surgical approaches: traditional open surgery, minimally invasive surgery, and robot-assisted surgery. Data is collected from 120 patients, with 40 in each surgical group, and the degree of mobility improvement post-surgery is recorded. The null hypothesis (H0) posits no significant difference in mobility improvement between the three surgical techniques, while the alternative hypothesis (H1) suggests at least one surgical approach differs significantly. ANOVA calculates an F-statistic based on the variance in mobility improvement scores among the surgical technique groups, and the comparison of this F-statistic to a critical value, along with a p-value assessment, guides the interpretation. If the F-statistic exceeds the critical value and the p-value is less than 0.05, the null hypothesis is rejected, indicating a statistically significant difference in post-operative mobility improvement among the surgical techniques. In this example this information equips orthopaedic surgeons with insights to select the most effective surgical approach, ultimately enhancing patient mobility and recovery after total knee replacement therapy (4, 6).
- **Regression analysis** is a statistical method that aims to predict the extent of change in one variable when another variable is altered by one unit. Regression analysis plays a crucial role in investigating the intricate relationship between patient-specific variables and outcomes. In any one of the above scenarios, researchers seek to identify which factors significantly influence the time required for patients to regain full mobility post-surgery. They collect data from a cohort of 200 patients, documenting their age, pre-operative health status, body mass index (BMI), and the duration of post-operative recovery. Employing multiple linear regression analysis, researchers model the intricate interplay between these variables while controlling for potential confounding factors. The results

offer valuable insights, potentially revealing that older patients with higher pre-operative health scores tend to experience longer recovery times, while BMI may exert a less pronounced effect. In this scenario such findings provide orthopaedic surgeons with a clearer understanding of which patients may require tailored post-operative care and rehabilitation, ultimately enhancing the overall success of their surgeries(4-6).

Conclusion

Statistical understanding is paramount for assessing the efficacy of different types of surgical intervention. Researchers often compare groups to determine if an intervention leads to significant improvements beyond chance. This entails managing Type 1 and Type 2 errors, where sample size plays a pivotal role in achieving a balance between power and resource efficiency. Various study designs, including RCTs and Prospective Cohort Studies are employed, depending on research questions and ethical considerations. Recognizing the nature of the data, understanding data distribution, and selecting appropriate statistical tests are crucial steps in this process. Properly addressing the null hypothesis and interpreting the p-value are pivotal in determining treatment effectiveness. This statistical literacy is essential for evidence-based decision-making in orthopaedic practice.

References

1. **Mc S, Ramakrishna H**. Knowledge of Surgeons About Statistics: an Online Survey. Indian Journal of Surgery. 2021;84:1-5.
2. **Haddad F.** Statistics and Data Management: A Practical Guide for Orthopaedic Surgeons. Ann R Coll Surg Engl. 93: Copyright © 2011 by the Annals of The Royal College of Surgeons of England.; 2011. p. 329.
3. **Cassidy LD.** Basic concepts of statistical analysis for surgical research. The Journal of surgical research. 2005;128(2):199-206.
4. **Hennekens CH, Buring JE, Mayrent SL.** Epidemiology in Medicine: Little, Brown; 1987.
5. **Kirkwood BR, Sterne JAC.** Essential Medical Statistics: Wiley; 2010.
6. **Swinscow TDV.** Study design and choosing a statistical test. Campbell MJ, editor: University of Southampton; 2021. 215-26 p.
7. **Ceresoli M, Abu-Zidan FM, Staudenmayer KL, Catena F, Coccolini F.** Statistics and Research Methods for Acute Care and General Surgeons: Springer International Publishing; 2022.
8. **Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M.** Selection of appropriate statistical methods for data analysis. Annals of cardiac anaesthesia. 2019;22(3):297-301.
9. **Ranganathan P.** An Introduction to Statistics: Choosing the Correct Statistical Test. Indian journal of critical care medicine : peer-reviewed, official publication of Indian Society of Critical Care Medicine. 2021;25(Suppl 2):S184-s6.
10. **Nayak BK, Hazra A.** How to choose the right statistical test? Indian journal of ophthalmology. 2011;59(2):85-6.