Paper AS15

More Than a p-Value: Understanding the Differences Between Clinical and Statistical Significance

Elnaz Ghadimi, Cytel Inc., Cambridge, USA
Jasperlynn Kao, Cytel Inc., Cambridge, USA

ABSTRACT

It is possible for a result to be statistically significant but not clinically significant. Conversely, a result can be clinically significant but not statistically significant. Both statistical and clinical significance are crucial for a comprehensive understanding of research findings. Statistical significance helps determine whether an effect exists, while clinical significance evaluates whether the effect is meaningful in a practical sense. To illustrate these concepts, we will explore two real-world examples: one demonstrating a clinically significant finding and another highlighting a statistically significant result, to better understand how these measures impact research and practice.

INTRODUCTION

In clinical research, the interpretation of results often hinges on one value representing one word: “significance”. The problem, however, is that there is much more that lies below the surface of this singular value and this singular term that should be taken into consideration when drawing research conclusions. In particular, there are two pivotal concepts that we will focus on in this paper: statistical significance and clinical significance. While these terms are related, they are not synonymous, and understanding the distinction between them is vital for making informed decisions in healthcare. This paper aims to dissect these concepts through illustrative examples, underscoring their implications for research and practice.

STATISTICAL SIGNIFICANCE

Statistical significance refers to whether the observed effect or relationship in a study is unlikely to have occurred due to chance. This is commonly assessed using a p-value, which represents the probability of obtaining the observed results, or more extreme results, if the null hypothesis (which is no effect) is true. This is typically assessed using a p-value, where a p-value less than a predetermined threshold (commonly 0.05) suggests that the null hypothesis can be rejected. However, statistical significance does not address the magnitude or practical importance of the effect.

ASSESSING STATISTICAL SIGNIFICANCE

Statistical significance is assessed through methods like p-values, confidence intervals, and hypothesis tests (e.g., t-tests, chi-square tests, ANOVA), which determine whether an observed effect or relationship is likely to be due to chance. A result is considered statistically significant if the p-value is below a predefined threshold (usually 0.05), indicating that the null hypothesis can be rejected.

CLINICAL SIGNIFICANCE: DEFINITION AND IMPLICATIONS

Clinical Significance pertains to whether the size of the effect or difference observed is substantial enough to have real-world implications in a clinical setting. This concept focuses on the practical importance and potential impact on patient care and outcomes. It addresses whether the observed effect translates into a tangible benefit for patients, such as improved quality of life, reduced morbidity, or increased survival. Clinical significance often involves subjective judgment and depends on the context of the research and the population studied.

ASSESSING CLINICAL SIGNIFICANCE

Clinical significance is assessed based on factors like the magnitude of the effect (e.g., a clinically meaningful improvement in symptoms), the potential for patient benefit through use of clinical benchmarks or Minimum Clinically Important Difference (MCID), and real-world applicability. Unlike statistical significance, which focuses on whether an effect is unlikely to have occurred by chance, clinical significance considers whether the effect is large enough to make a meaningful impact on patient health, well-being, or quality of life. It often involves clinical judgment, expert opinion, and context-specific factors.

STATISTICAL SIGNIFICANCE VS. CLINICAL SIGNIFICANCE

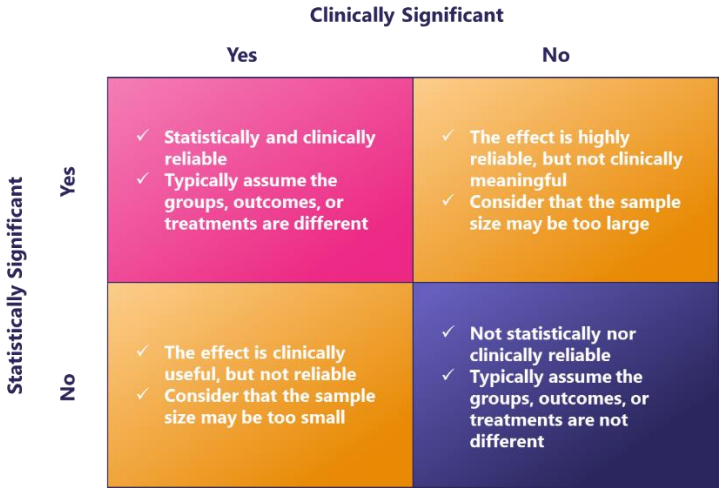

Figure 1. Statistical Significance vs Clinical Significance (adapted from Krousel-Wood, Marie, et al.)

The relationship between statistical and clinical significance can lead to four possible outcomes (***Figure 1***). When a result is both statistically and clinically significant, it means the effect is unlikely due to chance (low p-value) and is meaningful in a real-world context, such as a drug significantly reducing mortality. A result that is statistically significant but not clinically significant suggests that while the effect is unlikely due to chance, its magnitude is too small to be meaningful, such as a cholesterol-lowering drug showing a statistically significant reduction of LDL cholesterol by 2 mg/dL compared to placebo. In this case, the large sample size may have detected a very small difference that lacks practical impact on patient health. Conversely, a result that is clinically significant but not statistically significant indicates a meaningful effect that may not reach statistical significance due to small sample size or variability, such as a promising trend in patient survival that lacks sufficient power. Finally, when a result is neither statistically nor clinically significant, it suggests that any observed effect is likely due to chance and lacks practical importance, meaning further investigation may not be warranted. Understanding these distinctions is crucial for interpreting study findings accurately.

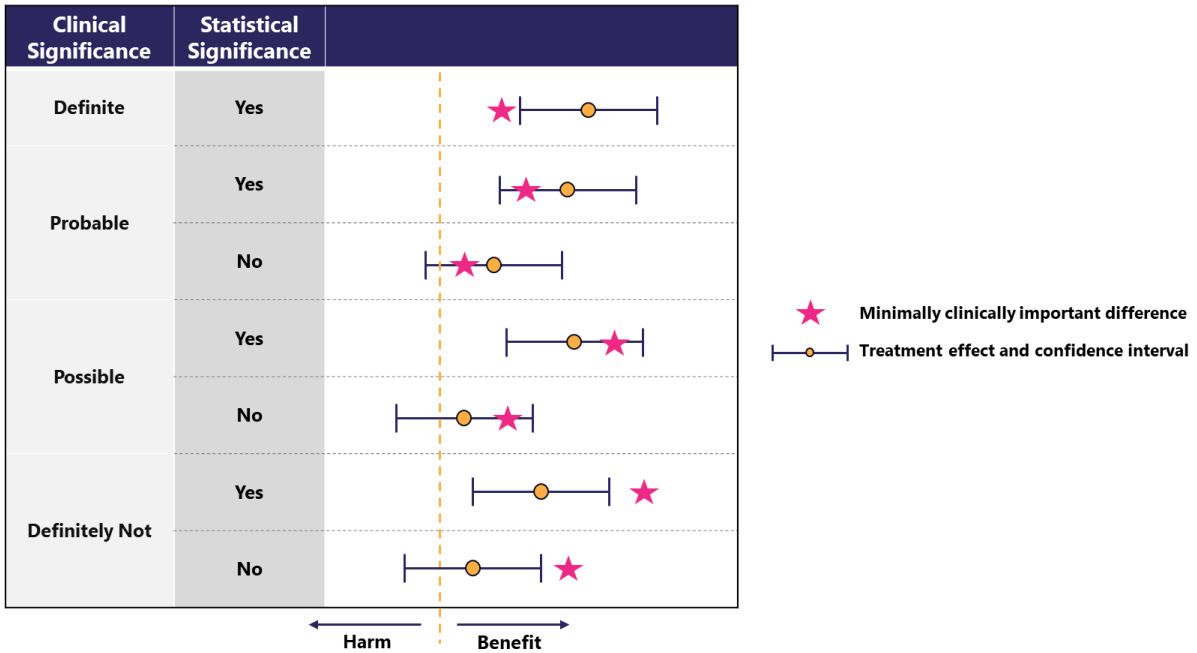

Figure 2. Relationship between clinical significance and statistical significance (adapted from Howard, A., Fuchsia et al.)

Clinical significance can be assessed using the relationship between the MCID of the treatment effect and the Confidence Interval (CI). The MCID represents the smallest change in the treatment effect that would be considered meaningful from a clinical perspective, while the CI provides a range of values within which the true treatment effect is likely to fall, based on the data. By comparing the MCID with the CI, clinical significance is categorized into four levels (***Figure. 2***):

1. **Definite** – This level indicates that the MCID is smaller than the lower limit of the CI of the treatment effect. In this case, the treatment effect is robust enough to exceed the threshold of clinical importance, and it is highly likely that the treatment is effective.
2. **Probable** – Here, the MCID is greater than the lower limit of the CI but still smaller than the treatment effect. This suggests that while the treatment shows promise, there is some uncertainty, and the effect might not be consistently meaningful across all patients or conditions.
3. **Possible** – In this category, the MCID is less than the upper limit of the CI but greater than the treatment effect. This indicates that although there may be some clinical benefit, it is not conclusively significant, and the treatment may not achieve a clinically meaningful difference in all cases.
4. **Definitely Not** – The MCID is greater than the upper limit of the CI of the treatment effect. This suggests that the treatment effect is unlikely to be clinically meaningful, and the treatment is considered ineffective or irrelevant for most patients.

This approach helps ensure that findings are not only statistically valid but also practically relevant to patient care.

CASE STUDY 1: HYPERTENSION STUDY - EXAMPLE OF STATISTICAL SIGNIFICANCE WITHOUT CLINICAL SIGNIFICANCE

Suppose a clinical trial evaluates a new antihypertensive drug aimed at reducing blood pressure in individuals with mild hypertension. The study compares the treatment group to a placebo group over 12 weeks and finds a statistically significant reduction in systolic blood pressure (SBP). The treatment group shows a decrease of 2 mmHg in SBP, compared to the placebo group, with a p-value of 0.01, which is below the commonly accepted threshold of 0.05, indicating statistical significance.

Although this result is statistically significant, the 2 mmHg reduction in SBP is small and unlikely to have a meaningful impact on patient outcomes. The MCID for blood pressure reduction is typically considered to be around 5-10 mmHg, with reductions of this magnitude having a proven impact on reducing the risk of cardiovascular events like stroke and heart attack. A reduction of only 2 mmHg does not meet this threshold and may not provide any meaningful clinical benefit in terms of reducing long-term health risks.

Therefore, while the treatment shows statistical significance, the actual reduction in blood pressure is too small to be considered clinically relevant, making the result clinically insignificant.

CASE STUDY 2: DIABETES STUDY - EXAMPLE OF CLINICAL SIGNIFICANCE WITHOUT STATISTICAL SIGNIFICANCE

Consider a study assessing the impact of a new SGLT2 inhibitor on change in hemoglobin A1c (HbA1c) levels from baseline to week 12 for 40 subjects. The treatment group shows a reduction in HbA1c of 0.6% compared to the placebo group showing a reduction in HbA1c of 0.3%, with a p-value of 0.12.

The MCID in HbA1c reduction is generally considered to be around 0.3% to 0.5%.

- ≥0.3% HbA1c reduction is often cited as the smallest meaningful change that may have some clinical impact.
- ≥0.5% HbA1c reduction is typically used in clinical trials and regulatory assessments (e.g., FDA and EMA) as a threshold for a clinically significant improvement in glycemic control.

Despite not achieving statistical significance for HbA1c reduction in this study, a 0.6% decrease in just 12 weeks is considered clinically meaningful, as even small reductions can lower long-term complications, particularly in patients with poorly controlled diabetes.

KEY CONSIDERATIONS

To avoid misalignment between statistical and clinical significance, researchers should consider the following strategies when conducting a study and interpreting the results:

1. **Define Clinically Meaningful Thresholds**: Establish thresholds for what constitutes a meaningful effect size before conducting the study. This ensures that statistical analyses are interpreted in the context of practical relevance.
2. **Ensure Adequate Study Power**: Design studies with sufficient sample sizes to detect clinically meaningful effects. Power calculations should be based on the anticipated effect size that is clinically significant.

3. **Employ Robust Study Designs**: Minimize variability and biases by using rigorous methodologies. High-quality designs reduce the likelihood of failing to detect meaningful effects or overemphasizing trivial ones.
4. **Use Comprehensive Reporting**: Present both statistical and clinical significance transparently. Include effect sizes, confidence intervals, and a discussion of clinical implications alongside p-values to provide a fuller picture of the findings.
5. **Contextualize Results:** Interpret findings within the broader context of the disease, treatment goals, and patient population. Engage clinical experts to assess the practical implications of statistically significant results.

CONCLUSION

Both statistical and clinical significance are essential for a comprehensive understanding of research findings. Statistical significance provides confidence in the existence of an effect, while clinical significance evaluates its practical impact. Researchers and clinicians must consider both measures to draw balanced and actionable conclusions. By defining clinically meaningful thresholds, ensuring adequate study power, employing robust designs, and contextualizing results, we can better align statistical and clinical significance. These practices ensure that research findings translate effectively into improved patient care.

REFERENCES (HEADER 1)

Howard, A. F., Fuchsia, G., Goddard, K., Samargandi, O., & Hasan, H. (2018). Clinical significance in pediatric oncology randomized controlled treatment trials: A systematic review. Trials, 19, Article 2925. https://doi.org/10.1186/s13063-018-2925-8.

Krousel-Wood, Marie & Chambers, Richard & Muntner, Paul. (2007). Clinicians' Guide to Statistics for Medical Practice and Research: Part II. The Ochsner journal. 7. 3-7. https://pmc.ncbi.nlm.nih.gov/articles/PMC3096336/.

Mehta, V. (2017). Statistical significance vs. clinical significance. SlideShare. https://www.slideshare.net/vinip3012/statistical-significance-vs-clinical-significance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Elnaz Ghadimi
Senior Biostatistician
Elnaz.ghadimi@cytel.com

Jasperlynn Kao
Associate Director, Biostatistics
Jasperlynn.Kao@cytel.com

Cytel, Inc.
675 Massachusetts Ave
Cambridge, MA 02139
www.cytel.com