Beyond P-Values: Unpacking the Clinical Relevance of Statistical Significance

Usman Hider

February 24, 2024

Beyond P-Values: Unpacking the Clinical Relevance of Statistical Significance

Usman Hider

Department of Health Science, University of Malashiya, Asia

Abstract:

In contemporary scientific research, statistical significance, often indicated by p-values, has been a cornerstone in determining the validity of findings. However, reliance solely on p-values to infer clinical relevance may lead to misinterpretation and inappropriate decision-making. This paper seeks to explore the limitations of p-values in conveying the significance of research findings in clinical contexts and advocates for a comprehensive approach that considers both statistical significance and clinical relevance. Firstly, we elucidate the concept of statistical significance and its role in hypothesis testing. While p-values provide a measure of the strength of evidence against a null hypothesis, they do not inherently indicate the magnitude or practical importance of an effect. Consequently, a statistically significant result may not necessarily translate to clinical significance. We then discuss the importance of effect size estimation and confidence intervals in assessing the practical relevance of study findings. Effect size measures provide quantitative estimates of the magnitude of an effect, facilitating meaningful interpretation in clinical practice. In conclusion, while statistical significance remains essential in scientific inquiry, its interpretation should be complemented by a thorough assessment of clinical relevance. Moving beyond p-values entails embracing a holistic approach that incorporates effect size estimation, confidence intervals, and contextual considerations, thereby fostering more robust and meaningful evidence-based practice in clinical research.

Keywords: statistical significance, p-values, clinical relevance, effect size, confidence intervals, hypothesis testing, evidence-based practice, research interpretation, healthcare decision-making

Introduction

Statistical significance is a fundamental concept in scientific research, particularly in the field of hypothesis testing. It serves as a measure of the reliability and credibility of research findings, indicating whether the observed results are likely to have occurred by chance or if they reflect a

true effect. At its core, statistical significance is concerned with assessing the probability of obtaining results as extreme as, or more extreme than, the observed data under the assumption that the null hypothesis is true. The null hypothesis, denoted as H0, represents the default assumption or status quo in a hypothesis test. It posits that there is no significant difference or effect present in the population under study. Conversely, the alternative hypothesis, denoted as Ha, proposes the existence of a significant effect, deviation from the null hypothesis, or relationship between variables. Hypothesis testing involves collecting data from a sample and using statistical methods to determine whether the evidence supports rejecting the null hypothesis in favor of the alternative hypothesis. This process typically revolves around calculating a test statistic, such as a t-test, z-test, or chi-square test, which quantifies the discrepancy between the observed data and what would be expected under the null hypothesis [1], [2].

The significance level, often denoted as α (alpha), is predetermined by researchers and represents the threshold for rejecting the null hypothesis. Commonly used values for α include 0.05 or 0.01, indicating a 5% or 1% chance, respectively, of falsely rejecting the null hypothesis when it is actually true. If the calculated p-value (probability value) associated with the test statistic is less than or equal to the chosen significance level, the results are deemed statistically significant, leading to rejection of the null hypothesis. Interpreting the p-value requires careful consideration of its meaning. Specifically, the p-value represents the probability of obtaining results as extreme as, or more extreme than, the observed data, assuming that the null hypothesis is true. A small p-value suggests strong evidence against the null hypothesis, indicating that the observed results are unlikely to have occurred by chance alone. Conversely, a large p-value indicates weak evidence against the null hypothesis, implying that the observed results are consistent with random variation or sampling variability. However, it is crucial to recognize that statistical significance does not necessarily equate to practical or clinical significance. While a statistically significant result indicates that an effect exists, it does not provide information about the magnitude or importance of that effect in real-world terms. Additionally, statistical significance can be influenced by factors such as sample size, study design, and variability within the data [3].

Limitations of Relying Solely on P-values

While p-values serve as a widely accepted metric for determining statistical significance, relying solely on them to infer clinical relevance can be problematic for several reasons.

1. **Lack of Effect Size Information**: P-values only indicate whether an observed effect is statistically significant or not; they do not provide information about the size or magnitude of the effect. Consequently, a small p-value may be associated with a statistically significant but trivial effect size, which may have little or no practical significance in clinical practice.

2. **Sensitivity to Sample Size**: P-values are highly sensitive to sample size, meaning that larger sample sizes can result in smaller p-values even for effects that are not clinically meaningful. This sensitivity can lead to statistically significant results that have little practical relevance due to the influence of large sample sizes [4].

3. **Multiple Comparisons**: In studies where, multiple hypotheses are tested simultaneously, conducting numerous statistical tests increases the likelihood of obtaining false-positive results (Type I errors). Although adjusting the significance threshold (e.g., Bonferroni correction) can mitigate this issue, it may also increase the risk of false negatives (Type II errors) and overlook clinically relevant findings.

4. **Publication Bias**: There is a tendency in academic publishing to favor the publication of studies with statistically significant results, leading to a potential bias in the literature known as publication bias. This bias can distort the perception of the true prevalence or magnitude of effects, as studies with non-significant findings may remain unpublished or overlooked.

5. **Contextual Factors Ignored**: P-values do not account for important contextual factors that may influence the clinical relevance of study findings, such as patient demographics, disease severity, treatment adherence, and practical considerations like cost-effectiveness and patient preferences. Ignoring these factors can result in misinterpretation of statistical significance and inappropriate application of research findings in clinical settings [5].

Effect Size Estimation and Confidence Intervals

Effect size estimation and confidence intervals play crucial roles in evaluating the practical significance of research findings, offering valuable insights beyond mere statistical significance.

1. **Effect Size Estimation**: Effect size measures quantify the magnitude of an observed effect, providing information about the practical importance of the findings. Unlike p-values, which focus solely on statistical significance, effect size estimation enables researchers and

practitioners to gauge the strength and relevance of an effect in real-world terms. Common effect size measures include Cohen's d for comparing means, odds ratios for assessing associations, and correlation coefficients for measuring relationships between variables. By considering effect size, researchers can prioritize interventions or treatments with the greatest potential impact on clinical outcomes, enhancing evidence-based decision-making in healthcare settings [6].

2. **Confidence Intervals (CIs)**: Confidence intervals provide a range of plausible values for the true effect size, conveying both the precision and uncertainty associated with study results. Unlike point estimates, which offer only a single value, confidence intervals offer a more comprehensive understanding of the variability inherent in the data. Wider confidence intervals indicate greater uncertainty, while narrower intervals suggest more precise estimates. By incorporating confidence intervals into data interpretation, researchers can better assess the reliability and stability of their findings, facilitating robust conclusions about the practical significance of observed effects.

3. **Interpretation of Clinical Relevance**: Effect size estimation and confidence intervals complement each other in the assessment of practical significance. A statistically significant result with a large effect size and narrow confidence intervals provides compelling evidence of both statistical and clinical significance, indicating a substantial and reliable impact on clinical outcomes. Conversely, a statistically significant result with a small effect size and wide confidence intervals may have limited practical relevance, warranting cautious interpretation and further investigation. By examining effect size estimates alongside confidence intervals, researchers can discern meaningful patterns or trends in the data, informing the development of targeted interventions or treatment strategies tailored to meet the needs of patients [7], [8].

Evaluating Clinical Relevance

The evaluation of clinical relevance goes beyond statistical measures and requires a comprehensive consideration of contextual factors that influence the applicability and impact of research findings in real-world settings.

1. **Patient Demographics and Characteristics**: Understanding the demographic characteristics of the patient population under study is essential for assessing the generalizability of research

findings to diverse patient groups. Factors such as age, gender, ethnicity, socioeconomic status, and comorbidities can influence treatment outcomes and response to interventions. Researchers must account for these demographic variables to ensure that study results are relevant and applicable across different patient populations.

2. **Disease Severity and Clinical Context**: The severity and stage of the disease or condition being studied can significantly impact the clinical relevance of research findings. Interventions that demonstrate efficacy in mild or early-stage cases may not necessarily translate to similar benefits in severe or advanced stages of the disease. Moreover, the clinical context, including standard treatment protocols, available resources, and healthcare infrastructure, must be considered when interpreting study results and determining their relevance to clinical practice.

3. **Treatment Adherence and Patient Preferences**: Factors related to treatment adherence and patient preferences play a crucial role in determining the practical significance of research findings. Interventions that are effective in controlled research settings may not be feasible or acceptable to patients in real-world practice. Considering patient preferences, values, and beliefs is essential for developing patient-centered treatment plans that align with individual needs and priorities.

4. **Cost-Effectiveness and Resource Allocation**: Assessing the cost-effectiveness of interventions is vital for informing healthcare decision-making and resource allocation. Even if an intervention demonstrates statistical and clinical efficacy, its implementation may be limited by cost considerations, especially in resource-constrained healthcare settings. Researchers and policymakers must weigh the costs and benefits of interventions to optimize healthcare delivery and maximize patient outcomes [9], [10].

5. **Potential Risks and Harms**: Evaluating the clinical relevance of research findings requires careful consideration of potential risks and harms associated with interventions. While an intervention may demonstrate efficacy in improving clinical outcomes, it may also pose risks or adverse effects that outweigh its benefits. Researchers must conduct thorough risk-benefit analyses to ensure that interventions are safe and appropriate for use in clinical practice.

Decision-Making in Healthcare Settings

In healthcare settings, informed decision-making relies on a combination of statistical analyses and clinical judgment to ensure that interventions are evidence-based, contextually relevant, and aligned with patient needs and preferences [11].

Complementary Nature of Statistical Analyses and Clinical Judgment: Statistical analyses provide valuable insights into the quantitative aspects of research findings, including measures of effect size, significance levels, and confidence intervals. However, statistical analyses alone may not capture the complexity and nuances of clinical practice. Clinical judgment, informed by years of professional experience and expertise, allows healthcare practitioners to interpret statistical findings within the broader context of patient care, considering factors such as clinical history, comorbidities, and individual patient preferences.

Integration of Quantitative and Qualitative Evidence: A holistic approach to decision-making involves integrating both quantitative evidence from statistical analyses and qualitative evidence derived from clinical observations, patient narratives, and shared decision-making processes. By combining these sources of information, healthcare providers can develop a comprehensive understanding of the benefits, risks, and potential outcomes associated with different treatment options, facilitating patient-centered care and personalized treatment plans [12].

Shared Decision-Making with Patients: Informed decision-making in healthcare goes beyond the expertise of healthcare providers and encompasses active involvement and collaboration with patients in the decision-making process. Shared decision-making recognizes patients as partners in their care, empowering them to make informed choices based on their values, preferences, and goals. Healthcare providers play a critical role in facilitating shared decision-making by presenting evidence-based information in a clear and understandable manner, addressing patient concerns, and guiding patients through the decision-making process [13], [14].

Continuous Monitoring and Evaluation: Informed decision-making in healthcare is an iterative process that requires continuous monitoring and evaluation of treatment outcomes, patient responses, and emerging evidence. Healthcare providers must remain vigilant and adaptable, adjusting treatment plans as needed based on new information, changing patient circumstances, and evolving best practices. By staying abreast of the latest research findings and clinical guidelines, healthcare providers can optimize patient care and outcomes over time.

Ethical Considerations and Patient Safety: In all decision-making processes, healthcare providers must uphold ethical principles and prioritize patient safety and well-being. This includes ensuring that interventions are evidence-based, clinically appropriate, and aligned with the principles of beneficence, non-maleficence, and respect for patient autonomy. Ethical decision-making requires careful consideration of the potential risks and benefits of interventions, as well as the values and preferences of individual patients [15].

Conclusion

In conclusion, the process of evaluating research findings and making informed decisions in healthcare settings is multifaceted and requires a balanced approach that integrates statistical analyses with clinical judgment, patient preferences, and ethical considerations. While statistical significance serves as an important metric for assessing the reliability of research findings, it must be interpreted in conjunction with effect size estimation, confidence intervals, and contextual factors to determine clinical relevance. Moreover, shared decision-making with patients plays a pivotal role in ensuring that treatment plans are aligned with individual preferences, values, and goals. A holistic approach to decision-making in healthcare emphasizes the importance of continuous monitoring and evaluation, ethical principles, and responsiveness to evolving evidence and patient needs. By combining quantitative evidence with qualitative insights and engaging patients as partners in their care, healthcare providers can deliver personalized, evidence-based interventions that optimize patient outcomes and promote patient-centered care. Ultimately, informed decision-making in healthcare is not solely about statistical significance or clinical judgment, but rather about striking a delicate balance between evidence-based practice, patient-centered care, and ethical considerations. By embracing a holistic approach that integrates multiple sources of information and actively involves patients in the decision-making process, healthcare providers can enhance the quality, safety, and effectiveness of patient care, fostering better health outcomes and improved patient satisfaction.

References

[1] Bleakley, C. M., Matthews, M., & Smoliga, J. M. (2021). Most ankle sprain research is either false or clinically unimportant: A 30-year audit of randomized controlled trials. *Journal of Sport and Health Science*, *10*(5), 523-529.

[2] Heston, T. F. (2023). Statistical Significance Versus Clinical Relevance: A Head-to-Head Comparison of the Fragility Index and Relative Risk Index. *Cureus*, *15*(10).

[3] Heston, T. F. (2023). Safety of large language models in addressing depression. *Cureus*, *15*(12).

[4] Heston, T. F. (2023). The robustness index: going beyond statistical significance by quantifying fragility. *Cureus*, *15*(8).

[5] Matthews, R. A. (2018). Beyond 'significance': principles and practice of the analysis of credibility. *Royal Society Open Science*, *5*(1), 171047.

[6] Greenland, S. (2019). Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values. *The American Statistician*, *73*(sup1), 106-114.

[7] Cristea, I. A., & Ioannidis, J. P. (2018). P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PloS one*, *13*(5), e0197440.

[8] Heston T F (October 26, 2023) Statistical Significance Versus Clinical Relevance: A Head-to-Head Comparison of the Fragility Index and Relative Risk Index. Cureus 15(10): e47741. doi:10.7759/cureus.47741

[9] Li-wei, H. L., Nemati, S., Moody, G. B., Heldt, T., & Mark, R. G. (2014, September). Uncovering clinical significance of vital sign dynamics in critical care. In *Computing in Cardiology 2014* (pp. 1141-1144). IEEE.

[10] Matthews, R. (2021). The p-value statement, five years on. *Significance*, *18*(2), 16-19.

[11] Vakhitova, Z. I., & Alston-Knox, C. L. (2018). Non-significant p-values? Strategies to understand and better determine the importance of effects and interactions in logistic regression. *PloS one*, *13*(11), e0205076.

[12] Heston T F (December 18, 2023) Safety of Large Language Models in Addressing Depression. Cureus 15(12): e50729. doi:10.7759/cureus.50729

[13] Heston T F (August 30, 2023) The Robustness Index: Going Beyond Statistical Significance by Quantifying Fragility. Cureus 15(8): e44397. doi:10.7759/cureus.44397

[14] Lovell, D. P. (2020). Null hypothesis significance testing and effect sizes: Can we 'effect' everything… or… anything?. *Current opinion in pharmacology*, *51*, 68-77.

[15] Matthews, R. A. (2019). Moving towards the post p< 0.05 era via the analysis of credibility. *The American Statistician*, *73*(sup1), 202-212.