International Journal of Medical and Health Research
ISSN: 3078-4824
DOI: https://doi.org/10.61424/ijmhr.v4i3.1146
Date Published Online: October 28, 2026
Journal Homepage: https://bluemarkpublishers.com/index.php/IJMHR

IJMHR
BLUEMARK PUBLISHERS

| RESEARCH ARTICLE

The Importance and Application of P-Value and Confidence Interval in Advanced Health Sciences Research

Muhammad Ilyas
PhD Nursing (Scholar), Lincoln University College, Petaling Jaya, Malaysia
Corresponding Author: Muhammad Ilyas, **E-mail**: ilyas.phdscholar@lincoln.edu.my

| ABSTRACT

P value has been the most commonly employed statistical concept in health sciences research to indicate relationship between exposure and outcomes, despite the availability of more descriptive statistics like risk ratios and confidence intervals (CIs). P-value was first introduced by Ronald Fisher in his 1925 "Statistical Methods for Research Workers", suggested that P = 0.05 be used as a flexible measure of statistical inference rather than a strict decision threshold when determining whether a deviation should be considered significant. CIs, which offer a range of reasonable values for the population parameter instead of a straightforward yes-or-no choice, evolved as a complementary approach. Both of these concepts together serve as the foundation for statistical inference in epidemiological and biomedical research, yet they have both been misunderstood and abused for many years. The study aims to evaluate and assess the importance and application of P-value and CI in advanced health sciences research. This review article systematically examines and synthesizes the existing literature on the importance and application of P-valve and CI in advanced health sciences research. The choice between P-value and CI should depend on context. P-values suit testing a specific null hypothesis, while CIs better describe the precision of an estimate in observational settings. CIs provide more valuable information than P-values alone because they offer the direction and size of the effect rather than just a binary significant/non-significant conclusion. It is advised to report both P-values and confidence intervals since they provide complimentary information. While CIs display the range of likely effect sizes, P-values reveal the distance from a null-hypothesized value.

| KEYWORDS

Importance, application, p-value, confidence interval, health sciences, research.

| ARTICLE INFORMATION

ACCEPTED: September 17, 2026 **PUBLISHED:** October 28, 2026 **DOI:** https://doi.org/10.61424/ijmhr.v4i3.1146

1. Introduction

P-value and Confidence Interval (CI) are the foundational ideas of quantitative research, especially in the health sciences, epidemiology, and clinical trials. They give researchers a framework for figuring out whether observed associations, differences, or treatment effects are more likely to represent real phenomena than chance (Altman N, & Krzywinski M., 2024). Researchers create an alternative hypothesis (H_1), which proposes the existence of an effect or difference, and a null hypothesis (H_0), which assumes no effect or difference, within the framework of Null Hypothesis Significance Testing (NHST). The evidence against the null hypothesis is then evaluated using statistical tests (Bonovas, S., & Piovani, D., 2023).

Copyright: © 2026 The Author(s). This open access article is distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC- BY-NC-SA) License (https://creativecommons.org/licenses/by-nc-sa/4.0/), published by Bluemark Publishers.

The probability of getting outcomes as extreme as or more extreme than those found in a study if the null hypothesis is true is known as the p-value. The observed results are less consistent with the null hypothesis when the p-value is less (Holmberg C., 2024). A limit of $p < 0.05$ has historically been used to announce statistical significance; however, new research highlights that this cutoff should not be taken as a tight boundary between "important" and "unimportant" findings (Altman N, & Krzywinski M., 2024).

The abuse and misunderstanding of p-values has been brought to light by recent methodological discussions. Frequently held fallacies include assuming that statistical significance inevitably implies practical or therapeutic importance or interpreting the p-value as the likelihood that the null hypothesis is true (Patel, S. & Green, A., 2025). While non-significant findings may still be therapeutically useful, particularly when sample sizes are small, modern scholars contend that statistically significant results may have little practical significance if the impact size is tiny (Ahmed, E. S., & Butt, M. N., 2025).

According to the current statistical guidelines, p-values should be interpreted in conjunction with effect sizes, confidence intervals, and subject-matter expertise. In order to enhance study interpretation and decision-making, a number of recent evaluations further recommend supplementing conventional significance tests using Bayesian methods and clinical relevance metrics (Holmberg C., 2024). In the health sciences, where evidence-based decisions should consider both statistical evidence and practical implications for patient care, this approach is especially crucial (Ahmed, E. S., & Butt, M. N., 2025).

In general, although p-values are still frequently employed in biomedical and health sciences research, new research urges researchers to adopt a more thorough interpretation of study results that considers effect magnitude, uncertainty, and clinical relevance rather than relying only on statistical significance (Holmberg C., 2024 & Ahmed, E. S., & Butt, M. N., 2025).

2. What is P-valve

The probability of data this extreme if the null hypothesis is correct is known as a p-value. In hypothesis testing, a p-value is a number between 0 and 1 that indicates how well the observed data match a null hypothesis, which is typically a hypothesis of no effect or no difference (Chén, O. Y. et al., 2023 & Bonovas, S., & Piovani, D., 2023). More precisely, it is the likelihood that, under a given statistical model, the observed result or one more extreme would occur if that null model were true (Mansournia, M. et al., 2022 & Bonovas, S., & Piovani, D., 2023).

2.1 Core Meaning of P-value

A small p-value is seen as stronger evidence against the null hypothesis since it indicates that the observed data would be somewhat exceptional if the null hypothesis were true (Thiese, M., 2016). Although it does not prove the null hypothesis is correct, a large p-value indicates that the data are more consistent with the null model (Robinson, R. & Haviland, J., 2021 & Mansournia, M. et al., 2022).

According to a number of studies, the standard cutoff $p < 0.05$ is conventional rather than sacrosanct, and it is frequently applied excessively strictly (Kwak, S., 2023 & Bonovas, S., & Piovani, D., 2023). P-values are helpful, but only when understood in relation to study design, effect size, assumptions, and reproducibility, according to evidence found in numerous medical and statistical publications (Chén, O. Y. et al., 2023).

2.2 Common Misinterpretations

*The probability that the null hypothesis is true is not represented by a p-value (Ioannidis, J., 2018).
*The magnitude or significance of an effect cannot be determined by a p-value (Robinson, R., & Haviland, J., 2021).
*A non-significant p-value may indicate poor power or small samples rather than no effect (Robinson, R., & Haviland, J., 2021).
*Even in cases when the effect is negligible, P-values may decrease as sample numbers increase (Tsushima, E., 2022).

2.3 Why P-value is Controversial

P-values are frequently utilized in biomedical research; however, they are also frequently misinterpreted and overemphasized, according to numerous studies (Ioannidis, J., 2018 & Kafadar, K., 2021). Additionally, the research associate's publication bias, selective reporting, and low reproducibility with strict significance-threshold thinking (Ioannidis, J., 2018 & Robinson, R., & Haviland, J., 2021).

While some authors advocate switching from binary "significant/not significant" thinking to confidence intervals, effect sizes, Bayesian methods, or compatibility interpretations, others argue that the default threshold should be lowered from 0.05 to 0.005 (Ioannidis, J., 2018 & Mansournia, M. et al., 2022). Simply, a p-value is not a direct probability that the hypothesis is correct or incorrect, but rather a measure of how unexpected the findings are under a null hypothesis. According to the recent literature, p-values are still valuable, but they should never be used in isolation (Hamasaki, T. et al., 2021 & Lohse, K. R., & Kliethermes, S. A., 2025).

According to current research in the health sciences, statistical significance and p-values are still significant, but mostly as measures of how incompatible data are with a null hypothesis rather than as independent evidence of truth, effect size, or clinical significance (Mansournia, M., & Nazemipour, M. 2024). The key takeaway from recent reviews, commentary, and reporting guidelines is to pair exact p-values with effect sizes, confidence intervals, research design, and clinical relevance in order to go beyond the $p<0.05$ dichotomy (Bujang, M., 2025 & Ferreira, M. H. et al., 2026).

P-values show how incompatible the data are with a certain statistical model, typically the null hypothesis; they do not represent effect size, clinical significance, or the likelihood that a hypothesis is true (Sharma, H., 2021). Because it oversimplifies uncertainty, encourages false dichotomies, and can skew interpretation and reproducibility in biomedical research, the use of $p<0.05$ as a binary decision criterion is frequently criticized (Boscardin, C. K. et al., 2024 & Ferreira, M. H. et al., 2026). According to Mansournia, M., & Nazemipour, M. 2024 & Boscardin, C. K. et al., 2024, it is best practice to provide precise p-values together with effect sizes and confidence intervals and to interpret results in light of clinical significance, precision, and study limitations.

P-values are unable to provide answers to the questions that clinicians are most concerned about. They are unable to determine whether a treatment's benefits outweigh its drawbacks, whether an effect is clinically significant, or whether a non-significant result indicates no effect (Ciapponi, A. et al., 2021 & Mansournia, M., & Nazemipour, M. 2024). In the health sciences, this distinction is crucial because, although clinically significant effects may not reach conventional significance in small or heterogeneous studies, minor effects may become statistically significant in large samples (Sharma, H., 2021 & Patel, S., & Green, A., 2025).

Because it provides a standardized assessment of how surprising the observed data would be under a selected null model, a p-value is still helpful in advanced health research (Sharma, H., 2021 & Bujang, M., 2025). Additionally, a number of recent studies contend that rather than being utilized in isolation, p-values can continue to be complementary evidence when interpreted with confidence ranges, effect sizes, or Bayesian approaches (Chén, O. Y. et al., 2023 & Ferreira, M. H. et al., 2026).

2.4 Common Problems regarding P-value

*Clinically useful results that barely miss significance may be obscured by strict $p<0.05$ thresholds (Patel, S., & Green, A., 2025).

*In biomedical literature, selective analyses and p-hacking produce an excessive number of false-positive results (Bonovas, S., & Piovani, D., 2023 & Goel, H. et al., 2024).

*Recent reporting guidelines contradict the common misinterpretation of large p-values as "no effect" (Mansournia, M., & Nazemipour, M. 2024).

*Misuse occurs in many fields, such as public health, toxicology, and nursing (Rovetta, A., 2023).

2.5 Better Reporting Approaches

A number of useful changes for health-science reporting have been discussed in recent articles. The precise p-value, predefined clinically significant differences, precision-enhancing confidence intervals, and direct discussion of impact magnitude are all advocated (Tsushima, E., 2022 & Mansournia, M., & Nazemipour, M. 2024).

In order to lessen binary thinking and enhance decision-making, some studies go further and suggest compatibility language, S-values, Bayes factors, multiple interval levels, or hybrid frequentist-Bayesian techniques (Greenland, S. et al., 2022 & Rovetta, A., 2023). Others link this reform agenda to preregistration, openness, Open Science, and more thorough reporting of assumptions, missing data, confounding, and interaction analyzes (Mansournia, M., & Nazemipour, M. 2024 & Ferreira, M. H. et al., 2026).

P-values and Statistical significance are important in modern health-science research because they play a limited but helpful function in evaluating data-model compatibility; nonetheless, they shouldn't be used alone to draw conclusions. Contextual interpretation is preferred over threshold-based assessment in the present literature, particularly when the research is intended to guide clinical or public health decisions (Bujang, M., 2025 & Ferreira, M. H. et al., 2026).

3. What is Confidence Interval (CI)

Confidence interval (CI) provides precision and clinical significance information to point estimates and P values by offering a range of reasonable values for the true population parameter, thereby quantifying the uncertainty surrounding effect estimates (Hespanhol, L. et al., 2019 & Schober, P. et al., 2020). Although its usage has been promoted for decades as a better option than isolated hypothesis testing, adoption in biological disciplines is still uneven (De Oliveira Santana Amaral, E., & Line, S. 2021).

3.1 Advantages of CI Over P-Value

While P-values alone classify data as significant or non-significant without revealing therapeutic importance, CI communicates an effect's magnitude, direction, and precision (Sharma, P. K., & Yadav, M., 2023). A 5% relative risk reduction (95% CI 1.0–9.0) vs a 30% reduction (95% CI 26.0–34.0) are two examples of interventions that may both attain statistical significance but differ significantly in practical consequence; this difference is only apparent through the CI (Trkulja, V., & Hrabac, P., 2019). Even when the point estimate is statistically significant, a broad confidence interval (CI) indicates imprecision, which frequently reflects insufficient sample size (El-Masri, M., 2017).

The result corresponds to statistical significance at $\alpha = 0.05$ when a 95% CI eliminates the null value (such as an odds ratio of 1.0), but the CI also shows the likely range of effect sizes (Schober, P. et al., 2020). Overconfident decisions in public health are lessened by interpreting CIs as ranges of compatibility with various impact sizes rather than only determining whether the null is excluded (Rovetta, A. et al., 2024).

3.2 Applications of CI in Health Sciences Research

CIs are increasingly a crucial component of contemporary trial designs, such as superiority, non-inferiority, and equivalency trials, where conclusions are based on whether CI boundaries cross predetermined margins rather than just P values. To prove drug formulation equivalency in bioequivalency testing, the 90% CI of ratios for pharmacokinetic parameters must be between 80 and 125% (Hazra, A., 2017). Because typical CIs may have inaccurate coverage probability when trial adjustments are not taken into consideration, regulatory bodies like the FDA and EMA mandate pre-specified CI approaches for adaptive trial designs (Robertson, D. S. et al., 2024).

*CIs are applicable to a variety of parameters, including as means, odds ratios, correlation coefficients, proportions, and diagnostic accuracy metrics (Schober, P. et al., 2020).

*To express accuracy across pooled evidence, CIs from individual trials should be provided and displayed in meta-analyses (Sim, J., & Reid, N., 1999).

*Modified CI techniques based on quadratic-form statistics perform better than traditional normal approximations for small-sample or rare-event data (Wu, M. et al., 2025).

3.3 Reporting Practices and Gaps

Health sciences disciplines differ greatly in their adoption of CIs. According to a 2024 analysis of 83 RCTs published in the Indian Journal of Anesthesia, 84.3% of them reported CIs, a significant improvement ascribed to increasing methodological understanding (Gupta, L. et al., 2025). Conversely, among 100 physical therapy trials published in 2021, only 47% recorded confidence intervals (CIs) for between-group differences, and only 6% both reported and evaluated them (Fernández Hernando, D. et al., 2024). Complementary statistical tests (effect size or CI) continue to be underutilized in clinical research domains, according to a more comprehensive scientometric examination of around 120,000 biomedical articles from 2019–2020 (De Oliveira Santana Amaral, E., & Line, S. 2021).

Despite increasing usage, many misconceptions still exist. 95% of intervals from hypothetical repeated samples would capture the true parameter; the CI does not represent the range of effects observed in 95% of individuals, nor is there a 95% probability that any single computed interval contains the true parameter (Schober, P. et al., 2020 & Gupta, L. et al., 2025). If underlying statistical assumptions are broken, narrow intervals are not always a sign of great precision (Rovetta, A. et al., 2024). To increase reporting frequency and interpretation quality, biostatistics training and editorial mandates are advised (De Oliveira Santana Amaral, E., & Line, S. 2021 & Gupta, L. et al., 2025).

Because it combines impact magnitude, precision, and clinical interpretability in a way that P values cannot, CI continues to be the most complete single reporting tool in health sciences research (Gupta, L. et al., 2025). However, researchers must evaluate interval boundaries in context rather than limiting them to another significance test in order to fully realize their relevance (Hespanhol, L. et al., 2019).

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD: https://orcid.org/0009-0006-1795-884X
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.
Artificial Intelligence (AI) Use Disclosure: The authors declare that no artificial intelligence tools were used in the preparation of this manuscript.

References

[1] Ahmed, E. S., & Butt, M. N. (2025). The misunderstood P-value: Why statistical significance is not enough in clinical practice. *British Journal of Anaesthesia,* 134(4), 909–913. https://doi.org/10.1016/j.bja.2025.01.008
[2] Altman N, & Krzywinski M. (2024). Understanding p-values and significance. Laboratory Animals; 58(5):443-447. doi:10.1177/00236772241247106
[3] Bonovas, S., & Piovani, D. (2023). On p-Values and Statistical Significance. *Journal of Clinical Medicine, 12.* https://doi.org/10.3390/jcm12030900
[4] Boscardin, C. K., Sewell, J. L., Tolsgaard, M. G., & Pusic, M. V. (2024). How to Use and Report on p-values. Perspectives on Medical Education, 13, 250 - 254. https://doi.org/10.5334/pme.1324
[5] Bujang, M. (2025). The Dilemma and Wisdom in Translating p Values: A Collaborative Approach to Strengthening Scientific Validity. BioMed Research International, 2025. https://doi.org/10.1155/bmri/6703756
[6] Chén, O. Y., Bodelet, J. S., Saraiva, R. G., Phan, H., Di, J., Nagels, G., Schwantje, T., Cao, H., Gou, J., Reinen, J. M., Xiong, B., Zhi, B., Wang, X., & Vos, M. (2023). The roles, challenges, and merits of the p value. Patterns, 4. https://doi.org/10.1016/j.patter.2023.100878
[7] Ciapponi, A., Belizán, J., Piaggio, G., & Yaya, S. (2021). There is life beyond the statistical significance. Reproductive Health, 18. https://doi.org/10.1186/s12978-021-01131-w
[8] De-Oliveira S A, E., & Line, S. (2021). Current use of effect size or confidence interval analyses in clinical and biomedical research. Scientometrics, 126, 9133 - 9145. https://doi.org/10.1007/s11192-021-04150-3

[9] El-Masri, M. (2017). Confidence in confidence intervals! Canadian Journal of Nursing Research, 49, 3 - 4. https://doi.org/10.1177/0844562117692406
[10] Fernández H, D., Elkins, M. R., & Freire, A. P. C. F. (2024). Reporting of confidence intervals, achievement of intended sample size, and adjustment for multiple primary outcomes in randomised trials of physical therapy interventions: an analysis of 100 representatively sampled trials. *Brazilian journal of physical therapy, 28 3,* 101079. https://doi.org/10.1016/j.bjpt.2024.101079
[11] Goel, H., Raheja, D., & Nadar, S. K. (2024). Evidence-based medicine or statistically manipulated medicine? Are we slaves to the P-value? *Postgraduate medical journal*. https://doi.org/10.1093/postmj/qgae012
[12] Greenland, S., Mansournia, M., & Joffe, M. (2022). To curb research misreporting, replace significance and confidence by compatibility: A Preventive Medicine golden jubilee article. Preventive medicine, 107127. https://doi.org/10.1016/j.ypmed.2022.107127
[13] Gupta, L., Chaudhary, K., & Bharti, D. (2025). Confidence interval reporting in randomised controlled trials: A critical analysis of IJA publications. *Indian Journal of Anaesthesia, 69,* 1069 - 1072. https://doi.org/10.4103/ija.ija_811_25
[14] Hamasaki, T., Bretz, F., LaVange, L., Müller, P., Pennello, G., & Pinheiro, J. (2021). Editorial: Roles of Hypothesis Testing, p-Values and Decision Making in Biopharmaceutical Research. Statistics in Biopharmaceutical Research, 13, 1 - 5. https://doi.org/10.1080/19466315.2021.1874803
[15] Hazra, A. (2017). Using the confidence interval confidently. Journal of thoracic disease, 9 10, 4125-4130. https://doi.org/10.21037/jtd.2017.09.14
[16] Hespanhol, L., Vallio, C. S., Costa, L. M., & Saragiotto, B. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian journal of physical therapy*. https://doi.org/10.1016/j.bjpt.2018.12.006
[17] Holmberg C. (2024). Why we need to discuss statistical significance and p-values (again): Addressing the underlying issue of different probability interpretations and actionable recommendations. *Nordic Journal of Nursing Research;44*. doi:10.1177/20571585241253177
[18] Ioannidis, J. (2018). The Proposal to Lower P Value Thresholds to .005. JAMA, 319 14, 1429-1430. https://doi.org/10.1001/jama.2018.1536
[19] Kafadar, K. (2021). EDITORIAL: Statistical significance, P-values, and replicability. The Annals of Applied Statistics. https://doi.org/10.1214/21-aoas1500
[20] Kwak, S. (2023). Are Only p-Values Less Than 0.05 Significant? A p-Value Greater Than 0.05 Is Also Significant! *Journal of Lipid and Atherosclerosis*, 12, 89 - 95. https://doi.org/10.12997/jla.2023.12.2.89
[21] Lohse, K. R., & Kliethermes, S. A. (2025). Approaching Significance: Statistical Guidance for Authors and Reviewers. *Journal of Neurologic Physical Therapy, 49*, 240 - 247. https://doi.org/10.1097/npt.0000000000000526
[22] Mansournia, M., & Nazemipour, M. (2024). Recommendations for accurate reporting in medical research statistics. Lancet, 403 10427, 611-612. https://doi.org/10.1016/s0140-6736(24)00139-9
[23] Mansournia, M., Nazemipour, M., & Etminan, M. (2022). P-value, compatibility, and S-value. Global Epidemiology, 4. https://doi.org/10.1016/j.gloepi.2022.100085
[24] Patel, S., & Green, A. (2025). Death by p-value: the overreliance on p-values in critical care research. Critical Care, 29. https://doi.org/10.1186/s13054-025-05307-9
[25] Robertson, D. S., Burnett, T., Choodari-Oskooei, B., Dimairo, M., Grayling, M. J., Pallmann, P., & Jaki, T. (2024). Confidence Intervals for Adaptive Trial Designs I: A Methodological Review. Statistics in Medicine, 44. https://doi.org/10.1002/sim.70174
[26] Robinson, R., & Haviland, J. (2021). Understanding Statistical Significance and Avoiding Common Pitfalls. Clinical oncology (Royal College of Radiologists (Great Britain)). https://doi.org/10.1016/j.clon.2021.06.008
[27] Rovetta, A. (2023). A Framework to Avoid Significance Fallacy. Cureus, 15. https://doi.org/10.7759/cureus.40242
[28] Rovetta, A., Mansournia, M., Vitale, A., & Accurso, V. (2024). How to read P-values and confidence intervals in public health studies. *Saudi Journal of Anaesthesia, 18,* 459 - 460. https://doi.org/10.4103/sja.sja_128_24
[29] Schober, P., & Vetter, T. (2020). Confidence Intervals in Clinical Research. Obstetric Anesthesia Digest. https://doi.org/10.1097/01.aoa.0000719460.05083.4b
[30] Sharma, H. (2021). Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. *Saudi Journal of Anaesthesia, 15,* 431 - 434. https://doi.org/10.4103/sja.sja_158_21
[31] Sharma, P. K., & Yadav, M. (2023). Confidence Interval: Advantages, Disadvantages and the Dilemma of Interpretation. Reviews on recent clinical trials. https://doi.org/10.2174/0115748871266250231120043345Tsushima, E. (2022). Interpreting Results from Statistical Hypothesis Testing: Understanding the Appropriate P-value. Physical therapy research, 25 2, 49-55. https://doi.org/10.1298/ptr.r0019
[32] Sim, J., & Reid, N. (1999). Statistical inference by confidence intervals: issues of interpretation and utilization. Physical therapy, 79 2, 186-95. https://doi.org/10.1093/ptj/79.2.186
[33] Thiese, M., Ronna, B. B., & Ott, U. (2016). P value interpretations and considerations. *Journal of thoracic disease, 8 9*, E928-E931. https://doi.org/10.21037/jtd.2016.08.16
[34] Trkulja, V., & Hrabac, P. (2019). Confidence intervals: what are they to us, medical doctors? *Croatian Medical Journal, 60*, 375 - 382. https://doi.org/10.3325/cmj.2019.60.375

[35] Wu, M.-M., Xu, M., & Kim, D. (2025). Confidence Intervals Based on the Modified Chi-Squared Distribution and Its Applications in Medicine. *Journal of Statistical Theory and Practice, 20*. https://doi.org/10.1007/s42519-026-00579-y