**ISSN- 0301-1216**
*Indian J. Prev. Soc. Med. Vol. 56, No.4, 2025*

## EDITORIAL

# Beyond Statistical Significance: Moving Past p< 0.05: Advancing Inference with Second-Generation P-Values

Abhishek Jaiswal

## ABSTRACT

For decades, the p-value has been the dominant tool for statistical inference in biomedical and epidemiological research. Despite its ubiquity, the p-value has been widely criticized for fostering dichotomous thinking, encouraging misinterpretation, and obscuring clinical relevance. Second-generation p-values (SGPVs) have emerged as an alternative, offering a more nuanced approach by explicitly considering the overlap between confidence intervals and clinically defined thresholds of trivial effects. Unlike traditional p-values, which test a point null hypothesis, SGPVs incorporate an interval null, thereby distinguishing between inconclusive findings and true evidence of clinically unimportant effects. This paper outlines the limitations of conventional p-values, explains the framework of SGPVs, and discusses their advantages, challenges, and potential applications in clinical research, reproducibility studies, and high-dimensional data analysis. By shifting emphasis from statistical to practical significance, SGPVs promise greater transparency, interpretability, and rigor in scientific inference.

**Authour(s) Details:**

Assistant Professor, Deptt. of Community Medicine, IMS, BHU, Varanasi, **Email:** jaiswal.aiims@gmail.com **Phone:** 7982298020, 9205624317; **ORCID ID:** https://orcid.org/0000-0002-1289-5425

**Corresponding Address:** Assistant Professor, Department of Community Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi- 221005, India

**Citation:** Jaiswal Abhishek. Beyond Statistical Significance: Moving Past p< 0.05: Advancing Inference with Second-Generation P-Values. Indian J Prev Soc Med, 2025; 56 (4): **454-456.DOI: https://doi.org/10.5281/zenodo.18045104**

**Sequence of Article:** **Submission:** 19.08.2025 **Accepted:** 10.10.2025 **Published:** 31.12.2025

***Prior Publication: Nil; Source of Funding: Nil; Conflicts of Interest: None, Article # 460***

### Introduction

The p-value has long been regarded as the cornerstone of statistical inference. In medical research, it is frequently employed to guide clinical decision-making and determine whether an observed effect is "statistically significant." However, reliance on the conventional threshold of $p < 0.05$ has drawn considerable criticism, as it often conflates statistical with clinical importance and contributes to widespread misinterpretation.[1-3]

A conventional p-value represents the probability of observing data as extreme as, or more extreme than, the actual data, assuming the null hypothesis is true.[1] Yet this quantity is often misunderstood as the probability that the null hypothesis itself is true.[2] More importantly, the dichotomous classification of results as "significant" or "non-significant" fails to convey the size or relevance of an effect.[3] Small effects may be statistically significant in large trials but clinically trivial, while clinically meaningful findings may be dismissed in underpowered studies.

In response to these concerns, Blume and colleagues introduced the concept of second-generation p-values (SGPVs), which aim to provide a more informative summary of evidence by shifting from a point null hypothesis to an interval null hypothesis.[4,5] This approach explicitly incorporates clinical or scientific relevance into the inferential process. Recent appeals in the scientific community have called for moving beyond the rigid dichotomy of "p <0.05" towards estimation-based inference that integrates clinical judgment.[6]

**Limitations of Traditional P-Values.** The main shortcomings of conventional p-values include:

1. A statistically significant result may correspond to an effect size too small to matter clinically. Conversely, a non-significant result may obscure a potentially important effect.
2. Large p-values are often mistaken as evidence for the absence of an effect, when they may instead indicate insufficient study power.[7]
3. The arbitrary cutoff of $p<0.05$ encourages "black-or-white" thinking, overlooking shades of uncertainty.[8]
4. Pressure to achieve significance has fueled practices such as p-hacking, selective reporting, and overemphasis on marginally significant results.[9,10]
5. Traditional p-values cannot differentiate between results compatible with no effect and those that are simply underpowered.[7]

**Concept and Calculation of Second-Generation P-Values**: SGPVs address these issues by comparing confidence intervals with a clinically defined interval null hypothesis. The null is defined not as a single value (e.g., exactly zero effect), but as a range of effect sizes deemed practically irrelevant.[4,11] The SGPV ($P_\delta$) is the proportion of the confidence interval that overlaps with this null range, adjusted for interval widths.[9] Its interpretation is straightforward:

- $P_\delta = 1$: Entire confidence interval lies within the null range → strong evidence for a trivial effect.
- $P_\delta = 0$: Entire confidence interval lies outside the null range → strong evidence for a meaningful effect.
- $0 < P_\delta < 1$: Partial overlap → results are inconclusive.

Formula: $P_\delta = |I \cap H_0| / |I|$

This framework forces researchers to specify, *a priori*, what constitutes a clinically unimportant effect, thereby linking statistical inference to practical decision-making.

**Example 1:** Consider a scenario for a new blood pressure lowering drug. Imagine a clinical trial testing a new drug to lower Systolic Blood Pressure (SBP), on a very large number of patients (N = 10,000 patients). With the result showing the drug lowers blood pressure by a very small, consistent amount. A drop of 0.1 mmHg is biologically meaningless. Let's assume doctors agree that any change less than 3 mmHg is trivial. Interval Null Hypothesis ($H_0$): [−3, +3]. Any value inside this range is "practically zero." The study runs, and because the sample size is huge, we get a very precise estimate. Observed Mean Difference: -1.0 mmHg (a drop of 1 unit). 95% Confidence Interval (I): [−1.2, −0.8]. As This interval does not include 0, the traditional p-value is highly significant (<0.001). Traditional Conclusion: "The drug works! The result is statistically significant." However, this is misleading. The drug "works," but it lowers BP by a useless amount (1 mmHg). (Figure 2)

Now for Calculation of SGPV, we have to ask Logic: What fraction of the data-supported hypotheses are in the trivial zone? Calculate the Widths: Data Interval (I): [−1.2, −0.8], Width = |−0.8−(−1.2) |=0.4. Null Interval ($H_0$): [−3, +3]. Calculate the Overlap ($I \cap H_0$): Does the data interval [−1.2, −0.8] fall inside the null zone [−3, +3]? Yes, entirely. Every plausible value for the drug's effect is between -1.2 and -0.8, which is well inside the "trivial" zone. Overlap Width = 0.4. Compute SGPV: $P_\delta$= 0.4/0.4=1. This Confirms the null, and therefore it can be said that the effect is precisely estimated to be trivial.

**Example- 2:** Now take another example of same drug with smaller sample size of 100. Suppose we observed the mean difference of -3.5 mmHg with 95% C.I. of -6 to -1 mmHg. As this interval also doesn't contain '0', the traditional p-value will come significant (<0.05). This gives a green light, ignoring the fact that the drug might only lower BP by 1 points (upper limit of the confidence interval). (Figure 2)

Now for calculation of SGPV, Data Interval (I): [−6, −1]. Width = |−1−(−6) |=5. Null Interval ($H_0$): [−3, +3]. Calculate the Overlap ($I \cap H_0$): We need to find the part of our data interval [−6, −1] that "bleeds" into the trivial zone [−3, +3]. Overlap: From -3 (the edge of the null zone) to -1 (the edge of our data). Overlap Interval: [−3, −1]. Overlap Width: |−1−(−3) |=2. Compute SGPV: $P_\delta$= 2/5=0.4. This confirms the result to be Inconclusive." It warns you: "While the drug lowers BP, there is a 40% chance the effect is trivial (overlap width 2 / total width 5). These examples illustrate how SGPVs combine effect magnitude, precision, and clinical thresholds.

### Advantages of SGPVs:

1. Emphasis on Clinical importance: By requiring explicit definition of trivial effects, SGPVs align results with patient-centered outcomes.[4,11]
2. Distinguish null vs. inconclusive finding: Unlike traditional p-values, SGPVs separate genuine null effects ($p\delta=1$) from insufficient evidence ($0<p\delta<1$).[4,7]
3. Transparency and reproducibility: Pre-specifying the null interval reduces scope for post-hoc interpretation.[4,9]
4. Compatibility with multiple frameworks: SGPVs can be applied with confidence intervals, likelihood intervals, or Bayesian credible intervals.[5]
5. Natural adjustment for multiple testing: The method reduces false positives in high-dimensional research without arbitrary corrections.[9,12]

### Challenges and Limitations

Despite their appeal, SGPVs face practical hurdles:

- **Defining the null interval** requires consensus and clinical judgment, which can be subjective.[9]
- **Limited awareness and adoption** mean most researchers remain unfamiliar with the method.[13]
- **Software availability** is growing (e.g., R and Stata packages), but integration into standard statistical workflows is still evolving.[14,15]
- Critics note similarities to equivalence testing frameworks, though SGPVs offer a continuous measure of evidence rather than a binary test outcome.[9]

**Applications and Future Directions**: SGPVs hold particular promise in: (1) **Clinical trials:** They help differentiate between trivial and meaningful treatment effects.[14] (2) **Reproducibility studies:** Clarify whether nonsignificant findings reflect genuine null effects or lack of precision.[7] (3) **High-dimensional research:** Such as genomics, where small but significant effects may not be practically relevant.[15]

**Figure 1:** Conceptual diagram showing the disadvantages of conventional p-value, advantages of SGPV, and barriers to adoption of SGPV

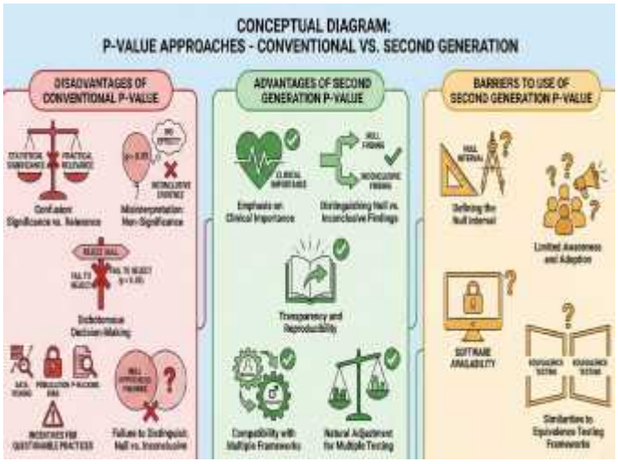
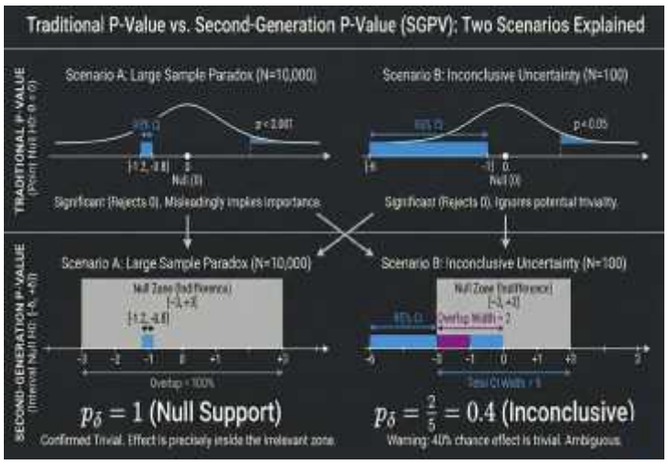



**Figure-2:** Difference between p-value and SGPV in two scenarios of the examples




Future directions may include development of standardized reporting guidelines and incorporation into CONSORT and STROBE frameworks. Integration into biostatistics curricula and regulatory guidance which will further support wider adoption.

## Conclusion

The second-generation p-value represents an important evolution in statistical inference. By moving beyond the rigid dichotomy of traditional p-values, it emphasizes clinical and practical relevance, distinguishes between null and inconclusive results, and improves transparency in research. While challenges remain—particularly in defining null intervals and promoting widespread uptake—SGPVs provide a more robust and meaningful framework for interpreting scientific evidence. Their integration into routine practice has the potential to strengthen both the rigor and reproducibility of medical and epidemiological research.

## References

1. Goodman SN. A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008;45(3):135-140. doi: 10.1053/j.seminhematol.04. 003.
2. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337-350. doi:10.1007/s10654-016-0149-3.
3. Gelman A, Stern H. The difference between "significant" and "not significant" is not itself statistically significant. Am Stat. 2006; 60(4): 328-331. doi:10.1198/000313006X152649.
4. Blume JD, D'Agostino McGowan L, Dupont WD, Greevy RA. Second-generation p-values: improved rigor, reproducibility, and transparency in statistical analyses. PLoS One. 2018;13(3): e0188299. doi: 10.1371/journal.pone.0188299.
5. Blume JD, Greevy RA, Welty VF, Smith JR, Dupont WD. An introduction to second-generation p-values. Am Stat. 2019;73(Suppl 1):157-167. doi:10.1080/00031305.2018.1537893.
6. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305-307. doi:10.1038/ d41586-019-00857-9.
7. Dienes Z. How to use Bayes factors to test hypotheses. J Math Psychol. 2014; 60:78-89. doi: 10.1016/j.jmp.2014.06.010.
8. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. Am Stat. 2016;70(2):129-133. doi:10.1080/ 00031305.2016.1154108.
9. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8): e124. doi: 10.1371/journal.pmed.0020124.
10. Gelman A, Loken E. The garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition." Psychol Methods. 2014;20(4):455-467. doi:10.1037/ met0000069.
11. Lakens D, Scheel AM, Isager PM. Equivalence Testing and the Second-Generation P-Value. Meta-Psychology. 2018;2(2). Available from: https://open.lnu.se/index.php/ metapsychology/ article/view/ 933.
12. D'Agostino McGowan L, Dupont WD, Blume JD. ProSGPV: an R package for variable selection with second-generation p-values. F1000Research. 2022; 11:58. doi:10.12688/f1000research.75973.1.
13. Staffa SJ, Zurakowski D. Guidelines for improving the use and presentation of P values. J Thorac Cardiovasc Surg. 2021;161(4):1367-1372. doi: 10.1016/j.jtcvs.2020.02.110.
14. Bormann S-K. A Stata implementation of second-generation p-values. Stata J. 2020;20(3):704-722.
15. Welty VF, Blume JD. SGPV: An R package for second-generation p-values. R J. 2020;12(1):153-16.