

Bayesian Reanalysis of Statistically Nonsignificant Outcomes in Plastic Surgery Clinical Trials

Gordon C. Wong, MBBS, MRCS*

Cynthia Huang, BS*

Joseph N. Fahmy, MD, MS*

Casey Zhang, BS†

Teun Teunis, MD, PhD†

Kevin C. Chung, MD, MS*

Background: Statistically nonsignificant randomized clinical trial (RCT) results are challenging to interpret, as they are unable to prove the absence of a difference between treatment groups. Bayesian analysis offers an alternative statistical framework capable of providing a comprehensive understanding of nonsignificant results.

Methods: This cross-sectional study conducted a post hoc Bayesian analysis of statistically nonsignificant outcomes from RCTs published in *Plastic and Reconstructive Surgery* from 2013 to 2022. Bayes factors representing the probability of the absence of a difference, or the null hypothesis of no difference, were calculated and examined. *P* values and Bayes factors of these outcomes were also compared with assessment of their association.

Results: In 73 studies with 176 statistically nonsignificant outcomes, 160 (91%) indicated evidence for the absence of a difference (Bayes factor > 1). For 110 (63%) of these, the Bayes factor was between 1 and 3, indicating weak evidence for the absence of a difference; 16 (9.1%) results supported the presence of a difference (Bayes factor < 1). A greater *P* value was independently associated with a larger Bayes factor ($\beta = 2.6$, $P < 0.001$).

Conclusions: Nearly two-thirds of nonsignificant RCT outcomes provided only weak evidence supporting the absence of a difference. This uncertainty poses challenges for clinical decision-making and highlights the inefficiency in resource utilization. Integrating Bayesian statistics into future trial design and analysis could overcome these challenges, enhancing result interpretability and guiding medical practice and research. (*Plast Reconstr Surg Glob Open* 2024; 12:e6370; doi: 10.1097/GOX.00000000000006370; Published online 13 December 2024.)

INTRODUCTION

Randomized controlled trials (RCTs) are crucial for guiding treatment decisions, with the majority using frequentist statistics. Under frequentist methods, the *P* value indicates the probability of observing the data, or more extreme results, under a hypothesis of no difference. A threshold of less than 0.05 typically indicates statistical significance, favoring one treatment over another. However,

encountering a *P* value greater than 0.05 suggests non-significance, which does not definitively indicate the absence of a difference but rather that the findings are inconclusive.¹ This highlights the principle that “absence of evidence is not evidence of absence,” especially because larger sample sizes might reveal statistically significant differences.²

The challenge of establishing conclusive evidence for the absence of a difference between two treatments within the frequentist framework complicates patient management decisions. It also raises uncertainties regarding the value of repeating an RCT with a larger sample size. Given these limitations, alternative statistical approaches that could enhance the insights gleaned from RCTs can be considered.³ Bayesian analysis provides an alternative approach to statistical inference, expressing experiment outcomes as probabilities rather than *P* values. Trial

From the *Section of Plastic Surgery, Department of Surgery, University of Michigan Medical School, Ann Arbor, MI; and †Department of Plastic Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA.

Received for publication August 13, 2024; accepted October 15, 2024.

Drs. Teunis and Chung contributed equally to this work as co-corresponding authors.

Copyright © 2024 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](#), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. DOI: 10.1097/GOX.00000000000006370

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

data are combined with prior knowledge to estimate the probability of different treatment effect sizes (posterior probability) (Fig. 1). Prior knowledge, often termed the “prior,” represents the researcher’s initial belief or knowledge about the likelihood of different hypotheses before observing the data. To achieve an unbiased result, a non-informative prior is used, reflecting minimal prior knowledge or influence on the analysis.

Under a Bayesian framework, the probability of equal effect sizes in both treatment groups (ie, null hypothesis) and the probability of a difference between treatment groups (ie, alternative hypothesis) can be expressed as a ratio known as the Bayes factor (Fig. 2). To illustrate this concept, consider a hypothetical trial comparing endoscopic carpal tunnel release to the open method. The null hypothesis assumes no difference between the two techniques, whereas the alternative hypothesis posits that the endoscopic method reduces recovery time. In this scenario, imagine the Bayes factor (BF_{01}) is 3. This means that under the observed data, the null hypothesis is 3 times more likely than the alternative hypothesis. Bayes factors offer 2 significant advantages over P values. First, frequentist statistics only permit the rejection of the null hypothesis but not its acceptance,⁴ whereas Bayes factors can assess the likelihood of no difference between treatment groups. Second, Bayesian factors offer insights into the strength or magnitude of the evidence-supporting hypotheses, providing a more comprehensive understanding of study outcomes. Currently, most clinicians are unfamiliar with Bayesian analysis.^{3,5,6}

By reanalyzing trials with nonsignificant results using Bayesian approaches, this study aims to examine the relationship between Bayes factors and P values in nonsignificant trial results and their impact on interpreting trial outcomes. This will familiarize physicians and researchers with probabilistic thinking and enable better assessment of the evidence in nonsignificant results within Plastic Surgery. Therefore, we asked the following: (1) What is the Bayes factor in nonsignificant *Plastic and Reconstructive*

Takeaways

Question: What is the Bayes factor of nonsignificant outcomes in plastic surgery clinical trials? Are P values independently associated with Bayes factor in these trials?

Findings: Bayes factors for most nonsignificant outcomes indicated only weak evidence for the absence of a difference between study interventions. Higher P values were independently associated with large Bayes factors.

Meaning: Bayesian analysis can complement traditional methods, enhancing the interpretation of clinical trial outcomes and reducing resource wastage in research.

Surgery (PRs) trials, and (2) are P values independently associated with Bayes factor in these trials?

METHODS

This cross-sectional study examines the relationship between P values and Bayes factors for statistically nonsignificant findings in RCTs within the *PRs* journal. Possessing the highest impact factor in its specialty, *PRs* was chosen to assess the quality of evidence in plastic surgery. On August 6, 2023, we systematically searched PubMed and MEDLINE databases using key terms including “randomized,” “clinical,” and “trial,” along with the inclusion of *PRs* as the journal source, and publication dates from 2013 to 2022. Randomized clinical trials reporting at least 1 statistically nonsignificant outcome ($P \geq 0.05$) were included. Exclusion criteria were reply letters, observational and educational studies, systematic reviews, and practice alerts. P values for all studies were recalculated to ensure veracity of outcome results. Outcomes that did not have reproducible P values because of a lack of reported data (eg, SDs) were excluded. Discordant P values, defined as a difference of greater than or equal to 0.1, were also excluded (Fig. 3).

The P value of the mean difference between study groups, group sizes, treatment effect sizes of the

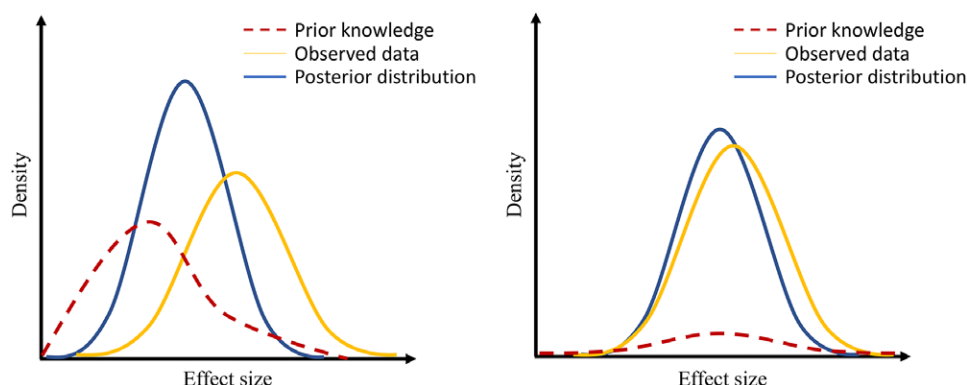


Fig. 1. Hypothetical example of the effect of prior knowledge on the posterior distribution. The graph on the left depicts how prior knowledge is combined with the observed data to derive the overall distribution of the treatment effect size (posterior distribution). The graph on the right illustrates how a weakly informative or noninformative prior produces an unbiased result approximating the probability distribution of the observed data.

intervention and comparator, and their respective SDs were collected to facilitate the calculation of Bayes factors and the recalculation of P values. Other recorded variables included total sample size, trial type, statistical test used, intervention types, comparator types, outcome types, and whether a power analysis was conducted.

To minimize bias, we collected data using a double-screening approach. Two independent reviewers (G.C.W. and C.H.) screened all the publications retrieved and applied the inclusion and exclusion criteria (Fig. 1). Reviewers started by screening the abstracts for the study type. All articles meeting the criteria were read in full. Many articles reported more than one statistically non-significant outcome. To prevent overcorrelation and dependency of outcome measures within our sample, we

included up to three statistically nonsignificant outcomes per study. Only outcomes calculated with an independent t test, chi-square test, or Fisher exact test, necessary for Bayes factor calculation, were considered. For studies with multiple nonsignificant outcomes, a standardized protocol for outcome selection was developed to ensure consistency. In cases where a study presented more than 3 nonsignificant outcomes, we chose to include the first 3 nonsignificant outcomes reported by chronological order in the text to standardize the data collection process across reviewers, based on the assumption that, generally, outcomes are reported in the order of importance as regarded by the study authors. When data were presented graphically without exact numbers, the relevant data were estimated from the graphs. After data collection was completed, the 2 independently collected datasets were compared and differences were resolved by consensus. Unresolved differences were escalated to a senior author (T.T.) for resolution.

We calculated Bayes factors using sample size and treatment effect size. Treatment effect size was calculated based on within or between-group differences in mean and SD for continuous variables and proportions for dichotomous variables. A BF_{01} of 1 indicates equal probability between hypotheses. BF_{01} values greater than 1 imply the evidence suggests there is no difference between groups, whereas values less than 1 indicate the evidence is in favor of a likely difference. The magnitude of BF_{01} reflects the strength of the evidence (Fig. 4). In terms of evidence for the hypothesis of no difference, values between 1 and

$$BF_{01} = \frac{\text{Probability that there is no difference between treatment groups}}{\text{Probability that there is a difference between treatment groups}}$$

Fig. 2. BF_{01} is the ratio of the probabilities of the null hypothesis (no difference) over the alternative hypothesis (presence of a difference) given the observed data. If BF_{01} is 1, both hypotheses are equally likely. If BF_{01} is greater than 1, there is a higher likelihood the null hypothesis is true. If BF_{01} is less than 1, the alternative hypothesis is more likely to be true.

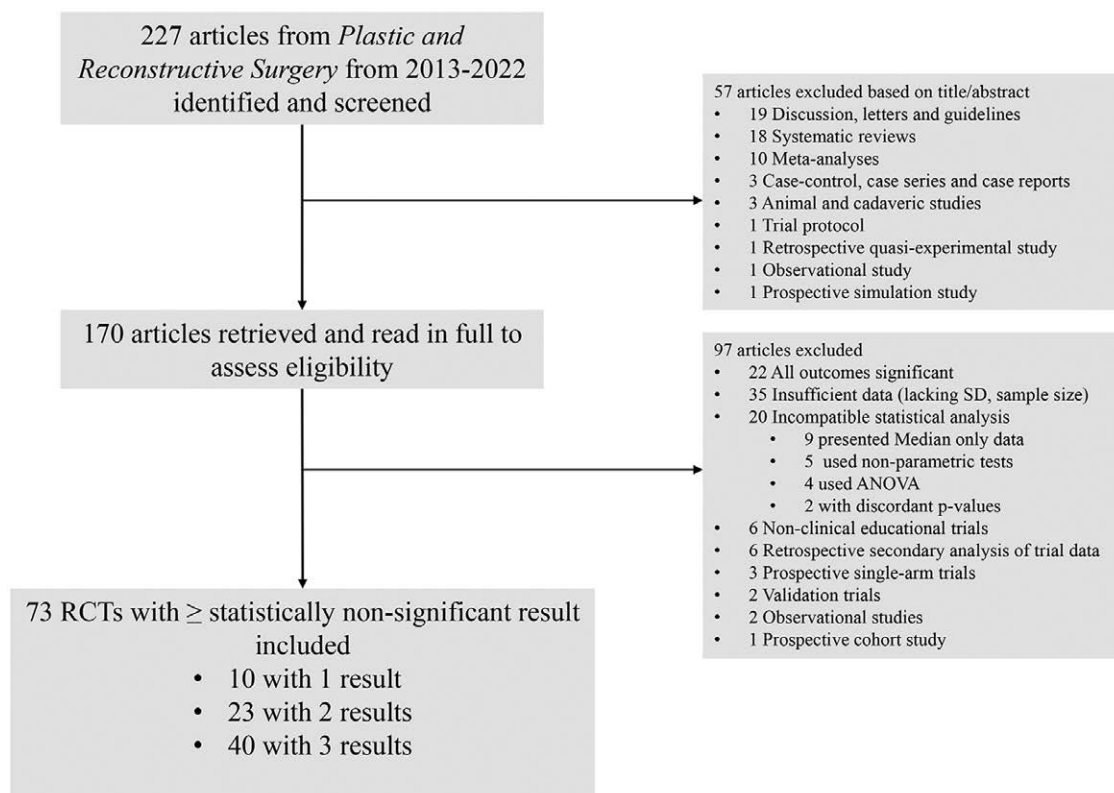


Fig. 3. Study selection flowchart.

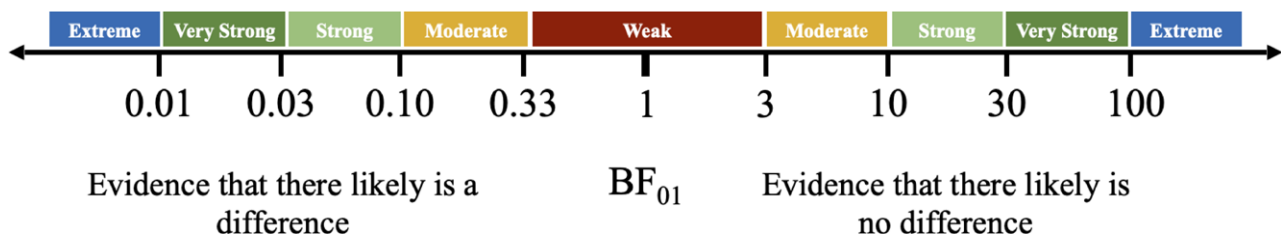


Fig. 4. Interpretation of Bayes factors.⁷⁻⁹

3 indicate weak evidence, whereas values greater than 3 suggest moderate evidence. Strong evidence is indicated by values greater than 10, very strong evidence by values exceeding 30, and values more than 100 imply decisive evidence.¹⁰ Conversely, values less than 1 indicate evidence suggesting the presence of a difference between the studied groups.

The sample size needed for this study was estimated through an a priori power analysis focusing on the simplified secondary study question of the relationship between *P* values and Bayes factor. A similar study reported a correlation coefficient of 0.29.¹¹ We anticipated a similar effect and powered our analysis to detect a correlation coefficient of 0.30. To achieve a power ($1-\beta$) of 0.80 and a significance level (α) of 0.05 for a 2-tailed test with an expected Pearson correlation coefficient of 0.30, a minimum sample of 85 outcomes was required.

Statistical analyses were performed using JASP version 0.18.1 (Department of Psychological Methods University of Amsterdam, Amsterdam, the Netherlands) and Stata 18 (College Station, TX). *P* values of the included outcomes were recalculated using the independent *t* test for continuous variables and the chi-square test for categorical variables. Using JASP, we generated Bayes factors with the Bayesian independent samples *t* test function for continuous variables and the Bayesian A/B test function for categorical data. A default weakly informative Cauchy prior (scale = 0.707) was chosen for the independent samples *t* test.¹² For the Bayesian A/B testing, we assumed a standard normal prior ($\mu = 0$, $\sigma = 1$), which is the default setting. Prior model probabilities for H_0 and H_1 were set to 0.5. Bayes factors were also stratified and analyzed by outcome type and the presence of a priori power analysis to evaluate trends in the level of evidence. Bivariate analyses were used to identify important factors influencing the Bayes factors. As some trials had more than 1 nonsignificant outcome, the nested structure of the data needed to be accounted for. Hence, we used a mixed-effects multilevel linear regression model to nest outcomes within the individual trials. *P* value, sample size, variable type, intervention category, comparator category, and trial type were analyzed against the Bayes factor. Significant variables on bivariate analysis ($P < 0.05$) were added to the final model sequentially. (See Appendix, Supplemental Digital Content 1, which displays individual bivariate analysis using mixed linear regression model accounting for nesting. <http://links.lww.com/PRSGO/D680>.)

RESULTS

The search retrieved 227 articles from the *PRIS* journal from 2013 to 2022. Data were collected from July to December 2023. After reading the titles and abstracts, 57 articles were found to not be RCTs and excluded. The remaining 170 articles were read in full. Articles that did not use the specified compatible statistical tests had incomplete data reporting, or only had significant outcomes were excluded. After exclusion, there were 73 articles yielding 176 nonsignificant outcomes (Fig. 1). Of the 176 outcomes analyzed, 52 were identified as primary, 75 as secondary, and 49 did not specify whether they were primary or secondary. Additionally, 56% of the included articles (41 of 73) conducted a priori power analysis.

The mean sample size of included trials was 108 (SD 106). 2-arm trials were the most common study design making up 86% (63 of 73) of all trials. Medical devices and equipment were the most frequently studied experimental treatment comprising 37% (27 of 73) of all trials, followed by surgical techniques and drugs both comprising 25% (18 of 73) each; 64% (47 of 73) of trials compared the test intervention with the usual care. The remaining trials used specific active interventions (active controls) and placebos equally. Treatment effect sizes were largely expressed in absolute terms (difference in means, percentage difference) with only a handful using relative terms (odds ratios, relative risk, effect size) (Table 1).

The mean *P* value was 0.47 ± 0.28 (range 0.054–1) and mean BF_{01} was 2.4 ± 1.6 (range 0.31–11). Of 176 outcomes analyzed, 91% (160 of 176) exhibited Bayes factors pointing toward no likely difference between treatment groups ($BF_{01} > 1$). The majority (63%, 110 of 176) suggested a weak level of evidence favoring the absence of a difference between treatment groups, whereas 27% (48 of 176) of outcomes provided moderate evidence suggesting no likely difference. Only 1.1% (2 of 176) displayed strong evidence for the same conclusion (Table 2). A small proportion of outcomes (9.1%, 16/176) displayed Bayes factors supporting a likely difference between treatment groups ($BF_{01} < 1$). However, all Bayes factors for these outcomes indicated weak evidence, with none reaching a level of moderate or stronger evidence for a likely difference between treatment groups. In comparison, when analyzing primary outcome variables alone, a similar distribution of Bayes factors across the evidence levels was observed (Table 3).

Stratifying by outcome type, the mean BF_{01} for primary outcomes was significantly lower than secondary outcomes

and unspecified outcomes ($P = 0.031$). The mean BF_{01} for primary outcomes was 1.9 ± 0.86 compared with 2.6 ± 1.8 for secondary outcomes and 2.5 ± 1.1 for unspecified outcomes (Table 4). Similarly, outcomes from studies that performed a priori power analyses demonstrated significantly lower mean BF_{01} ($P = 0.023$). The mean BF_{01} for outcomes that had a priori power analysis performed was 2.1 ± 1 compared with 2.7 ± 1.8 for outcomes that did not have a priori power analysis (Table 4).

Mixed linear regression analysis revealed that a higher P value was independently associated with a greater Bayes factor ($\beta = 2.5$; 95% confidence interval, 1.9–3.1; $SE = 0.31$; P value < 0.001). This finding suggests that there is a positive association between P values and Bayes factors, meaning that as the P value increases, the magnitude of the Bayes factor also increases. Specifically, for every unit increase in P value, the Bayes factor increases by approximately 2.6 units (Table 5).

DISCUSSION

In more than two-thirds of the nonsignificant trial outcomes, Bayesian analysis yielded only weak evidence of no difference, or the presence of a potential difference. This implies that a substantial proportion of the studied RCTs

failed to deliver conclusive evidence. When stratified by outcome type, primary outcomes displayed a lower mean Bayes factor compared with secondary or unspecified outcomes. Although this was statistically significant, it was not clinically significance as the mean Bayes factor for primary outcomes still fell squarely within the weak evidence category. In contrast to our findings, a review examining 43 nonsignificant outcomes from 36 RCTs published in the *New England Journal of Medicine* found that only 7% (3 of 43) demonstrated weak evidence of the absence of any difference with the rest indicating at least moderate evidence for the same conclusion. A larger review studied 162 nonsignificant RCT outcomes published in the *New England Journal of Medicine*, *BMJ*, *Lancet*, *JAMA*, and *Annals of Internal Medicine* using the likelihood ratio (a measure similar to Bayes factors). They found that 11% (18 of 169) demonstrated weak evidence indicating no difference, whereas the remaining outcomes showed at least moderate evidence for the same conclusion.¹³ Both reviews included studies with much larger sample sizes (mean 3020, range 33–18,446 and mean 5872, range 100–202,638), which likely accounts for the disparity with our findings (mean sample size 108, range 18–649). Plastic surgery RCTs face significant challenges in design, conduct, and recruitment. The inherent difficulties of conducting surgical trials—such as issues with patient autonomy, ethical considerations, and logistical complexities related to randomization and blinding—are compounded by the smaller patient populations typical of this niche specialty. These obstacles exacerbate recruitment issues and make conducting robust RCTs in plastic surgery more challenging, contributing to less conclusive results. Given that surgery inherently entails iatrogenic harm, the importance of obtaining conclusive evidence is even greater. Although multicenter collaboration and increased funding could alleviate some of these obstacles, these solutions are often difficult to implement.

A priori power analysis, an essential trial design step to ensure credible results and limit redundancy, was only performed in just over half of the included studies. In surgical RCTs, recruitment is a perennial challenge, and enrollment often targets the minimum sample size dictated by the power analysis. Although this leads to a study being efficient (neither overpowered nor underpowered), it can also lead to fragility and compromise reproducibility of results.¹⁴ One key distinction between Bayesian and frequentist approaches is that Bayesian analysis permits continuous data analysis, enabling new data to be integrated into the analysis as they are collected. Consequently, Bayes factors can be updated in real-time, enabling early trial termination when sufficient evidence is reached,

Table 1. Characteristics of Included Articles

Characteristic	No. (%)	
	Trials	Outcomes
Total	73	176
Mean sample size	108	
Trial type		
2-arm	63 (86)	149 (85)
Multiarm	9 (12)	24 (14)
Factorial	1 (1.4)	3 (1.7)
Type of trial intervention		
Medical device/equipment	27 (37)	66 (38)
Surgical technique	18 (25)	45 (26)
Drug	18 (25)	44 (25)
Delivery/dose	4 (5.5)	10 (5.7)
Other	6 (8.2)	11 (6.3)
Type of comparator		
Usual care	47 (64)	114 (65)
Active control	13 (18)	31 (18)
Placebo/sham	13 (18)	31 (18)
Variable type		
Continuous		100 (57)
Categorical		76 (43)
Scale of treatment effect		
Absolute (difference in means)		160 (91)
Relative (OR, RR, effect size)		16 (9.1)

OR, odds ratio; RR, relative risk.

Table 2. Distribution of Bayes Factors for All Variables

BF_{01}	Evidence Level	No. (%)	P^*
0.3–1	Weak evidence there is likely a difference	16 (9.1)	0.090 (0.065)
1–3	Weak evidence there is likely <i>no</i> difference	110 (63)	0.37 (0.44)
3–10	Moderate evidence there is likely <i>no</i> difference	48 (27)	0.61 (0.28)
10–30	Strong evidence there is likely <i>no</i> difference	2 (1.1)	0.90 (0.0005)

*Values are presented as median (IQR).

Table 3. Distribution of Bayes Factors for Primary Outcome Variables

BF ₀₁	Evidence Level	No. (%)	P*
0.3–1	Weak evidence there is likely a difference	7 (13)	0.088 (0.06)
1–3	Weak evidence there is likely <i>no</i> difference	37 (71)	0.37 (0.29)
3–10	Moderate evidence there is likely <i>no</i> difference	8 (15)	0.55 (0.5)
10–30	Strong evidence there is likely <i>no</i> difference	0 (0)	0 (0)

*Median (IQR).

Table 4. Mean BF01 and P Value Based on Outcome Type and if Power Analysis Conducted

	No. (%)	Mean BF01 (SD)	P	Mean P (SD)	P
Total	176	2.4 (1.4)	—	0.47 (0.28)	—
Outcome type					
Primary	52 (30)	1.9 (0.86)	0.025*	0.39 (0.27)	0.067*
Secondary	75 (43)	2.6 (1.8)		0.52 (0.3)	
Unspecified	49 (28)	2.5 (1.1)		0.46 (0.26)	
Power analysis					
Conducted	101 (57)	2.1 (1)	0.041†	0.47 (0.3)	0.96†
Not conducted	75 (43)	2.7 (1.8)		0.46 (0.26)	

*Kruskal-Wallis test.

†Mann-Whitney U test.

Table 5. Final Mixed Regression Model

BF ₀₁	Regression Coefficient	Standard Error	Z	P	95% Confidence Interval
P	2.5	0.31	8.2	<0.001	1.9 3.1

thereby conserving resources. If result fragility is a concern, researchers can opt to continue recruitment until the desired level of evidence is achieved. Conversely, if the Bayes factor remains inconclusive, trials can be extended until a conclusion can be drawn, reducing the waste associated with inconclusive results.¹⁵

We found an independent association between *P* values and Bayes factors, where an increase in *P* value corresponded with an increase in the Bayes factor. This indicates that these 2 statistical measures are not entirely independent and can provide complementary insights about the strength of evidence in a study. Although previous research has shown that *P* values can sometimes exaggerate the strength of evidence, these studies illustrate how Bayes factors can be used alongside *P* values to provide a more nuanced understanding of trial data, particularly when frequentist methods alone may not be conclusive.^{16–18} For example, an RCT comparing avulsion versus decompression of the zygomaticotemporal branch of the trigeminal nerve for the treatment of bitemporal migraines found similar nonsignificant reductions in migraine days for both treatments (*P* = 0.5).¹⁹ Although this *P* value might imply the need for a larger trial to confirm findings, the corresponding Bayes factor of 4.2 provides moderate evidence that no difference exists, indicating a repeat trial may not be necessary. In contrast, another RCT comparing closed-incision negative pressure therapy to adhesive strips for surgical site infections showed a nonsignificant result (*P* = 0.86) but a Bayes factor of 1.9, which indicates weak evidence of no difference, suggesting the trial is worth repeating.²⁰ These examples underscore how Bayesian analysis can provide clearer insights into trial results, offering an

advantage over traditional *P* values by facilitating more informed decisions about whether additional studies are warranted. By using Bayesian inference, researchers can reduce resource wastage and improve the interpretability of trial outcomes.

Our study has limitations. There is potential for selection bias as studies were excluded because of incomplete reporting or incompatible data. Therefore, our results may not apply to the field of plastic surgery as a whole, but they give a general sense of the studies published in the top plastic surgery journal. We minimized these biases by ensuring a comprehensive search strategy and rigorous inclusion criteria. Given that we included more than 1 outcome per trial, there may be some dependency in the data. However, we accounted for this by using a multilevel model to accommodate the nested structure of the data. Finally, Bayesian analysis uses prior knowledge to inform the analysis process. In our study, we intentionally adopted neutral uninformative priors. The choice of such priors has minimal impact on our results.

Bayesian statistics is a powerful tool that is gaining traction in medical research, yet it remains unfamiliar to many physicians. This article not only presents the results of our Bayesian reanalysis but also aims to bridge the knowledge gap by introducing the fundamental principles of Bayesian analysis (Fig. 5). Unlike traditional frequentist methods, which provide a binary outcome of statistical significance, the Bayesian framework presents data through direct probabilities and credible intervals, offering a more intuitive and informative approach. Bayesian methods can overcome several challenges commonly faced in plastic surgery clinical trials. For instance, Bayesian analysis permits continuous data monitoring, enabling trials to adapt

**Study Planning**

- Define research question
- Identify and specify prior knowledge (distribution) based on existing evidence
- Choose the study design, outcomes, endpoints, and define the minimal clinically important difference

**Data Collection and Analysis**

- Perform study and collect data
- As data is collected, combine observed data with prior distribution to obtain posterior distribution
- Continuously monitor and assess the results, allowing for potential adaptations or early stopping based on updated beliefs

**Interpretation and Reporting**

- Report the results in terms of posterior probabilities, credible intervals, or Bayes factors, emphasizing the clinical relevance
- Translate the posterior distributions into clinically meaningful conclusions
- Explain choice of priors, the data collected, and how the Bayesian analysis led to the conclusions

Fig. 5. Key steps for designing and conducting a clinical trial using Bayesian analysis.

based on accumulating evidence. This flexibility can prevent the premature conclusion of studies that have not yet reached sufficient evidence, or enable early termination when the evidence is strong, thereby conserving resources. [Figure 5](#) illustrates the fundamental tenets of designing, conducting, interpreting, and reporting a Bayesian trial. By familiarizing themselves with these principles, plastic surgeons and researchers will have the option of leveraging Bayesian methods to conduct more efficient and informative trials.

Bayesian analysis offers several advantages over traditional frequentist methods, particularly in its ability to present data more intuitively through direct probabilities and credible intervals. This approach can also make clinical trials more efficient by enabling continuous data evaluation and adaptive designs, potentially reducing resource wastage. By understanding both the benefits and limitations, researchers can better decide when to use Bayesian methods to enhance the quality and efficiency of their trials.

Kevin C. Chung, MD, MS

Section of Plastic Surgery
The University of Michigan Health System
1500 E. Medical Center Drive
2130 Taubman Center, SPC 5340
Ann Arbor, MI 48109-5340
E-mail: kecchung@med.umich.edu

Teun Teunis, MD, PhD

Department of Plastic Surgery
University of Pittsburgh Medical Center
3550 Terrace Street
675 Scaife Hall
Pittsburgh, PA 15261
E-mail: teunteunis@gmail.com

DISCLOSURE

Dr. Chung receives funding from the National Institutes of Health and book royalties from Wolters Kluwer and Elsevier. Dr. Fahmy receives institutional support through a T32 training grant from the National Institutes of Health (T32GM008616-23) and reports grants from the Plastic Surgery Foundation and American Foundation for Surgery of the Hand, outside of the submitted work. The other authors have no financial interest to declare in relation to the content of this article.

REFERENCES

1. Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev.* 2009;2009:MR000006.
2. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ.* 1995;311:485.
3. Wijesundera DN, Austin PC, Hux JE, et al. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol.* 2009;62:13–21.e5.
4. Fisher R. *The Design of Experiments.* 1st ed. Oliver and Boyd Ltd; 1935.
5. Tiemens B, Wagenvoerde R, Witteman C. Why every clinician should know Bayes' rule. *Health Prof Educ.* 2020;6:320–324.
6. Yarnell CJ, Granton JT, Tomlinson G. Bayesian analysis in critical care medicine. *Am J Respir Crit Care Med.* 2020;201:396–398.
7. Quintana DS, Williams DR. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry.* 2018;18:178.
8. Jeffreys H. *The Theory of Probability.* Oxford University Press; 1961.
9. Lee MD, Wagenmakers E-J. *Bayesian Cognitive Modelling: A Practical Course.* Cambridge University Press; 2014.
10. Jeffreys H. *Theory of Probability.* 3rd ed. Oxford University Press; 1998.
11. Hoekstra R, Monden R, Van Ravenzwaaij D, et al. Bayesian reanalysis of null results reported in medicine: strong yet variable

- evidence for the absence of treatment effects. Li X, ed. *PLoS One*. 2018;13:e0195474.
12. Gelman A, Jakulin A, Pittau MG, et al. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2:1360–1383.
13. Perneger T, Gayet-Ageron A. Evidence of lack of treatment efficacy derived from statistically nonsignificant results of randomized clinical trials. *JAMA*. 2023;329:2050–2056.
14. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294:218–228.
15. Teunis T, Samadi A, Chen N, et al. Research methodology: the Bayesian statistical framework and the future of trial design. *J Hand Surg Eur Vol*. 2023;48:593–597.
16. Goodman SN. Toward evidence-based medical statistics. 2: the Bayes factor. *Ann Intern Med*. 1999;130:1005–1013.
17. Sellke T, Bayarri MJ, Berger JO. Calibration of *P* values for testing precise null hypotheses. *Am Stat*. 2001;55:62–71.
18. Johnson VE. Revised standards for statistical evidence. *Proc Natl Acad Sci U S A*. 2013;110:19313–19317.
19. Guyuron B, Harvey D, Reed D. A prospective randomized outcomes comparison of two temple migraine trigger site deactivation techniques. *Plast Reconstr Surg*. 2015;136:159–165.
20. Muller-Sloof E, De Laat E, Kenç O, et al. Closed-incision negative-pressure therapy reduces donor-site surgical wound dehiscence in DIEP flap breast reconstructions: a randomized clinical trial. *Plast Reconstr Surg*. 2022;150:38S–47S.