



Artigo original

Impacto da mudança do valor da significância estatística de 0,05 para 0,005 nos resultados dos ensaios clínicos randomizados publicados em anestesia: estudo transversal

Impact of changing the statistical significance value from 0.05 to 0.005 on the results of published randomized controlled trials in anesthesia: cross-sectional study

Impacto de cambiar el valor de significancia estadística de 0,05 a 0,005 en los resultados de ensayos controlados aleatorios publicados en anestesia: estudio transversal

Fabiano Timbó Barbosa¹ - Orcid-Id 0000-0001-6630-0629

¹ Universidade Federal de Alagoas, Maceió, AL, BR.

Autor correspondente: Fabiano Timbó Barbosa. CAMPUS A.C. SIMÕES. Av. Lourival Melo Mota, s/n – Tabuleiro do Martins Maceió, AL CEP 57072-900.

Recebido em: 31/07/2023 ----Aprovado em: 10/02/2026----Publicado em: 10/03/2026

RESUMO

Justificativa e objetivo: A ausência de reprodutibilidade dos resultados dos ensaios clínicos randomizados (ECR) vem se do questionado nos últimos anos senso de causa multifatorial, mas uma maior importância tem sido dada à significância estatística. O objetivo desta pesquisa foi descrever o impacto da mudança do valor da significância estatística para 0,005 nos resultados dos ensaios clínicos publicados em anestesia. **Método:** Foram utilizados dados de artigos originais de ECRs publicados na revista Brazilian Journal of Anesthesiology selecionados aleatoriamente. A variável primária foi a frequência dos estudos que mantiveram a significância estatística após a mudança do valor de p para 0,005. Os resultados da presente pesquisa foram comparados com dados da literatura utilizando o qui-quadrado de aderência, com nível de significância de 5% e poder estatístico de 80% por meio do aplicativo BioEstat 5.0. **Resultado:** Foram analisados dados de 73 artigos de ECRs. Todos utilizaram 0,05 como o nível de significância estatística e 61,7% (45/73) apresentaram diferença estatística significativa no momento da inclusão nesta pesquisa. A frequência dos ECRs que mantiveram a significância estatística após a mudança do valor de p de 0,05 para 0,005 foi de 55,5% (25/45) não havendo diferença com dados da literatura ($p = 0,5428$). **Conclusão:** A mudança de nível de significância de 0,05 para 0,005 levaria a uma revisão nas conclusões de 44,5% ECRs publicados em anestesia. A pesquisa identificou erros na estatística básica. A maior reprodutibilidade dos ECRs podem não ocorrer após a mudança do nível de significância devido a presença de falhas na utilização da estatística básica.

ABSTRACT

Palavras-Chave

Anestesia;
Ensaio clínicos
randomizados;
Significância
estatística,

Keywords

Background and objective: The lack of reproducibility of the results of randomized clinical trials (RCTs) has been questioned in recent years as a sense of multifactorial causation, but greater importance has been given to statistical significance. The aim of this research was to describe the impact of changing the statistical significance value to 0.005 on the results of published clinical trials in anesthesia. **Method:** Randomly selected data from original articles of RCTs published in the Brazilian Journal of Anesthesiology were selected. The primary variable was the frequency of studies that maintained statistical significance after changing the p value to 0.005. The results of this research were compared with data from the literature using the chi-square of adherence, with a significance level of 5% and statistical power of 80% using the BioEstat 5.0 application. **Result:** Data from 73 RCT articles were analyzed. All used 0.05 as the level of statistical significance and 61.7% (45/73) showed a statistically significant difference at the time of inclusion in this research. The frequency of RCTs that maintained statistical significance after changing the p value from 0.05 to 0.005 was 55.5% (25/45), with no difference with data from the literature ($p = 0.5428$). **Conclusion:** Changing the significance level from 0.05 to 0.005 would lead to a revision in the conclusions of 44.5% published RCTs on anesthesia. The survey identified errors in basic statistics. The greater reproducibility of RCTs may not occur after changing the significance level due to flaws in the use of basic statistics.

Anesthesia;

Randomized controlled trials;

Statistical significance.

RESUMEN

Fundamento y objetivo: La falta de reproducibilidad de los resultados de los ensayos clínicos aleatorizados (ECA) ha sido cuestionada en los últimos años como un sentido de causalidad multifactorial, pero se ha dado mayor importancia a la significación estadística. El objetivo de esta investigación fue describir el impacto de cambiar el valor de significación estadística a 0,005 en los resultados de los ensayos clínicos publicados en anestesia. **Método:** Se seleccionaron datos seleccionados aleatoriamente de artículos originales de ECA publicados en el Brazilian Journal of Anesthesiology. La variable principal fue la frecuencia de los estudios que mantuvieron la significación estadística después de cambiar el valor de p a 0,005. Los resultados de esta investigación fueron comparados con datos de la literatura utilizando el chi-cuadrado de adherencia, con un nivel de significancia del 5% y poder estadístico del 80% utilizando la aplicación BioEstat 5.0. **Resultado:** Se analizaron los datos de 73 artículos de ECA. Todos utilizaron 0,05 como nivel de significación estadística y el 61,7% (45/73) mostró diferencia estadísticamente significativa al momento de la inclusión en esta investigación. La frecuencia de ECA que mantuvieron la significación estadística tras cambiar el valor de p de 0,05 a 0,005 fue del 55,5% (25/45), sin diferencia con los datos de la literatura ($p = 0,5428$). **Conclusión:** Cambiar el nivel de significación de 0,05 a 0,005 supondría una revisión de las conclusiones del 44,5% de los ECA publicados sobre anestesia. La encuesta identificó errores en las estadísticas básicas. La mayor reproducibilidad de los ECA puede no ocurrir después de cambiar el nivel de significación debido a fallas en el uso de estadísticas básicas.

Anestesia;

Ensayos controlados aleatorios;

Significación estadística.

INTRODUCTION

The randomized clinical trial (RCT) has been considered for many years as the type of study that provides the highest level of evidence with a high potential for recommendation for clinical practice. RCT results are not always reproducible in clinical practice, which leads to a contradiction in this statement identified by some authors as the research reproducibility crisis (1). The lack of reproducibility of RCT results is multifactorial, but greater importance has been given to statistical significance (2).

A p-value less than 0.05 has been used in health research as a limit for statistical significance, but it has been much questioned today because it was an arbitrary choice in the past and that has been maintained without questioning over the years (3). The correct interpretation of the p-value does not seem to be as widespread as its use as a limit for statistical significance, so that readers generally overrule other

important characteristics to overestimate the p-value (4, 5). Experts in the area of statistics suggested that changing the p-value from 0.05 to 0.005 may help resolve the issue of reproducibility of RCT results (4, 5, 6).

The omission of p-value results in published articles does not seem to be an acceptable alternative in current research practice, so the author of this research aimed to describe the impact of changing the statistical significance value to 0.005 on previously published RCT results in anesthesia. The methodological characteristics of the RCTs were analyzed in this research for a better perception of the scenario of interest to this research.

METHOD

The research used data from original articles of randomized clinical trials published in the Brazilian Journal of Anesthesiology (BJAN) and freely available on the World Wide Web. The journal was chosen because it is a reference in the field of anesthesia in Brazil and has a high impact factor, demonstrating that the results are widely disseminated worldwide. The research was not submitted to the ethics committee because it is a cross-sectional study with secondary data analysis.

The inclusion criterion was an original article from a randomized clinical trial published in the BJAN, available for access at <https://www.scielo.br/>. The exclusion criterion was published article without the exact description of the p value for the result of the primary variable of the study. The choice of participants was made by chance using dice. Two drawings were carried out: the first indicated the volume of the magazine; and the second indicated the number of the copy to be used. All original articles from RCTs present in the randomly selected issue corresponding to the volume and issue of the journal were selected for data collection.

The primary variable was the frequency of studies that maintained statistical significance after changing the p value to 0.005. Secondary variables were: type of anesthesia, conflict of interest, clinical outcome (simple or composite), statistical analysis (by protocol or by intention to treat), sample size calculation, description of data distribution analysis, description of statistical tests, multiple comparisons, and frequency of incorrect use of basic statistics. Secondary variables were collected from all randomized clinical trial articles.

The use of basic statistics was considered incorrect when: data from continuous variables had a symmetrical distribution and were not described using mean and standard deviation; the normality test did not precede the choice between parametric and non-parametric tests; there was no description of the use of the normality test; there was no description of the correction factor for multiple comparisons; the variable was categorical, the sample size was less than 30 participants and its analysis was performed using the chi-square; and when the variable was categorical, the sample size was greater than 30 and its analysis was performed using the Fischer test. The analysis of the application of basic statistics was considered inconclusive when: it was not possible to evaluate the distribution of variables, the p value was reported without the description of the statistical test used, the p value was not accurately described and when the statistical method of the included article did not present the description of the statistical tests. The description of the distribution of variables was considered adequate when based on normality tests or by the joint analysis of histogram graphs, kurtosis coefficient and symmetry coefficient. The criteria for considering inappropriate use of basic statistics were used by at least one author in previous research analyzing articles published in BJAN (7).

The sample size was estimated at 73 articles considering 5% the minimum proportion of original articles that would maintain statistical significance after changing the significance level to 0.005, a statistical power of 80%, statistical precision of 5% and statistical significance of 5%. Descriptive statistics were used in result section.

The Medline database, via Pubmed, was used to identify published articles that analyzed the change in the statistical significance level to 0.005 in other anesthesia journals. The terms used in combination were: p value 0.005 and anesthesia. The results of this research were compared with data from the literature identified by this search using the chi-square of adherence, with a significance level of 5% and statistical power of 80% using the BioEstat 5.0 software.

RESULTS

Data from 73 RCTs articles were analyzed. All used 0.05 as the level of statistical significance and 61.7% (45/73) showed a statistically significant difference in their primary variable at the time of their inclusion in this research.

The RCT articles included for analysis came from the following countries: Turkey 36.9% (27/73), Brazil 28.7% (21/73), India 8.2% (6/73), China 6.8% (5/73), Egypt 6.8% (5/73), Serbia 2.7% (2/73) and another 9.58% (Italy 1/73, Pakistan 1/73, France 1/73, Canada 1/73, Iran 1/73, Czech Republic 1/73 and Chile 1/73). The funding source was mentioned in 17.8% (13/73) of the articles included and there was no description of this item in 82.2% (60/73). The authors of 98.6% (72/73) of the articles included declared that they had no conflict of interest.

The type of anesthesia or topic of interest in RCT articles were: general anesthesia 34.2% (25/73), ultrasound and use of local anesthetic 21.9% (16/73), tracheal intubation 13.6% (10/73), neuraxial block 9.5% (7/73), pain 8.2% (6/73), volume replacement 4.1% (3/73), blood glucose 2.7% (2/73) and another 5.47% (sedation 1/73, analgesia 1/73, ventilation 1/73 and ventilation 1/73). The clinical outcome was classified as simple in 87.7% (64/73) of the included articles and composite in 12.3% (9/73).

The sample size calculation was performed in 78.1% (57/73) of the articles included and, among these, the sample calculation was followed in 98% (56/57). Statistical tests were described in all included articles (Table 1). Multiple comparisons occurred in 12.3% (9/73) and among these the correction factor was used in 77.7% (7/9). The statistical test to analyze the symmetrical distribution of data was described in 64.4% (47/73) included articles, but among these 19.1% (9/47) did not describe which test was used. The description of the type of analysis was absent in 98.7% (72/73) of the articles included.

Table 1 Description of statistical hypothesis tests

Statistical tests	Primary outcome	Secondary outcome
Student t test	52,1% (38/73)	9,5% (7/73)
ANOVA	19,1% (14/73)	13,6% (10/73)
Chi-square	17,8% (13/73)	52,1% (38/73)
Kruskal- Wallis	4,1% (3/73)	6,8% (5/73)
Wilcoxon	4,1% (3/73)	9,5% (7/73)
Fisher's Exact Test	1,3% (1:83)	32,8% (24/73)
ANCOVA	1,3% (1/73)	
Mann-Whitney		36,9% (27/73)
Signal test		1,3% (1/73)
Wald		1,3% (1/73)

Basic statistics were classified as incorrect in 61.6% (45/73) of the included articles. Reasons for incorrect use of basic statistics were: Fisher's exact test was used for more than 30 participants 44.4% (20/45); the choice of tests in the absence of description of verification of data normality 42.2% (19/45); the

chi-square test was used for less than 40 participants 4.4% (2/45); and multiple analysis was performed without using correction factor 4.4% (2/45).

The frequency of RCTs that maintained statistical significance after changing the p-value from 0.05 to 0.005 was 55.5% (25/45). The search for articles published in the Medline database resulted in 643 titles in total. The analysis of titles and abstracts showed that only one research analyzed the change in the significance level to 0.005 in anesthesia.⁸ The frequency of analyzed articles that maintained statistical significance in this research was 58%, with no statistically significant difference between this data and the result of this research ($p = 0.5428$).

DISCUSSION

The recommendation to change the overall significance level from 0.05 to 0.005 in clinical research is intended to ensure the reproducibility of research findings in clinical practice. The impact of this change has already been researched in other areas of medicine, concluding that approximately half of the results should be reinterpreted (2, 8, 9). The suggestion of changing the significance level and the observation of findings in other specialties led researchers from this field to create the hypothesis that half of RCTs in anesthesia have false-positive results. The impact of the change in significance level on the clinical practice of anesthesia remains unknown as the change has not been widely used in currently performed clinical research. We performed a review of RCTs published in BJAN and compared their results with data from another journal to predict the possible impact of this change on daily clinical practice.

The results of this research demonstrated that 55.5% of the RCTs maintained statistical significance with the threshold change to 0.005. Authors of another anesthesia survey analyzing data from other journals demonstrated that 58% of RCTs maintained significance (8). It is noticed that there is a trend in anesthesia research demonstrating that almost half of the RCTs will present false-positive results after changing the significance level. Data from other specialties analyzed in conjunction with anesthesia data suggest that the result is not unique to anesthesia research. Decision-making based on RCT articles with false-positive results can lead to harm to the patient and the health system, such as longer hospital stays, higher expenses and lower levels of patient satisfaction.

The present research carried out an analysis of the basic statistics of the included articles, demonstrating that there was a failure in at least two thirds of the published studies. I used the same

technique for analyzing statistical data from BJAN articles in a previous research and found that 56.42% incorrectly used basic statistics (7). Comparative analysis of data from the previous research and data from this research showed an increase in the frequency of inappropriate use of basic statistics in clinical research published in Brazil. The inappropriate use of statistics leads to errors in the interpretation of the results, making them questionable and making the reproducibility of the results unfeasible in clinical practice (10) The recommendations must be followed to avoid statistical bias and enable the reproducibility of the results.

The analysis of methodological items such as funding source, conflict of interests, calculation of sample size and analysis of the symmetry distribution of quantitative data showed that their description of test items was not present in 100% of the RCT articles. The least described item was the funding source. The absence of a description of each item does not guarantee that the item was correctly performed in the research practice. The lack of description of this information may prevent the reader from being able to carry out an adequate risk of bias analysis, assuming the conclusions as coming from truly positive results when, in fact, they may be false positives. The reproducibility of research findings does not depend exclusively on adjusting the level of statistical significance, and journals should value the complete description of other methodological items.

The main implication for the clinical practice of the findings of this research was the observation that almost half of what is currently used must be revised with the adoption of the new value of statistical significance, as research results may not be reproducible because they are false-positive. The conducts of clinical practice must be reviewed and based on the findings of the reinterpretation of the results with the p-value adjusted for the new level of statistical significance and the protocols must be updated considering the new conclusions.

Future research should be carried out with the view that the statistical significance value should be 0.005, but current research should consider better use of basic statistics so that the results are more reproducible from now on. A viable alternative to the alpha value threshold of 0.005 is to adjust the statistical significance value to the statistical power that was considered for the sample size calculation (11).

The limitations of this study were: being a descriptive study, evaluating already published studies and the analysis of only one journal in the field of anesthesia. The character of this study being descriptive did not allow the analysis in practice of each RCT, that is, how the authors evaluated the magnitude of the

effect and the association of this effect with the treatment employed, so that the disparities between the way the research was carried out and the way it was written may have compromised the results of this research. Failures in the reporting of original articles have been decreasing after the adoption of publication guidelines, but methodological failures remain, making it necessary to evaluate the methodological items described in each original article. The analysis of published studies instead of the analysis of research in practice did not allow verifying the impact of the change in the type II error, because in addition to the change in the p value, there must also be a sample size adequate to this change. Data analysis only from BJAN did not impact the results of this research, since the published articles included in this research come from different parts of the world and the data were compared with research published in another journal.

Changing the significance level from 0.05 to 0.005 would lead to a revision in the conclusions of 44.5% published RCTs in anesthesia. The survey identified errors in basic statistics. The greater reproducibility of RCTs may not occur after changing the significance level due to flaws in the use of basic statistics.

REFERENCES

1. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452-4.
2. Bruno AM, Shea AE, Einerson BD, Metz TD, Allshouse AA, Scott JR, et al. Impact of the p-Value Threshold on Interpretation of Trial Outcomes in Obstetrics and Gynecology. *Am J Perinatol*. 2021;38(12):1223-30.
3. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet*. 2002;359(9300):57-61.
4. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA* 2018;319:1429–30.
5. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70:129–33.
6. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2:6–10.
7. Barbosa FT, de Souza DA. Frequency of the adequate use of statistical tests of hypothesis in original articles published in the *Revista Brasileira de Anestesiologia* between January 2008 and December 2009. *Rev Bras Anesthesiol*. 2010;60(5):528-36.
8. Chuang Z, Martin J, Shapiro J, Nguyen D, Neocleous P, Jones PM. Minimum false-positive risk of primary outcomes and impact of reducing nominal P-value threshold from 0.05 to 0.005 in anaesthesiology randomised clinical trials: a cross-sectional study. *Br J Anaesth*. 2023;130(4):412-20.
9. Landy DC, Utset-Ward TJ, Lee MJ. What Are the Implications of Alternative Alpha Thresholds for Hypothesis Testing in Orthopaedics? *Clin Orthop Relat Res*. 2019;477(10):2358-63.

10. Rodrigues CFS, Lima FJC, Barbosa FT. Importance of using basic statistics adequately in clinical research. Rev Bras Anesthesiol. 2017;67(6):619-25.

11. Barbosa FT. The New alpha value for statistical significance in clinical trials. Rev. Portal: Saúde e Sociedade [Internet]. 3º de dezembro de 2022 [citado 21º de julho de 2023];7(Fluxo contínuo):9. Disponível em: <https://www.seer.ufal.br/index.php/nuspfamed/article/view/14075>

Como citar	Barbosa, F. T. (2026). Impacto da mudança do valor da significância estatística de 0,05 para 0,005 nos resultados dos ensaios clínicos randomizados publicados em anestesia: estudo transversal. <i>Revista Portal: Saúde E Sociedade</i> , 12(unico). https://doi.org/10.28998/rpss.v12iunico.15901
-------------------	--



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado

Conflito de interesses

Sem conflito de interesse

Financiamento

Sem apoio financeiro