



Editorial

Statistically Significant or Clinically Important?☆

¿Estadísticamente significativo o clínicamente relevante?



This editorial is the third and final installment of our 3-part series. The first explained the scientific method in order to contextualise the role of statistics, and discussed concepts of precision and internal and external validity.¹ Our aim in the second editorial was to make clear that a non-statistically significant difference is not synonymous with equivalence, and that treatments that are not statistically significantly different should not be considered “equal or equivalent”, for example, in a clinical trial comparing a new treatment with a standard treatment.²

This final editorial aims to show the difference between statistically significant and clinically important results.³

Let's take the example of a “primary endpoint”, which in this case is the change from baseline in forced expiratory volume in 1 second (FEV1) in the final visit,⁴ for example, in a comparative trial of 2 asthma treatments.

Supposing the clinical trial design is a randomized controlled trial with 2 parallel arms, we will obtain as a descriptive measure the mean of the FEV1 score in the intervention group and we will compare it with the mean in the control group to obtain the “difference of means between groups” in liters or milliliters (ml), which is a continuous quantitative variable as it can be measured in decimal points. In terms of statistical tests, the Student-Fisher t test is the most common procedure for comparing both means.

In Fig. 1, the threshold for the minimal clinically important difference (+100 ml) is highlighted with black bars, and the threshold for the classic null hypothesis of difference in ml = 0 is shown with a dashed-dotted line (highlighted in red in the electronic version of this editorial). Each confidence interval (CI) represents the lower and upper limits of the true difference of means in the population, in relation to the response to the treatments (with 95% confidence). In all four 95% CIs, the lower limit of each interval is greater than zero. If we divide the difference of means by the standard error, which we know that quantifies the role of chance in our results, we can see that the result would be a “Student's t” statistic greater than 1.96; with a p-value < 0.05. Thus, the four 95% CIs are statistically significant.

A “statistically significant” result only means that the standard error is small compared to the difference of means between groups. If the standard error is small enough, our results will always be statistically significant, but this does not necessarily imply that they

are clinically important. In our example, as mentioned above, a difference in FEV1 has to be at least 100 ml in order to have an impact on the quality of life of patients.

Continuing with the example of Fig. 1, we can see that while all 95% CIs are statistically significant, not all are conclusive in terms of the clinical importance of the treatment. Critical importance is interpreted based on the effect size (our difference of means), using the 95% CI limits in relation to the delta threshold corresponding to the “minimal clinically important difference” (MCID).^{5,6}

The first 95% CI (example 1) corresponds to a difference of +50 ml, with a lower limit and upper limit of +25 and +75 ml, respectively (taking into account the role of chance in our results with 95% confidence). In other words, the new treatment would be better than the standard treatment, but not enough to consider it clinically important, because its upper limit (+75 ml) is less than the minimal clinically important difference of +100 ml. This is the paradigmatic example of how something can be statistically significant, but clinically unimportant.

Remember that the standard error can be reduced by reducing variability or increasing sample size, so any difference in means that is not zero can become statistically significant. A totally unimportant difference of means of 0.05 ml will be statistically significant if, for example, we can manage to achieve an associated standard error of 0.025 ml.

Only one 95% CI (shown in example 4), in addition to being statistically significant, is conclusively clinically important because its lower limit is greater than 100 ml. In example 2, most of the interval agrees with the hypothesis that the difference is not clinically important. In example 3, most of the interval agrees that the effect is clinically important. However, as the two 95% CIs cross the threshold for the minimal clinically important difference, neither example would be clinically conclusive.

The same applies to measures of association. The clinical importance in this case must always be taken into account in the case of results close to 1 in measures of association, and especially in the case of odds ratios, because, as we know, this measure may overestimate the magnitude of the association compared to others such as the risk ratio.⁷

Therefore, in conclusion, any difference in means other than zero, or any measure of association other than 1, can become statistically significant by increasing the sample size or decreasing the variability of the data. A statistically significant result is not synonymous with a clinically important result, as not all statistically significant results will be clinically important, as this parameter must be interpreted on the basis of a specific threshold of clinical

☆ Please cite this article as: Santibáñez M, García-Rivero JL, Barreiro E. ¿Estadísticamente significativo o clínicamente importante? Arch Bronconeumol. 2020;56:615–616.

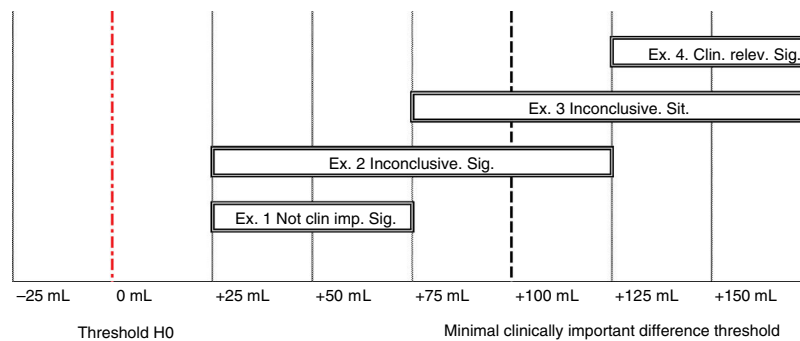


Fig. 1. Interpretation of the clinical importance of examples of the 95% confidence intervals, where the threshold for the minimum clinically important difference, has been set in a FEV1 difference of means greater than +100 mL. Sig. = statistically significant ($p < 0.05$).

importance. This problem and the different methodologies for setting a threshold for the minimal clinically important difference are of scientific interest^{8,9} but they go beyond the scope of this editorial. In the case of research on respiratory diseases, the symptom control questionnaires such as the Asthma Control Test (ACT) are worth remembering for their practical utility. Among their metrics, these validated tools report “responsiveness”,¹⁰ and establish the minimal clinically important difference which in this case is 3 points.¹¹

References

- Santibáñez M, García-Rivero JL, Barreiro E. No se debe empezar la casa por el tejado (si queremos publicar en unarevista de impacto). Arch Bronconeumol. 2020;56(2):70–1. <http://dx.doi.org/10.1016/j.arbres.2019.05.019>.
- Santibáñez M, García-Rivero JL, Barreiro E. p de significación: ¿mejor no usarla si se interpreta mal? Arch Bronconeumol. 2019. <http://dx.doi.org/10.1016/j.arbres.2019.11.003>.
- Chan KB, Man-Son-Hing M, Molnar FJ, Laupa-cis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. CMAJ. 2001;165:1197–202.
- Puente Maestú L, García de Pedro J. Lung function tests in clinical decision-making. Arch Bronconeumol. 2012;48:161–9.
- Kieser M, Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. Pharm Stat. 2005;4:101–7.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407–15.
- Viera AJ. Odds ratios and risk ratios: what’s the difference and why does it matter? South Med J. 2008;101:730–4.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77:371–83.
- Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. COPD. 2005;2:57–62.
- Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: reliability, validity, and responsiveness in patients not previously followed by asthma specialists. J Allergy Clin Immunol. 2006;117:549–56.
- Schatz M, Kosinski M, Yarlas AS, Hanlon J, Watson ME, Jhingran P. The minimally important difference of the Asthma Control Test. J Allergy Clin Immunol. 2009;124:719–23.

Miguel Santibáñez,^{a,*} Juan Luis García-Rivero,^b Esther Barreiro^{c,d,e}

^a Grupo de Investigación de Salud Global, Universidad de Cantabria, Instituto de Investigación Marqués de Valdecilla (IDIVAL), Santander, Cantabria, Spain

^b Servicio de Neumología, Hospital de Laredo, Laredo, Cantabria, Spain

^c Servicio de Neumología-Debilidad muscular y caquexia en las enfermedades respiratorias crónicas y el cáncer de pulmón, IMIM-Hospital del Mar, Barcelona, Spain

^d Departament de Ciències Experimentals i de la Salut (CEXS), Universitat Pompeu Fabra (UPF), Parc de Recerca Biomèdica de Barcelona (PRBB), Barcelona, Spain

^e Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III (ISCIII), Barcelona, Spain

* Corresponding author.

E-mail address: santibanezm@unican.es
(M. Santibáñez).