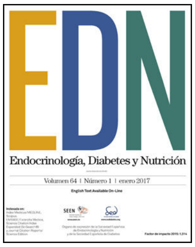
Endocrinol Diabetes Nutr. 2018;65(9):477–478

Endocrinología, Diabetes y Nutrición

www.elsevier.es/endo



EDITORIAL

It would be desirable to reduce the *p* value considered significant?☆

¿Sería conveniente reducir el valor p considerado significativo?

Ignasi Gich Saladich

Servicio de Epidemiología y Salud Pública, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

To be honest with the reader, I must point out that an important part of my work is based on the use of hypothesis testing and *p*-values; as a result, this text cannot be absolutely objective.

Suppose we have a group of subjects in whom we would like to determine whether a diet is able to modify their glycosylated hemoglobin (HbA1c) values. In order to study the effect, we would obtain the mean value before starting the diet and the mean value after it. Without resorting to more complicated analyses, the most usual procedure would be to compare the two mean values (with their standard deviations), for example: initial HbA1c value 6.8% and final value 5.9%, with comparison (in this case a repeated measures *t*-test) yielding a *p*-value of 0.026.

The question would be: Can we confirm that the diet has been beneficial? Clinically, the decrease in HbA1c is evident, and the result is moreover statistically significant. In other words, we have evidence that repetition of the study with another group of cases similar to those of our study, and using the same design with a similar number of cases, would yield similar values.

The cut-off point for rejecting equality is called the significance level, and is usually arbitrarily set at 5%. In our example, 0.026 is less than 0.05, which allows us to affirm that the difference was not unique to our subjects (in these individuals the difference is undeniable), and that it seems reasonable to assume that such a decrease would also be observed in the rest of the population from which our sample was extracted.

A more formal expression, usually found in the Material and Methods section of scientific articles, would be to state that the alpha value is set at 0.05. The lower the *p*-value, the greater the probability of rejecting the null hypothesis (i.e., equality), and thus the greater the probability of a true difference in what is being compared.

As I mentioned earlier, the value is arbitrary, and all guidelines explain that it can be modified, usually by lowering it.

It should be noted that the effect size (in the aforementioned example the decrease in HbA1c from 6.8% to 5.9%) is the most important datum and which always needs to be discussed, since we cannot limit ourselves to statistical considerations without also addressing the clinical relevance of the results.

Now we can remember the more formal definition of the *p*-value: the empirical probability of committing type I error, rejecting equality and accepting the existence of the difference in our comparison, even though we should not do so.

☆ Please cite this article as: Gich Saladich I. ¿Sería conveniente reducir el valor p considerado significativo? Endocrinol Diabetes Nutr. 2018;65:477-478.

E-mail address: igichs@santpau.cat

https://doi.org/10.1016/j.endien.2018.11.002

The main disadvantage of this approach is that it is not possible from this *p*-value to know the magnitude of the effect and therefore its clinical relevance, which as commented above is ultimately our true objective.

Furthermore, the mistaken interpretation that the *p*-value indicates the probability that H_0 (equality) is true leads us to draw wrong conclusions.

Talking about false interpretations, an article in which the last signing author was the recently deceased Dr. Altman (Grenland, 2016)[1] describes numerous errors and misinterpretations. Of all of them, I wish to highlight the statement that a significant result implies a clinically relevant outcome. Effect size cannot be interpreted from the *p*-value.

In view of the above, it has been repeatedly postulated over the years that we need to lower the significance level, as suggested by Ioannidis (JAMA, 2018).[2] However, lowering the *p*-value does not completely correct the problem. A proposed solution has been not to consider the *p*-value as significant or not, but to facilitate the effect size with its corresponding confidence interval, which allows us to assess the precision of the magnitude, affording a clearly more clinical perspective.

A recent initiative, signed by a large group of expert methodologists (Benjamin et al., 2018),[3] suggests a reduction in the commonly used significance level from 0.05 to 0.005. The underlying idea is to improve the reproducibility of the investigation, which suggests that the authors consider that investigators are not fully rigorous and that the results they obtain are not sufficiently robust to warrant the conclusions drawn. In this respect, Ioannidis wondered why most published research results are actually false (Ioannidis, 2005).[4]

A non-exhaustive list of the problems explaining the lack of validity of the conclusions, considering only the *p*-value, would include the elimination of values, the analysis of multiple variables, multiple comparisons, unforeseen comparisons, analyses of subgroups, expanding the sample until significant results are obtained, etc. In sum, there is a tendency to ''torture the data'' or to simplify interpretation of the results of a study based only on the *p*-value.

Reducing the significance level does not avoid the problem, but presumably makes it less prevalent, and is a simple and easy solution to use. In many of the commonly used statistical tests, setting the alpha value at 0.005, with the usual power of 80%, would require expanding the sample size by approximately 70% (Benjamin et al., 2018).[3] A significant advantage is the impact upon future research, since studies with few cases tend to exaggerate the estimates of effect size. Accordingly, the values which an investigator draws from a study with a large sample size will be more robust.

Another way of addressing the problem is to adopt the Bayesian approach, which involves adding to our analysis information from previous studies (called ''prior'' information), and which may improve the reproducibility of the conclusions, as evidenced by Nuzzo (2014).[5] This is a more complex approach that is hampered by the problem of the validity of the ''prior'' information used as reference.

As I have mentioned, reducing the *p*-value may in fact be distracting us from the true solution, as explained by the American Statistics Association (Wasserstein and Lazar, 2016).[6] This publication concluded that good statistical practice – as an essential element of good scientific practice – emphasizes the importance of good study design and conduction, variety of numerical summaries and data plots, understanding of the phenomenon being studied, interpretation of the results within context, complete reporting and adequate logical and quantitative understanding of what the data actually mean.

In view of the above, it seems more reasonable to choose a good design, with a clear explanation of the planned analyses and of the specific details used for sample size calculation – clearly indicating all exclusions (if any) and all calculations made, as well as the variables evaluated. In sum, the idea is to improve the quality and transparency of research, based on previous scientific protocols, with the adoption of a detailed statistical analytical plan.

No particular ''index'' should be allowed to replace scientific reasoning.

Acknowledgments

Thanks are due to Mr. Ivan Solà-Arnau, investigator of the Cochrane Latin American center in Barcelona, and to Dr. Carmen Fajardo, for her indications.

References

1. 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;31:337–50.
2. Ioannidis JPA. The proposal to lower *p* value thresholds to.005. JAMA. 2018;319:1429–30.
3. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018;2:6–10.
4. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2:e124.
5. Nuzzo R. Scientific method: statistical errors. Nature. 2014;506:150–2.
6. Wasserstein RL, Lazar NA. The ASA's statement on *p*-values: context, process, and purpose. Am Stat. 2016;70:129–33.