

Some Facts That You Might Be Unaware of About the P-Value

Inkyung Jung

Statistical Editor, Archives of Plastic Surgery
Department of Biostatistics and Medical Informatics, Yonsei University College of Medicine, Seoul, Korea



Recently, some articles on the P-value published in biomedical journals caught my eye. These articles included the words ‘misuse,’ ‘misconception,’ or ‘misinterpretation’ of the P-value in their titles or abstracts. Nuzzo [1] stated that P-values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume and added that ‘the P-value was never meant to be used the way it’s used today.’ Greenland et al. [2] provided an explanatory list of 18 misinterpretations of P-values and guidelines for improving statistical interpretation. Other interesting articles include a paper by Chavalarias et al. [3] and the accompanying editorial by Kyriacou [4] published in the *Journal of the American Medical Association*, and a review paper by van Rijn et al. [5]. Not only in biomedical communities, but also in a major statistical society, a statement on P-values has been issued in order to sound the alarm about the misuse of P-values [6].

Having been a statistician at medical schools for more than 10 years, I understand very well how obsessed many researchers are with P-values regarding the results of their studies. I have seen many times that some researchers try to obtain a P-value less than 0.05 from a P-value slightly greater than 0.05, such as 0.053 or 0.06, by deleting or adding some data. This happens, I think, because many researchers often believe that results with a P-value < 0.05, which is considered to be ‘statistically significant,’ are truly scientifically or substantially significant. However, this is one of the most notorious misinterpretations of the P-value.

With this in mind, how can we correctly interpret P-values? To answer this question, we should understand what a P-value really is. Informally, a P-value is the probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value [6]. I admit that this definition is not easy to understand. An important

point is that it should not be used as a definitive decision-making tool, yielding outcomes of yes or no. A P-value is a *probability*. It is a measure of summarizing the incompatibility between a particular set of data and a proposed model for the data (the null hypothesis) [6]. Ronald Fisher, who introduced the P-value, intended it as an informal way to judge whether evidence was significant in the sense of being worthy of a second look [1]. A very small P-value indicates that the null hypothesis is very incompatible with the data that have been collected. However, we cannot say with certainty that the null hypothesis is not true, or that the alternative hypothesis must be true [5].

Another important fact is that the P-value has nothing to do with the magnitude or the importance of an observed effect [5]. I have been surprised to see that many researchers interpret a result with a risk ratio of 0.59 with a P-value of 0.16 as non-significant or ‘no difference,’ while stating that a risk ratio of 0.83 with a P-value of 0.002 is highly significant. As argued by Wasserstein and Lazar [6], statistical significance is not equivalent to scientific, human, or economic significance. One must not interpret the results solely by the P-value. A small P-value could be simply due to a very large sample size regardless of the effect size. A P-value > 0.05 does not mean that no effect was observed, or that the effect size was small. One must look at the effect size and uncertainty measures (e.g., standard error and confidence interval) to evaluate whether the results are clinically or scientifically relevant.

Now, how do we avoid misusing P-values? One journal even prohibits P-values, saying that the null hypothesis significance testing procedure is invalid [7]. I think this is a rather extreme action. Still, it is collateral evidence of the rampant misuse of the P-value. As many statisticians have said, the P-value itself is not

the problem. One should clearly understand what a P-value really means and should not judge the results of a study or experiment by relying only on the P-value. There are other approaches that can supplement P-values, such as confidence, credibility, or prediction intervals; Bayesian methods; and alternative measures of evidence such as likelihood ratios or Bayes factors [6]. Chavalarias et al. [3] also recommended that rather than reporting isolated P-values, articles should include effect sizes and uncertainty metrics. I strongly encourage the readers of this article to read the papers listed in the references (and other relevant papers as well), which will lead them to a deeper understanding of P-values and other important statistical concepts.

REFERENCES

1. Nuzzo R. Scientific method: statistical errors. *Nature* 2014;506:150-2.
2. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337-50.
3. Chavalarias D, Wallach JD, Li AH, et al. Evolution of reporting p values in the biomedical literature, 1990-2015. *JAMA* 2016;315:1141-8.
4. Kyriacou DN. The enduring evolution of the p value. *JAMA* 2016;315:1113-5.
5. van Rijn MH, Bech A, Bouyer J, et al. Statistical significance versus clinical relevance. *Nephrol Dial Transplant* 2017 Jan 7 [Epub]. <http://doi.org/10.1093/ndt/gfw385>.
6. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016;70:129-33.
7. Trafimow D, Marks M. Editorial. *Basic Appl Soc Psych* 2015;37:1-2.

Correspondence: Inkyung Jung
Department of Biostatistics and Medical Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-2-2228-2494, Fax: +82-2-364-8037, E-mail: ijung@yuhs.ac

No potential conflict of interest relevant to this article was reported.

Received: 10 Mar 2017 • Revised: 11 Mar 2017 • Accepted: 11 Mar 2017
pISSN: 2234-6163 • eISSN: 2234-6171
<https://doi.org/10.5999/aps.2017.44.2.93> • Arch Plast Surg 2017;44:93-94
