



Why say “statistically significant” rather than just “significant” —a plea to rid the medical literature of linguistic ambiguity (a secondary publication)

Abe Fingerhut^{1,2}, Steve Wexner³, Kevin Behrns⁴, Alberto Arrezo⁵, Leo Buhler⁶, Nader Francis⁷,
Deborah S. Keller⁸, Wolfram Knoefel⁹, Paulina Salminen¹⁰, Lee Swanstrom^{11,12}, Des Winter¹³

¹Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine and Shanghai Minimally Invasive Surgery Center, Shanghai, China; ²Section for Surgical Research, Department of Surgery, Medical University of Graz, Graz, Austria; ³Ellen Leifer Shulman and Steven Shulman Digestive Disease Center, Cleveland Clinic Florida, Weston, FL, USA; ⁴University of Florida, Gainesville, FL, USA; ⁵Department of Surgical Sciences, University of Turin, Turin, Italy; ⁶University of Fribourg, Fribourg, Switzerland; ⁷Division of Surgery and Interventional Science, University College London, London, UK; ⁸Marks Colorectal Surgical Associates, Lankenau Medical Center, Wynnewood, PA, USA; ⁹University Hospital Dusseldorf, Dusseldorf, Germany; ¹⁰Division of Digestive Surgery and Urology, Department of Surgery, University of Turku, Turku University Hospital, Turku, Finland; ¹¹Oregon Health and Science University, Portland, OR, USA; ¹²IHU-Strasbourg, Strasbourg, France; ¹³St. Vincent's Private Hospital Dublin, Dublin, Ireland

Correspondence to: Abe Fingerhut, MD. Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine and Shanghai Minimally Invasive Surgery Center, Shanghai, China; Section for Surgical Research, Department of Surgery, Medical University of Graz, Graz, Austria. Email: abefingerhut@aol.com.

Keywords: Statistical significance; statistical test; linguistic ambiguity; medical writing; literature

Received: 09 August 2022; Accepted: 16 December 2022; Published online: 12 January 2023.

doi: 10.21037/ales-22-78

View this article at: <https://dx.doi.org/10.21037/ales-22-78>

As scientists and academics, we teach, conduct research and write scientific research papers (1). Whenever we try to determine whether an outcome, an exposure, or a difference between two or more variables, is meaningful, we perform a statistical test to determine the probability of our finding being due just to chance, or, of our finding having the property of being “statistically significant”, using the “P value”, or some alternative such as confidence, credibility, or prediction intervals.

The word “significant” in relation to statistical testing seems to have been first employed by the British economist and statistician Francis Edgeworth in the 1880's (2). Edgeworth used the word “significant” to mean “signifying” something causal and not accidental. In Ronald Fisher's famous description of the “tea test” (3), he stated that an outcome of an experiment “signified” or could help to interpret a result. Widely used thereafter by Karl Pearson and his mentees, slowly, but steadily, the statistical meaning of “significant” shifted and both the word “significant” and P values have—incorrectly—become synonymous with

“important” (4).

All too frequently, writers of scientific papers omit the adjective or adverb “statistical” or “statistically”, respectively, and we see a sentence with the word “significant”, standing alone. Our hypothesis is that stating that an outcome or a difference is “significant” without the modifier “statistically” is lexicographically incorrect and misleading because of the linguistic ambiguity created by the double meaning in the word “significant”.

There are two reasons why “statistical” or “statistically” should always be used with the words “significance” or “significant” in scientific papers. The first is that the P value corresponds to the result of a test—a statistical test. The P value is not a probability of something being true, or important, or a property or characteristic of the effect or population being studied. Assimilated to be the probability that the result (or difference) is not due to chance, statistical significance means, more scientifically speaking, if the “null hypothesis” (there really is no difference) is true, there is a low probability (usually set at 0.05 or less) of obtaining a

result (or difference) that large or larger. Whether we use P values or some alternative, the significance that we want to highlight is statistical significance.

The second is that confusion arises when writers omit “statistical”/“statistically”, thinking that everyone knows what they mean, i.e., “significant” in the statistical vernacular. However, without the modifier “statistical”/“statistically”, linguistic ambiguity of the word “significant” automatically and unconsciously invades the mind of the reader, who all too often takes the short cut of interpreting the result or difference as a characteristic, or an absolute, resulting from causality and sound methodology, whether a test, or even a P value, was provided or not. Indeed, the definition of the word “significant” as “being important” or “proven” is inevitably what the reader retains subconsciously when reading the word “significant” alone.

While we are aware of the arguments against the use of P values altogether, the current medical literature is not ready for change, at least, not in the near future: the phraseology concerning “statistical significance” is not going to disappear for many years to come.

Consequently, we would like to strongly suggest avoiding the use of the words “significant”, “significantly”, or “significance” in medical writing other than to designate “statistical significance”, and to prevent any possible confusion, the modifier, “statistical” or “statistically” should always be attached.

Acknowledgments

This article is the fruit of reflections and comments made during the Webinar on Peer Review sponsored by the journal *Surgery* on June 28, 2022 (available online <https://elsevier.zoom.us/rec/share/qlxM13AwjPKlofc4sUf4-Ed983rOKT8BtDxKhNOfhMGWjZCDQb9mlV2ZMDy8R7GU.OsSpGDve2d7VotPF?startTime=1656425313000,passcode:6LSC%20H4F>). It was originally published in *Surgery* (2022;172:1039-40. doi: 10.1016/j.surg.2022.08.019). In order to encourage dissemination, this article is granted the permission from all authors and the journal *Surgery* to be published in *Annals of Laparoscopic and Endoscopic Surgery*. Members of the Faculty of the Webinar (by alphabetical order): Alberto Arezzo, MD, PhD, Editor of the *Journal Minimally Invasive Therapy and Allied Technologies (MITAT)*, Associate Editor of the *Journal Techniques in Coloproctology (TCOL)*; Leo Buhler, MD, Editor in Chief of *Xenotransplantation*; Abe Fingerhut, MD, Co-editor in

Chief of *Annals of Laparoscopic and Endoscopic Surgery*; Nader Francis, FRCS, FEBS, PhD, Education and Training Subject Editor of *Surgical Endoscopy*; Susan Galandiuk, MD, Editor-In-Chief, *Diseases of the Colon & Rectum*; Wolfram Trudo Knoefel, MD, Member of Editorial Board of *Surgery*; Paulina Salminen, MD, PhD, Editor-in-Chief, *Scandinavian Journal of Surgery*; Lee Swanstrom, MD, Editor-in-chief of *Surgical Innovation*; Des Winter, MD, Editor-in-chief of *British Journal of Surgery (BJS)*.

Funding: None.

Footnote

Provenience and Peer Review: This article was a standard submission to the journal. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://ales.amegroups.com/article/view/10.21037/ales-22-78/coif>). AF serves as the co-Editor-in-Chief of *Annals of Laparoscopic and Endoscopic Surgery*. SW reports receiving consulting fees from ARC/Corvus, Astellas, Baxter, Becton Dickinson, GI Supply, ICON Language Services, Intuitive Surgical, Leading BioSciences, Livsmed, Medtronic, Olympus Surgical, Stryker, Takeda and receiving royalties from Intuitive Surgical, Karl Storz Endoscopy America Inc, Medtronic, Unique Surgical Innovations, LLC. KB reports that he serves as the co-Editor-in-Chief for *Surgery* and receives a stipend for the work. PS reports that she receives research grants from Sigríd Jusélius Foundation and Finnish Academy, and participation on board of DSMC BEST trial (sleeve vs. bypass) and Magnet study (magnetic anastomosis). LS reports that he received stipend as Editor-in-Chief of *Surgical Innovation*. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Rosengart TK, Mason MC, LeMaire SA, et al. The seven attributes of the academic surgeon: Critical aspects of the archetype and contributions to the surgical community. *Am J Surg* 2017;214:165-79.
2. Shafer G. On the nineteenth-century origins of significance testing and p-hacking. Available online: <http://www.probabilityandfinance.com>
3. Salsburg D. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. New York, NY, USA: 2001.
4. Best AM 3rd, Greenberg BL, Glick M. From tea tasting to t test: a P value ain't what you think it is. *J Am Dent Assoc* 2016;147:527-9. Erratum in: *J Am Dent Assoc* 2017;148:287.

doi: 10.21037/ales-22-78

Cite this article as: Fingerhut A, Wexner S, Behrns K, Arrezo A, Buhler L, Francis N, Keller DS, Knoefel W, Salminen P, Swanstrom L, Winter D. Why say “statistically significant” rather than just “significant”—a plea to rid the medical literature of linguistic ambiguity (a secondary publication). *Ann Laparosc Endosc Surg* 2023;8:8.