

EDITORIAL



The clinician's guide to p values, confidence intervals, and magnitude of effects

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2021, corrected publication 2021

Eye (2022) 36:341–342; <https://doi.org/10.1038/s41433-021-01863-w>

INTRODUCTION

There are numerous statistical and methodological considerations within every published study, and the ability of clinicians to appreciate the implications and limitations associated with these key concepts is critically important. These implications often have a direct impact on the applicability of study findings – which, in turn, often determine the appropriateness for the results to lead to modification of practice patterns. Because it can be challenging and time-consuming for busy clinicians to break down the nuances of each study, herein we provide a brief summary of 3 important topics that every ophthalmologist should consider when interpreting evidence.

P-VALUES: WHAT THEY TELL US AND WHAT THEY DON'T

Perhaps the most universally recognized statistic is the p -value. Most individuals understand the notion that (usually) a p -value < 0.05 signifies a statistically significant difference between the two groups being compared. While this understanding is shared amongst most, it is far more important to understand what a p -value does not tell us. Attempting to inform clinical practice patterns through interpretation of p -values is overly simplistic, and is fraught with potential for misleading conclusions. A p -value represents the probability that the observed result (difference between the groups being compared)—or one that is more extreme—would occur by random chance, assuming that the null hypothesis (the alternative scenario to the study's hypothesis) is that there are no differences between the groups being compared. For example, a p -value of 0.04 would indicate that the difference between the groups compared would have a 4% chance of occurring by random chance. When this probability is small, it becomes less likely that the null hypothesis is accurate—or, alternatively, that the probability of a difference between groups is high [1]. Studies use a predefined threshold to determine when a p -value is sufficiently small enough to support the study hypothesis. This threshold is conventionally a p -value of 0.05; however, there are reasons and justifications for studies to use a different threshold if appropriate.

What a p -value cannot tell us, is the clinical relevance or importance of the observed treatment effects. [1]. Specifically, a p -value does not provide details about the magnitude of effect [2–4]. Despite a significant p -value, it is quite possible for the difference between the groups to be small. This phenomenon is especially common with larger sample sizes in which comparisons may result in statistically significant differences that are actually not clinically meaningful. For example, a study may find a statistically significant difference ($p < 0.05$) between the visual

acuity outcomes between two groups, while the difference between the groups may only amount to a 1 or less letter difference. While this may be in fact a statistically significant difference, the difference is likely not large enough to make a meaningful difference for patients. Thus, p -values lack vital information on the magnitude of effects for the assessed outcomes [2–4].

OVERCOMING THE LIMITATIONS OF INTERPRETING P-VALUES: MAGNITUDE OF EFFECT

To overcome this limitation, it is important to consider both (1) whether or not the p -value of a comparison is significant according to the pre-defined statistical plan, and (2) the magnitude of the treatment effects (commonly reported as an effect estimate with 95% confidence intervals) [5]. The magnitude of effect is most often represented as the mean difference between groups for continuous outcomes, such as visual acuity on the logMAR scale, and the risk or odds ratio for dichotomous/binary outcomes, such as occurrence of adverse events. These measures indicate the observed effect that was quantified by the study comparison. As suggested in the previous section, understanding the actual magnitude of the difference in the study comparison provides an understanding of the results that an isolated p -value does not provide [4, 5]. Understanding the results of a study should shift from a binary interpretation of significant vs not significant, and instead, focus on a more critical judgement of the clinical relevance of the observed effect [1].

There are a number of important metrics, such as the Minimally Important Difference (MID), which helps to determine if a difference between groups is large enough to be clinically meaningful [6, 7]. When a clinician is able to identify (1) the magnitude of effect within a study, and (2) the MID (smallest change in the outcome that a patient would deem meaningful), they are far more capable of understanding the effects of a treatment, and further articulate the pros and cons of a treatment option to patients with reference to treatment effects that can be considered clinically valuable.

THE ROLE OF CONFIDENCE INTERVALS





Confidence intervals are estimates that provide a lower and upper threshold to the estimate of the magnitude of effect. By convention, 95% confidence intervals are most typically reported. These intervals represent the range in which we can, with 95% confidence, assume the treatment effect to fall within. For example, a mean difference in visual acuity of 8 (95% confidence interval: 6 to 10) suggests that the best estimate of the difference between the two study groups is 8 letters, and we have 95% certainty that the true value is between 6 and 10 letters. When interpreting this clinically, one can consider the different clinical scenarios at each end of the confidence interval; if the patient's

Received: 11 November 2021 Revised: 12 November 2021 Accepted: 15 November 2021
Published online: 26 November 2021

outcome was to be the most conservative, in this case an improvement of 6 letters, would the importance to the patient be different than if the patient's outcome was to be the most optimistic, or 10 letters in this example? When the clinical value of the treatment effect does not change when considering the lower versus upper confidence intervals, there is enhanced certainty that the treatment effect will be meaningful to the patient [4, 5]. In contrast, if the clinical merits of a treatment appear different when considering the possibility of the lower versus the upper confidence intervals, one may be more cautious about the expected benefits to be anticipated with treatment [4, 5].

CONCLUSION




There are a number of important details for clinicians to consider when interpreting evidence. Through this editorial, we hope to provide practical insights into fundamental methodological principals that can help guide clinical decision making. *P*-values are one small component to consider when interpreting study results, with much deeper appreciation of results being available when the treatment effects and associated confidence intervals are also taken into consideration.

Mark R. Phillips ^{1,1}, Charles C. Wykoff^{2,3}, Lehana Thabane ^{1,4}, Mohit Bhandari ^{1,5}, Varun Chaudhary ^{1,5}✉ and for the Retina Evidence Trials InterNational Alliance (R.E.T.I.N.A.) Study Group*
¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada. ²Retina Consultants of Texas (Retina Consultants of America), Houston, TX, USA. ³Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA. ⁴Biostatistics Unit, St. Joseph's Healthcare-Hamilton, Hamilton, ON, Canada. ⁵Department of Surgery, McMaster University, Hamilton, ON, Canada. *A list of authors and their affiliations appears at the end of the paper. ✉email: vchaudh@mcmaster.ca

REFERENCES

- Li G, Walter SD, Thabane L. Shifting the focus away from binary thinking of statistical significance and towards education for key stakeholders: revisiting the debate on whether it's time to de-emphasize or get rid of statistical significance. *J Clin Epidemiol*. 2021;137:104–12. <https://doi.org/10.1016/j.jclinepi.2021.03.033>
- Gagnier JJ, Morgenstern H. Misconceptions, misuses, and misinterpretations of p values and significance testing. *J Bone Joint Surg Am*. 2017;99:1598–603. <https://doi.org/10.2106/JBJS.16.01314>
- Goodman SN. Toward evidence-based medical statistics. 1: the p value fallacy. *Ann Intern Med*. 1999;130:995–1004. <https://doi.org/10.7326/0003-4819-130-12-199906150-00008>
- 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. <https://doi.org/10.1007/s10654-016-0149-3>

FOR THE RETINA EVIDENCE TRIALS INTERNATIONAL ALLIANCE (R.E.T.I.N.A.) STUDY GROUP

Varun Chaudhary ^{1,5}✉, Mohit Bhandari ^{1,5}, Charles C. Wykoff^{2,3}, Sobha Sivaprasad⁶, Lehana Thabane ^{1,4}, Peter Kaiser⁷, David Sarraf⁸, Sophie J. Bakri⁹, Sunir J. Garg¹⁰, Rishi P. Singh^{11,12}, Frank G. Holz¹³, Tien Y. Wong^{14,15} and Robyn H. Guymer^{16,17}

⁶NIHR Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, UK. ⁷Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ⁸Retinal Disorders and Ophthalmic Genetics, Stein Eye Institute, University of California, Los Angeles, CA, USA. ⁹Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA. ¹⁰The Retina Service at Wills Eye Hospital, Philadelphia, PA, USA. ¹¹Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ¹²Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. ¹³Department of Ophthalmology, University of Bonn, Bonn, Germany. ¹⁴Singapore Eye Research Institute, Singapore, Singapore. ¹⁵Singapore National Eye Centre, Duke-NUD Medical School, Singapore, Singapore. ¹⁶Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia. ¹⁷Department of Surgery (Ophthalmology), The University of Melbourne, Melbourne, VIC, Australia.

- Phillips M. Letter to the editor: editorial: threshold p values in orthopaedic research—we know the problem. What is the solution? *Clin Orthop*. 2019;477:1756–8. <https://doi.org/10.1097/CORR.0000000000000827>
- Devji T, Carrasco-Labra A, Qasim A, Phillips MR, Johnston BC, Devasevnapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020;369:m1714. <https://doi.org/10.1136/bmj.m1714>
- Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, et al. Minimal important difference estimates for patient-reported outcomes: a systematic survey. *J Clin Epidemiol*. 2020;0. <https://doi.org/10.1016/j.jclinepi.2020.11.024>

AUTHOR CONTRIBUTIONS

MRP was responsible for conception of idea, writing of manuscript and review of manuscript. VC was responsible for conception of idea, writing of manuscript and review of manuscript. MB was responsible for conception of idea, writing of manuscript and review of manuscript. CCW was responsible for critical review and feedback on manuscript. LT was responsible for critical review and feedback on manuscript.

COMPETING INTERESTS

MRP: Nothing to disclose. CCW: Consultant: Acuela, Adverum Biotechnologies, Inc, Aerieo, Alimera Sciences, Allegro Ophthalmics, LLC, Allergan, Apellis Pharmaceuticals, Bayer AG, Chengdu Kanghong Pharmaceuticals Group Co, Ltd, Clearside Biomedical, DORC (Dutch Ophthalmic Research Center), EyePoint Pharmaceuticals, Genentech/Roche, GyroscopeTx, IVERIC bio, Kodiak Sciences Inc, Novartis AG, ONL Therapeutics, Oxurion NV, PolyPhotonix, Recens Medical, Regeron Pharmaceuticals, Inc, REGENXBIO Inc, Santen Pharmaceutical Co, Ltd, and Takeda Pharmaceutical Company Limited; Research funds: Adverum Biotechnologies, Inc, Aerie Pharmaceuticals, Inc, Aerieo, Alimera Sciences, Allergan, Apellis Pharmaceuticals, Chengdu Kanghong Pharmaceutical Group Co, Ltd, Clearside Biomedical, Gemini Therapeutics, Genentech/Roche, Graybug Vision, Inc, GyroscopeTx, Ionis Pharmaceuticals, IVERIC bio, Kodiak Sciences Inc, Neurotech LLC, Novartis AG, Opthea, Outlook Therapeutics, Inc, Recens Medical, Regeneron Pharmaceuticals, Inc, REGENXBIO Inc, Samsung Pharm Co, Ltd, Santen Pharmaceutical Co, Ltd, and Xbrane Biopharma AB—unrelated to this study. LT: Nothing to disclose. MB: Research funds: Pendopharm, Bioventus, Acumed – unrelated to this study. VC: Advisory Board Member: Alcon, Roche, Bayer, Novartis; Grants: Bayer, Novartis – unrelated to this study.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Varun Chaudhary.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.