EDITORIALS

New Guidelines for Statistical Reporting in the *Journal*

David Harrington, Ph.D., Ralph B. D'Agostino, Sr., Ph.D., Constantine Gatsonis, Ph.D., Joseph W. Hogan, Sc.D., David J. Hunter, M.B., B.S., M.P.H., Sc.D., Sharon-Lise T. Normand, Ph.D., Jeffrey M. Drazen, M.D., and Mary Beth Hamel, M.D., M.P.H

Some *Journal* readers may have noticed more parsimonious reporting of P values in our research articles over the past year. For example, in November 2018, we published two reports from the Vitamin D and Omega-3 Trial (VITAL),[1,2] a two-by-two factorial, placebo-controlled, randomized trial assessing whether vitamin D or marine n–3 (also known as omega-3) fatty acids prevent cardiovascular disease or cancer. For the n–3 portion of the trial, Manson et al.[2] reported 2 prespecified primary outcomes and 22 prespecified and other secondary outcomes — not uncommon in large, expensive randomized or observational studies. The n–3 fatty acids did not significantly reduce the rate of either the primary cardiovascular outcome or the cancer outcome. If reported as independent findings, the P values for two of the secondary outcomes would have been less than 0.05; however, the article reported only the hazard ratios and confidence intervals for the intervention effects for those secondary outcomes, consistent with recently implemented *Journal* guidelines limiting the use of P values for secondary and other comparisons.

We have now clarified, expanded, and refined our statistical guidelines for authors to cover both clinical trials and observational studies. The new guidelines discuss many aspects of the reporting of studies in the *Journal*, including a requirement to replace P values with estimates of effects or association and 95% confidence intervals when neither the protocol nor the statistical analysis plan has specified methods used to adjust for multiplicity. *Journal* editors and statistical consultants have become increasingly concerned about the overuse and misinterpretation of significance testing and P values in the medical literature. Along with their strengths, P values are subject to inherent weaknesses, as summarized in recent publications from the American Statistical Association.[3,4]

P values indicate how incompatible the observed data may be with a null hypothesis; "P<0.05" implies that a treatment effect or exposure association larger than that observed would occur less than 5% of the time under a null hypothesis of no effect or association and assuming no confounding. Concluding that the null hypothesis is false when in fact it is true (a type I error in statistical terms) has a likelihood of less than 5%. When P values are reported for multiple outcomes without adjustment for multiplicity, the probability of declaring a treatment difference when none exists can be much higher than 5%. When 10 tests are conducted, the probability that at least one of the 10 will have a P value less than 0.05 may be as high as 40% when the null hypothesis of no difference is true. Even when no adjustment for multiplicity is needed, P values do not represent the probability that the null hypothesis is false: P<0.05 does not imply that the probability of the null hypothesis is less than 5%. Because P values provide no information about the variability of an estimated association (its standard error), nonsignificant P values do not distinguish between group differences that are truly negligible and group differences that are noninformative because of large standard errors. P values provide no information about the size of an effect or an association.

The use of P values to summarize evidence in a study requires, on the one hand, thresholds that have a strong theoretical and empirical justification and, on the other hand, proper attention to the error that can result from uncritical interpretation of multiple inferences.[5] This inflation due to multiple comparisons can also occur when comparisons have been conducted by investigators but are not reported in a manuscript. A large array of methods to adjust for multiple comparisons is available and can be used to control the type I error probability in an analysis when specified in the design of a study.[6,7] Finally, the

The New England Journal of Medicine
Downloaded from nejm.org at IDAHO STATE UNIVERSITY on July 17, 2019. For personal use only. No other uses without permission.
Copyright © 2019 Massachusetts Medical Society. All rights reserved.

notion that a treatment is effective for a particular outcome if P<0.05 and ineffective if that threshold is not reached is a reductionist view of medicine that does not always reflect reality.

Despite the difficulties they pose, P values continue to have an important role in medical research, and we do not believe that P values and significance tests should be eliminated altogether. A well-designed randomized or observational study will have a primary hypothesis and a prespecified method of analysis, and the significance level from that analysis is a reliable indicator of the extent to which the observed data contradict a null hypothesis of no association between an intervention or an exposure and a response. Clinicians and regulatory agencies must make decisions about which treatment to use or to allow to be marketed, and P values interpreted by reliably calculated thresholds subjected to appropriate adjustments have a role in those decisions.

The *Journal*'s revised policies on P values rest on three premises: it is important to adhere to a prespecified analysis plan if one exists; the use of statistical thresholds for claiming an effect or association should be limited to analyses for which the analysis plan outlined a method for controlling type I error; and the evidence about the benefits and harms of a treatment or exposure should include both point estimates and their margins of error.

We acknowledge that our new guidelines may present challenges in their use and interpretation, especially for authors and readers who are accustomed to thinking of P values or confidence intervals as a bright-line marker for a conclusion or a claim. We also understand that the results reported in a manuscript submitted to the *Journal* today may have come from a trial designed a decade ago. We are willing to work with authors within our new guidelines to maintain appropriate reporting of results. Finally, the current guidelines are limited to studies with a traditional frequentist design and analysis, since that matches the large majority of manuscripts submitted to the *Journal*. We do not mean to imply that these are the only acceptable designs and analyses. The *Journal* has published many studies with Bayesian designs and analyses[8-10] and expects to see more such trials in the future. When appropriate, our guidelines will be expanded to include best practices for reporting trials with Bayesian and other designs.

Disclosure forms provided by the authors are available with the full text of this editorial at NEJM.org.

From the Department of Data Sciences, Dana–Farber Cancer Institute (D.H.), Boston University (R.B.D.), Harvard T.H. Chan School of Public Health (D.H., D.J.H.), and the Department of Health Care Policy, Harvard Medical School, and the Department of Biostatistics, Harvard T.H. Chan School of Public Health (S.-L.T.N.) — all in Boston; the Department of Biostatistics and Center for Statistical Sciences, Brown University School of Public Health, Providence, RI (C.G., J.W.H.); and Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom (D.J.H.).

1. Manson JE, Cook NR, Lee I-M, et al. Vitamin D supplements and prevention of cancer and cardiovascular disease. N Engl J Med 2019;380:33-44.
2. Manson JE, Cook NR, Lee I-M, et al. Marine n–3 fatty acids and prevention of cardiovascular disease and cancer. N Engl J Med 2019;380:23-32.
3. Wasserstein RL, Lazar NA. The ASA's statement on *p*-values: context, process, and purpose. Am Stat 2016;70:129-33.
4. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "*p* < 0.05." Am Stat 2019;73:Suppl 1:1-19.
5. National Academies of Sciences, Engineering, and Medicine. Reproducibility and replicability in science. Washington, DC: National Academies Press, 2019.
6. Dmitrienko A, Bretz F, Westfall PH, et al. Multiple testing methodology. In: Dmitrienko A, Tamhane AC, Bretz F, eds. Multiple testing problems in pharmaceutical statistics. New York: Chapman and Hall/CRC Press, 2009:35-98.
7. Dmitrienko A, D'Agostino RB Sr. Multiplicity considerations in clinical trials. N Engl J Med 2018;378:2115-22.
8. Rugo HS, Olopade OI, DeMichele A, et al. Adaptive randomization of veliparib–carboplatin treatment in breast cancer. N Engl J Med 2016;375:23-34.
9. Park JW, Liu MC, Yee D, et al. Adaptive randomization of neratinib in early breast cancer. N Engl J Med 2016;375:11-22.
10. Popma JJ, Deeb GM, Yakubov SJ, et al. Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. N Engl J Med 2019;380:1706-15.

DOI: 10.1056/NEJMe1906559
Copyright © 2019 Massachusetts Medical Society.

HIV-1 Epidemic Control — Insights from Test-and-Treat Trials

Salim S. Abdool Karim, M.B., Ch.B., Ph.D.

Debate ensued when mathematical models[1] predicted that universal testing and treatment could achieve epidemic control within a few years in a high-burden setting. However, the idea rapidly gained new credence when antiretroviral therapy (ART)–induced viral suppression was shown to be highly effective in preventing transmission of human immunodeficiency virus type 1 (HIV-1) in a clinical trial.[2]

Four large, cluster-randomized, controlled trials, each with a different approach to maximizing viral suppression, set out to assess the effect of universal testing and treatment on community HIV incidence. In 2018, the Treatment as Prevention (TasP) trial[3] involving 28,419 persons in 22 communities showed no effect on HIV transmission in a rural South African district in which earlier

The New England Journal of Medicine
Downloaded from nejm.org at IDAHO STATE UNIVERSITY on July 17, 2019. For personal use only. No other uses without permission.
Copyright © 2019 Massachusetts Medical Society. All rights reserved.