has equity in Virta Health; and serves on the advisory board of Simply Good Foods. ECW received consulting fees from Hill United Health and founded Adapt Your Life, Inc (equity interest)—both companies founded on low-carbohydrate-diet principles; and received royalties for books that recommend a carbohydrate-restricted diet. WSY received grants to study low-carbohydrate (and other) eating patterns; consulted for Guideline Central on a clinical guideline about low-carbohydrate nutrition; and serves as scientific advisor for dietdoctor.com. CBE received grants to study the carbohydrate-insulin model from the National Institutes of Health (USA) and philanthropies unaffiliated with the food industry. The other authors report no conflicts of interest.

David S Ludwig
Louis J Aronne
Arne Astrup
Rafael de Cabo
Lewis C Cantley
Mark I Friedman
Steven B Heymsfield
James D Johnson
Janet C King
Ronald M Krauss
Daniel E Lieberman
Gary Taubes
Jeff S Volek
Eric C Westman
Walter C Willett
William S Yancy, Jr
Cara B Ebbeling

From the New Balance Foundation Obesity Prevention Center, Boston Children's Hospital, MA, USA (CBE; DSL, e-mail: david.ludwig@childrens.harvard.edu); Department of Pediatrics, Harvard Medical School, Boston, MA, USA (DSL, CBE); Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA (DSL, WCW); Comprehensive Weight Control Center, Weill Cornell Medicine, New York, NY, USA (LJA); Department of Nutrition, Exercise and Sports, University of Copenhagen, Denmark (AA); Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA (RdC); Department of Medicine, Weill Cornell Medicine, New York, NY, USA (LCC); Monell Chemical Senses Center, Philadelphia, PA, USA (MIF); Nutrition Science Initiative, San Diego, CA, USA (MIF, GT); Metabolism & Body Composition Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA (SBH); Diabetes Research Group, Life Sciences Institute, University of British Columbia, Vancouver, Canada (JDJ); Institute for Personalized Therapeutic Nutrition, Vancouver, Canada (JDJ); Department of Nutritional Sciences & Toxicology, University of California Berkeley, Berkeley, CA, USA (JCK); Departments of Pediatrics and Medicine, University of California San Francisco, San Francisco, CA, USA (RMK); Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA (DEL); Department of Human Sciences, Ohio State University, Columbus, OH, USA (JSV); and Department of Medicine, Duke University School of Medicine, Durham, NC, USA (ECW, WSY). .

References

1. Ludwig DS, Aronne LJ, Astrup A, de Cabo R, Cantley LC, Friedman MI, Heymsfield SB, Johnson JD, King JC, Krauss RM, et al. The carbohydrate-insulin model: a physiological perspective on the obesity pandemic. Am J Clin Nutr 2021;114(6):1873–85.
2. Schwartz MW, Seeley RJ, Zeltser LM, Drewnowski A, Ravussin E, Redman LM, Leibel RL. Obesity pathogenesis: an Endocrine Society Scientific Statement. Endocr Rev 2017;38(4):267–96.
3. American Association for the Advancement of Science (AAAS). EurekAlert! Insulin regulates body fat independent of dietary carbohydrates. Science [Internet] Available from: https://www.eurekalert.org/news-releases/919847 (accessed 3 November 2021).
4. Ludwig DS. Second opinion: let's focus more on what we eat, not how much. Medpage Today. 18 October 2021 [Internet]. Available from: https://www.medpagetoday.com/opinion/second-opinions/95104.
5. Toates FM, Booth DA. Control of food intake by energy supply. Nature 1974;251(5477):710–1.
6. Booth DA, Toates RM. No paradox in the control of energy intake. Nature 1978;275(5678):345.
7. Ebbeling CB, Feldman HA, Klein GL, Wong JMW, Bielak L, Steltz SK, Luoto PK, Wolfe RR, Wong WW, Ludwig DS. Effects of a low carbohydrate diet on energy expenditure during weight loss maintenance: randomized trial. BMJ 2018;363:k4583.
8. Ludwig DS, Lakin PR, Wong WW, Ebbeling CB. Scientific discourse in the era of open science: a response to Hall et al. regarding the carbohydrate-insulin model. Int J Obes 2019;43(12):2355–60.
9. Ludwig DS, Greco KF, Ma C, Ebbeling CB. Testing the carbohydrate-insulin model of obesity in a 5-month feeding study: the perils of post-hoc participant exclusions. Eur J Clin Nutr 2020;74(7):1109–12.
10. Ludwig DS, Dickinson SL, Henschel B, Ebbeling CB, Allison DB. Do lower-carbohydrate diets increase total energy expenditure? An updated and reanalyzed meta-analysis of 29 controlled-feeding studies. J Nutr 2021;151(3):482–90.

doi: https://doi.org/10.1093/ajcn/nqab385.

ASN guidelines on *P* values

Dear Editor:

The recently proposed guidelines by the American Society for Nutrition (ASN) journals on *P* values (1) correctly point out some of the problems with *P* values, but directly contradict the explicit and well-considered recommendation to abandon statistical significance testing by the American Statistical Association (2) and many other learned critiques by statisticians as reviewed by Hurlbert et al. (3) and Greenland et al. (4). It is not helpful for researchers to be confronted with conflicting recommendations. The ASN vision reflects a dominant but erroneous way of thinking and ignores the wealth of modern ideas on scientific inference. To justify significance testing, ASN guidelines (1) state that medical and nutritional research often requires making a binary choice (e.g., to declare a treatment effective or not, to recommend 1 set of nutritional recommendations or another, to further investigate or move on to another question). Even so, we would hope that a binary decision is not made only based on the presence or absence of an effect, but also and primarily on the magnitude of the effect. Dichotomizing *P* values implies that biology is discontinuous, which is seldom the case. It is disheartening to see that, after decades of progress in thinking about these issues, this misleading and simplified approach is being promoted by ASN.

The ASN guidelines (1) point out some well-known problems that have collectively been referred to as the 4 horsemen of the reproducibility apocalypse (5): publication bias, insufficient sample size, *P*-hacking, and HARKing (hypothesizing after results are known). They fail to address what is arguably the most important issue—namely, that most researchers interpret *P* values using flawed inferential reasoning (i.e., it is not the use, but the misuse of *P* values that is the main problem). They fail to consider a *P* value as a *conditional* probability (i.e., the probability of findings in the sample at least as extreme as observed, given that, in truth, there is no association in the sampled population). They are also looking "in

the wrong direction," from hypothesized effect to data instead of the other way around.

To illustrate this point, consider a similar problem in clinical practice: a physician reports to a patient that the result of her diagnostic test was positive. When the patient asks whether the test result could be wrong, she would be poorly served with an answer that the test is highly specific (i.e., given the absence of disease, it is very unlikely that the test result is positive). The patient's primary interest concerns the question: Given that my test result is positive, what is the probability of truly having the disease? This positive predictive value depends not only on specificity but also on sensitivity and the a priori probability of disease. A clinician weighs the evidence of a test result in view of her combined understanding of the biology of the disease, patient characteristics, and the pre-test probability of disease. That is why the interpretation is not left to the laboratory technician. Similarly, statistically significant results should be interpreted taking the prior expectation and plausibility of the null hypothesis into account. By intuition, people usually get this right. For example, consider a trial report with a statistically significant benefit obtained with a homeopathic, super-diluted remedy. People who do not believe in homeopathy are unlikely to believe the test results. Statistical testing is like interpreting a diagnostic test result by looking only at its specificity—that is, under the null hypothesis of no disease. Interpretation of statistical tests should also take into account the plausibility or likelihood of the alternative hypothesis, which depends on external or subjective knowledge. That is also why interpretation of study results should not be left to a simplistic statistical rule.

Statistical adjustment for multiple comparisons, as recommended by default in the ASN guidelines (1), results in an increased probability of false-negative results. It also undermines the interpretation of related endpoints (6). It is equivalent to a physician finding an abnormally low hemoglobin concentration in a patient but no longer judging it worthy of treatment because she also found iron deficiency. In their Figure 2, Sorkin et al. (1) show that the probability of at least 1 false-positive result occurring increases with the number of tests performed. This is true when test results are independent. Because, in practice, outcomes are typically related, the default should be to not adjust, and if adjustment is nonetheless done, it should be justified. Many other commentaries support this view, again summarized by Hurlbert et al. (3).

To assist in the interpretation of significance, the ASN guidelines (1) recommend that *P* values should be reported with a statement of the sample size, an estimate of the treatment effect, and its variability. This is 1 option, but it is very cumbersome and we do not believe that adding more statistical information would help the general reader in interpretating (non)significance. Why not demand instead that effects are reported with CIs? Contrary to what is stated in the AJCN guidelines (1), however, CIs do not give a range in which the true value of a parameter θ is expected to lie. It is not Bayesian; a 95% CI does not mean that the probability that the true value of the parameter is in the interval is 95%. Instead, as conceived by Neyman (7), a 95% CI encompasses a range of hypothesized effect sizes that have a *P* value exceeding 0.05—that is, hypothesized effect sizes within this range would be compatible with the sample estimate x_0 if the *P* value would be set at 0.05. In mathematical notation: $Pr(\theta|x_0) \neq Pr(x_0|\theta)$. Some additional pitfalls in the interpretation of CIs are outlined by Greenland et al. (4).

In conclusion, we agree that *P* values should not be banned. But, they should generally not be dichotomized, they should never be reported as (non)significant, and they should not be used unless there are good reasons for doing so. Even better is to separate results into a point estimate and its corresponding 95% CI. Because all information about statistical precision is contained in CIs, it is not necessary to additionally report *P* values.

The authors' responsibilities were as follows—HV: wrote the manuscript; EF, PvV, and AMP: edited the manuscript; and all authors: read and approved the final manuscript. The authors report no conflicts of interest.

Hans Verhoef
Edith Feskens
Pieter van 't Veer
Andrew M Prentice

From Wageningen University, Division of Human Nutrition and Health, Wageningen, The Netherlands (EF, PvV; HV, e-mail: hans.verhoef@wur.nl); and MRC Unit, The Gambia at London School of Hygiene and Tropical Medicine, Fajara, Banjul, The Gambia (AMP).

References

1. Sorkin JD, Manary M, Smeets PAM, MacFarlane AJ, Astrup A, Prigeon RL, Hogans BB, Odle J, Davis TA, Tucker KL, et al. A guide for authors and readers of the American Society for Nutrition journals on the proper use of *P* values and strategies that promote transparency and improve research reproducibility. Am J Clin Nutr 2021;114:1–6.
2. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p < 0.05." Am Stat 2019;73:1–19.
3. Hurlbert SH, Levine RA, Utts J. Coup de grâce for a tough old bull: "statistically significant" expires. Am Stat 2019;73(Suppl 1):352–7.
4. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31(4):337–50.
5. Bishop D. Rein in the four horsemen of irreproducibility. Nature 2019;568(7753):435.
6. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. Lancet North Am Ed 2005;365(9470):1591–5.
7. Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. Phil Trans Royal Soc of London Ser A 1937;236:333–80.

doi: https://doi.org/10.1093/ajcn/nqab370.

Reply to Verhoef et al.

Dear Editor:

We thank our colleagues, Verhoef et al., for their thoughtful reply (1) to our article, "A guide for authors and readers of the American Society for Nutrition Journals on the proper use of *P* values, and transparency, to improve research reproducibility" (2).

Our colleagues state that we "directly contradict the explicit and well-considered recommendation to abandon statistical significance testing by the American Statistical Association." We do not. In the American Statistical Association (ASA) Statement on Statistical Significance and *P*-Values (3), the ASA does not state that *P* values should be banned but rather that they should be used in proper context.

JDS was supported by the Baltimore VA Medical Center GRE CC, NIA AG028747, and NIDDK P30 DK072488; CPD was supported by K24 DK104676 and P30 DK04056, BBH was supported by VA RR&D grant 5I21RX003169-02; TAD was supported by HD-085573 and HD-099080 and USDA CRIS 3092-51000-060.