

Erratum to: Letter to the Editor: The ongoing tyranny of statistical significance testing in biomedical research

Tyler J. VanderWeele

Published online: 29 December 2010
© Springer Science+Business Media B.V. 2010

Erratum to: Eur J Epidemiol (2010) 25:843–845
DOI 10.1007/s10654-010-9507-8

We apologize for a number of misspellings of Dr. VanderWeele's surname and inaccuracy in the 9th paragraph of The Author's Reply; 'least inclined' should have read 'most inclined'.

The Publisher apologizes for these errors and also for inconvenience and/or embarrassment caused to the authors. The correct full text read as follows:

The Authors Reply

On a use of the null P -value

We thank Tyler VanderWeele [1] for his kind words about our recent paper on statistical significance [2]. We will thank him even more in the future if he stops suggesting that we are intent on "dismissing the P -value entirely." We oppose some uses of P -values, not P -values themselves. If any of us does issue a call for P -values to be banished, it will require no reading between the lines to discern.

Long ago, one of us [3] suggested that interpretation might be enhanced by inspecting the graph of all P -values for all hypothetical values of the measure one is estimating. Our support for the continued reporting of point estimates and confidence intervals [2] was tantamount to encouraging

the reporting of three P -values: $P = 1$ for the point estimate and $P = 0.05$ for each of the interval's limits (if the confidence level is 95%). How these recommendations can be viewed as aligning us with those who would instruct authors "to refrain from giving P -values entirely" is a mystery.

To be fair, VanderWeele does not defend the reporting of all P -values. He has a specific P -value in mind, one so special it has become known as "the" P -value. It is the P -value for the null hypothesis. We have no objection to the reporting of this or any other P -value. What matters to us at present is what VanderWeele would have us do with it.

As we have tried to stress, we believe it is wise for epidemiologic researchers to focus on estimates: point estimates, interval estimates or entire P -value or likelihood functions. Estimates are what systematic reviewers seek when they review a literature. When they do not find them and find null P -values instead, they are understandably disappointed, for they know their review will not reach as deep an understanding of the state of epidemiologic research on the topic as it might have.

When we look at the forest plot in the Figure [4], we see 13 estimates. Apparently, VanderWeele sees something else: up to 13 "findings." We are not sure how he defines this key term in his framework. It would appear that a finding cannot be a null P -value, as he urges us to use null P -values in coming to judgments about findings. It would also seem that findings, as the objects of those judgments, cannot be the judgments themselves. Our best guess is that VanderWeele's "findings" are estimates, though we suspect that not every estimate qualifies.

VanderWeele seems to view chance and bias as competing, alternative explanations for findings. Chance is to be considered first, for reasons that are obscure to us, with the aim of "dismissing" it, for reasons that seem just as obscure. Given a very low null P -value, he would have us

The online version of the original article can be found under doi:10.1007/s10654-010-9507-8.

T. J. VanderWeele (✉)
Harvard School of Public Health, Boston, MA, USA
e-mail: tvanderw@hsph.harvard.edu

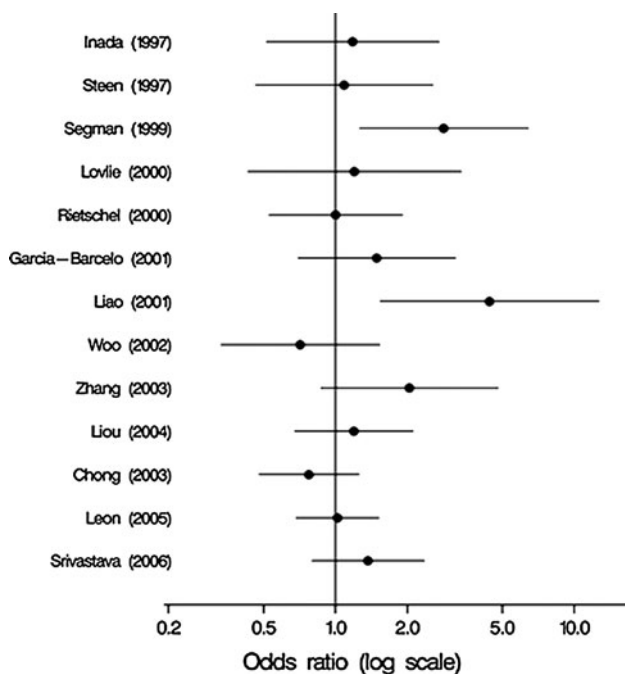


Fig. 1 Odds ratio estimates and 95% confidence intervals from case-control studies of tardive dyskinesia prevalence contrasting the Serine/Glycine and Serine/Serine rs6280 polymorphisms in the dopamine receptor 3 gene [4]

“rather focus” on bias than on chance, as though each finding must have one or the other (or, presumably, causality) as its explanation.

Our working model is that chance and bias are always part of the explanation for every estimate and, of course, that causality often plays a role as well. In our view, the ever-present influences of chance and bias on estimates should always be considered, in no fixed order of priority. Our aim is not to “dismiss” any of these influences, but to gauge them as to their magnitude and, in the case of biases, their direction.

Of the 13 estimates in the Figure [4], we would say that the estimate of Leon (2005) is the least influenced by chance. It has the narrowest confidence interval. We are not even sure that VanderWeele would consider this estimate a “finding.” We suspect not.

We would say that chance played the greatest role in helping produce the estimate of Liao (2001), which has the widest confidence interval. VanderWeele, in contrast, would be most inclined to “dismiss” chance as “a potential explanation for th[is] finding,” as it has the lowest of all 13 null P -values ($P_{\text{null}} = 0.006$).

The difference in outlooks could not be sharper. We would encourage epidemiologists to focus on estimates, to assess the precision and validity of every estimate, and to use the width of a confidence interval to gauge an estimate’s precision. We would discourage epidemiologists from focusing on findings, from considering chance and bias as competing explanations for findings, and from using null P -values to dismiss chance before considering bias (Fig. 1).

We have expressed some critical thoughts on VanderWeele’s considered views of these matters. We hope no one, most of all he, will impute to us a suggestion that his views should be banned or “dismissed entirely.”

References

1. VanderWeele TJ. Re: “The ongoing tyranny of statistical significance testing in biomedical research.” *Eur J Epidemiol.* 2010; 25:843–5.
2. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. *Eur J Epidemiol.* 2010;25:225–30.
3. Poole C. Beyond the confidence interval. *Am J Public Health.* 1987;77:195–9.
4. Tsai HT, North KE, West SL, Poole C. The DRD3 rs6280 polymorphism and prevalence of tardive dyskinesia: a meta-analysis. *Am J Med Genet Part B.* 2010;153B:57–66.

C. Poole

Department of Epidemiology,
UNC Gillings School of Global Public Health,
University of North Carolina,
Chapel Hill, NC, USA
e-mail: cpoole@unc.edu

O. Kuss

Institut für Medizinische Epidemiologie,
Biometrie und Informatik,
Univeritätsklinikum und Medizinische Fakultät,
Martin-Luther-Universität Halle-Wittenberg,
Halle, Germany
e-mail: oliver.kuss@medizin.uni-halle.de

A. Stang

Institut für Klinische Epidemiologie,
Medizinische Fakultät,
Martin-Luther-Universität Halle-Wittenberg,
Halle, Germany
e-mail: andreas.stang@medizin.uni-halle.de