International Epidemiological Association

International Journal of Epidemiology, 2020, 1–2
doi: 10.1093/ije/dyaa096
Com Inv Only

Com Inv Only

Commentary: Cynical epidemiology

Jay S. Kaufman

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Ave West, Montreal, Quebec H3A1A2, Canada. E-mail: jay.kaufman@mcgill.ca

Editorial decision 1 May 2020; Accepted 4 May 2020

We know very well by now that the institutions managing the epidemiological research environment—agencies, schools and journals—do not sufficiently incentivize getting the best answers to causal questions.[1,2] Sadly, they do not even incentivize asking well-formulated causal questions in the first place.[3] The dominant metrics of success are publication counts, impact factors and external funds received, none of which necessarily reflect the priority of sober and honest accounting of study limitations and biases.[4] If you are a NASA engineer and a spaceship crashes because you messed up, heads are going to roll. But if you are an epidemiologist who told people to eat margarine when they were better off eating butter, you never have to give any money back to the funders. For outright fraud, maybe. Yet for simple incompetence, or even stubborn and wilful incompetence, there are generally no retractions, no penalties, no demotions, no apologies. Thus we perversely incentivize 'spin', obscuring true weaknesses and limitations, selecting the most 'exciting' results to highlight (often gauging excitement by the smallness of the *P*-value), and anything else short of outright fabrication that gets the submission past the reviewers and into a 'top' journal.[5] The review process itself is played like a football game, dodging and weaving through the opposing team's defences to score a goal. This does not look like a scientific community dedicated to deducing the best answers to the best questions. Rather, this is a sadly cynical portrayal of our field, and one consistent with the paper published in this issue by Blum *et al*.[6]

The authors describe a systematic literature search for all published papers reporting E-values through to the end of 2018. As they explain, the E-value is a sensitivity analysis for uncontrolled confounding that is dumbed down to a single number.[7] Under some rather bizarre conditions, it represents the minimum strength of an unmeasured confounder that could nullify the reported finding. The inventors of the E-value have been forthright about the rather unrealistic set-up required to boil three parameters of an unmeasured confounder (association with exposure, association with outcome and target-population-specific prevalence) into a single number.[8,9] They argue, however, that sensitivity analyses for residual confounding are not widely reported in the biomedical literature, and reducing this problem to the simplicity of a single number could make this consideration more widely accessible. The scalar E-value is obviously an imitation of the wildly popular *P*-value, another single number that is meant to index some aspect of validity. The twist that made the *P*-value especially malign, however, was its arbitrary categorization, for example at 0.05 or 0.005.[10,11] Thus, it is ironic that Blum *et al*. criticize the E-value for not making that same egregious mistake: '[T]here is no clear demarcation of what magnitude of E-values is large enough to herald protection from confounding', they complain. Of course there is not, nor should there be.[12]

The E-value has not been around for very long, and so Blum *et al*. found only 87 publications, and matched these 1:1 by journal and issue to have a comparison set of papers that did not employ E-values. They achieved 67 matched pairs, which is not much data to support secure inferences. The authors report various analyses from this small dataset, with a focus on how authors use E-values as a way of deflecting concerns about residual confounding. The 'standard of care' is for epidemiologists to treat their observational studies as conditionally randomized, with no attention to residual confounding whatsoever. Sometimes

in the discussion section they will admit to lacking some other key covariate, but this is generally dismissed qualitatively. Indeed, in the comparison series, only three of the matched papers averred that residual confounding was unlikely to challenge the validity of a reported finding, whereas 52 made no comment of any kind about unmeasured confounding. This contrasts with the E-value papers, for which 44 stated that the validity of the reported effect was not threatened by residual confounding. Even with small numbers, this is a big discrepancy, and it rings true. Authors want to publish their papers, and the usual way, if the main claim is for a causal effect of the exposure, is to ignore residual confounding altogether. For authors claiming effects who report E-values, do we really expect them to propose that their reported effect is entirely spurious? That is not how one dodges and weaves through the defence to score a goal. Unless of course the main claim of the paper is exoneration of an exposure, in which case weaponizing the E-value is exactly what we can expect.

The numbers reported by Blum *et al*. make E-values sound deleterious, but it is not so obvious. E-Value advocates argue that just getting authors to discuss unmeasured confounders is a step in the right direction.[8] Indeed, uncontrolled potential confounders that potentially threatened the main conclusions were listed in 26 of the matched E-value papers, but only in 16 of the non-E-value papers. Among all 87 E-value papers, 19 related E-value magnitudes to expected strengths of specific confounders. Of course, it would be great if all E-value papers did this, but almost a quarter of papers is better than almost none, which is what happened in the control series. And more than half of the E-value papers presented the results of at least one other sensitivity analysis.

The authors clearly do not like E-values and 'spin' the discussion to conclude that they do more harm than good, proposing that they act as an 'alibi' for unmeasured confounders. They agree that current practice is bad, but argue that 'facile automation' makes things worse. They may be right, but it is too soon to tell. These are early adopters, and practice will improve over time as the critiques become required reading in epidemiology training and reviewers become better informed. Maybe E-values will serve as a stepping stone for some researchers who will then move on to more realistic and informative sensitivity analyses. Just getting consideration of residual confounding into more papers may be a benefit, even if ultimately dismissed. We need more data, but one thing is already certain: new tools get used in existing contexts. Say you have a system that incentivizes carpenters to build houses that look nice to the casual observer but have weak foundations and soon collapse. You cannot hand the carpenters a new kind of hammer and then complain that the weak foundation is the fault of this new hammer. It's the research context, the human system, not the technical method that is at fault. You have to somehow change this context so that carpenters will want to bother making strong foundations. Then, after you fix the flawed human institutions that warped carpenters' priorities, you can finally start to judge whether the new hammer is better than the old one.[13]

Acknowledgement

Sander Greenland generously read a draft version and offered helpful suggestions.

Conflict of Interest

None declared.

References

1. Cole P, Delzell E, Rodu B. Moneychangers in the Temple. *Epidemiology* 2000;**11**:84–90.
2. Lindner MD, Torralba KD, Khan NA. Scientific productivity: an exploratory study of metrics and incentives. *PLoS One* 2018;**13**:e0195321.
3. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol* 2016;**26**:674–80.
4. Galea S. A typology of nonfinancial conflict in population health research. *Am J Public Health* 2018;**108**:631–32.
5. Chiu K, Grundy Q, Bero L. (Spin' in published biomedical literature: a methodological systematic review. *PLoS Biol* 2017;**15**:e2002173.
6. Blum MR, Tan YJ, Ioannidis J. Use of E-values for addressing confounding in observational studies—an empirical assessment of the literature. *Int J Epid* 2020; https://doi.org/10.1093/ije/dyz261.
7. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med* 2017;**167**:268.
8. VanderWeele TJ, Mathur MB, Ding P. Correcting misinterpretations of the E-value. *Ann Intern Med* 2019;**170**:131–2.
9. Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990;**1**:239–46.
10. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019;**567**:305–7.
11. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. *Nat Hum Behav* 2018;**2**:4.
12. Greenland S. Invited Commentary: The need for cognitive science in methodology. *Am J Epidemiol* 2017;**186**:639–45.
13. Greenland S. For and against methodologies: some perspectives on recent causal and statistical inference debates. *Eur J Epidemiol* 2017;**32**:3–20.