



Published in final edited form as:

Environ Int. 2022 February ; 160: 107032. doi:10.1016/j.envint.2021.107032.

How to report E-values for meta-analyses: Recommended improvements and additions to the new GRADE approach

Maya B. Mathur^{*,1}, Tyler J. VanderWeele²

¹Quantitative Sciences Unit and Department of Pediatrics, Stanford University

²Department of Epidemiology, Harvard University

Abstract

In a recent concept paper (Verbeek et al., 2021), the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group provides a preliminary proposal to improve its existing guidelines for assessing sensitivity to uncontrolled confounding in meta-analyses of nonrandomized studies. The new proposal centers on reporting the E-value for the meta-analytic mean and on comparing this E-value to a measured “reference confounder” to determine whether residual uncontrolled confounding in the meta-analyzed studies could or could not plausibly explain away the meta-analytic mean. Although we agree that E-value analogs for meta-analyses could be an informative addition to future GRADE guidelines, we suggest improvements to the Verbeek et al. (2021)’s specific proposal regarding: (1) their interpretation of comparisons between the E-value and the strengths of associations of a reference confounder; (2) their characterization of evidence strength in meta-analyses in terms of only the meta-analytic mean; and (3) the possibility of confounding bias that is heterogeneous across studies.

Keywords

meta-analysis; bias; confounding; observational studies; sensitivity analysis

In a recent concept paper (Verbeek et al., 2021), the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group provides a preliminary proposal to improve its existing guidelines (Schünemann et al., 2019) for assessing sensitivity to uncontrolled confounding in meta-analyses of nonrandomized studies. Verbeek et al. (2021) propose reporting the E-value (VanderWeele and Ding, 2017) for the meta-analytic mean; as we have shown, this E-value represents the average strengths of association across studies, on the risk ratio (*RR*) scale, that uncontrolled confounder(s) would need to have with studies’ exposures and outcomes in order to shift the meta-analytic mean to the null (Mathur and VanderWeele, 2020a; VanderWeele and Ding, 2017). Verbeek et al. (2021) then propose comparing this E-value to the strengths of association that a measured “reference confounder” has with the exposure and with the outcome to determine

*Correspondence to: Maya B. Mathur (mmathur@stanford.edu), Quantitative Sciences Unit, 1701 Page Mill Road, Palo Alto, CA, 94304.

Conflicts of interest

The authors have no conflicts to disclose.

whether residual uncontrolled confounding in the meta-analyzed studies could or could not plausibly explain away the meta-analytic mean.

We believe that Verbeek et al. (2021)'s proposal does improve upon the current GRADE guidelines, in which the meta-analyst can choose to upgrade the evidence certainty rating if the meta-analytic mean is greater than $RR = 2$, on the assumption that meta-analyses with larger mean estimates are categorically more robust to uncontrolled confounding than those with smaller estimates (Schünemann et al., 2019). We agree with Verbeek et al. (2021) that this existing approach is limited and that using quantitative sensitivity analyses, such as E-value analogs for meta-analysis, can better inform judgments of robustness to uncontrolled confounding.

However, we do have three central suggestions to improve upon limitations of Verbeek et al. (2021)'s proposal, namely regarding their interpretation of comparisons between the E-value and the strengths of associations of a reference confounder; their characterization of evidence strength in meta-analyses in terms of only the meta-analytic mean; and the possibility of confounding bias that is heterogeneous across studies.

Point #1: Verbeek et al. (2021) provide incorrect interpretations of comparisons between the E-value and the strengths of association of a reference confounder.

Verbeek et al. (2021) suggest identifying a “reference confounder”, which is the single known confounder that has the strongest known strengths of association on the risk ratio (RR) scale with the exposure and with the outcome. They recommend using these known strengths of confounding associations, which they term RR_{CE} and RR_{CO} , to calculate the bias factor (VanderWeele and Ding, 2017; Ding and VanderWeele, 2016), which they incorrectly define as “the amount by which a confounder would reduce the observed effect size given known relations between confounder and outcome and exposure” (Verbeek et al., 2021). They recommend dividing the observed meta-analysis mean on the risk ratio scale by this bias factor to “see if the reference confounder reduces the observed RR to the null value or below a relevant threshold” and interpret the results as follows: “...if the RR [for the confounding associations RR_{CE} and RR_{CO}] is larger than [the E-value] it is likely that a residual confounder could reduce the observed RR to null as well”.

This interpretation is incorrect. As we have emphasized throughout our work on the E-value, the bias factor represents the *maximum possible* bias that *could* be generated by uncontrolled confounder(s) with specified joint *residual* confounding associations (e.g., RR_{CE} and RR_{CO}) *above and beyond controlled confounders* (VanderWeele and Ding, 2017; Ding and VanderWeele, 2016; VanderWeele et al., 2019a). In other words, the bias factor and resulting E-value conservatively consider hypothetical worst-case uncontrolled confounder(s); however, any given uncontrolled confounder(s) may or may not produce that much bias, depending on (for example) the confounders' prevalences. If, as Verbeek et al. (2021) stipulate, the meta-analyzed studies “should have adjusted for the effects of all known critical confounders”, the strongest known, controlled confounder among this rich

set could very well have strong associations with the exposure and outcome. However, because of the studies' extensive control of known confounders, it may nevertheless be the case that any remaining uncontrolled confounder(s) have considerably weaker strengths of association *above and beyond* those produced by the reference confounder and other controlled confounders. Furthermore, even if uncontrolled confounder(s) do have confounding strengths of associations as large as RR_{CE} and RR_{CO} , this does not mean that such uncontrolled confounders necessarily "would reduce" the meta-analytic mean to the null (Verbeek et al., 2021), only that they could potentially do so.

On the other hand, Verbeek et al. (2021) state that "if the reference confounder will not be able to reduce the observed effect size to null, then it is unlikely that an unknown residual confounder will be able to do this". Statements of this form are on stronger ground than their logical inverses, discussed above, but still require careful interpretation. As Verbeek et al. (2021) briefly allude later, such judgments of robustness to uncontrolled confounding based on reference confounders may be reasonable if the meta-analyzed studies have good existing control of confounding. However, if the meta-analyzed studies have relatively limited control of confounding, it is quite possible that the strongest known confounder from among this limited set has weaker confounding associations than the potentially numerous uncontrolled confounders, especially when the uncontrolled confounders' effects are considered jointly (VanderWeele, under review; VanderWeele et al., 2019b).

As an additional caveat, estimates of reference confounder associations depend on which other confounders were included in analysis (VanderWeele and Mathur, 2020). Particularly if the meta-analyzed studies control for different sets of confounders, caution would be warranted when using reference confounder associations as a benchmark for the plausible strengths of associations for uncontrolled confounders across studies (VanderWeele and Mathur, 2020).

Our proposed modification regarding Point #1.

We previously gave suggestions for how one might report on known confounders and their strengths of associations with the exposure and outcome (VanderWeele and Mathur, 2020). Considering the E-value with reference to these known confounding associations is perhaps most informative if the meta-analyzed studies have good existing confounding control and when the associations of the known confounder(s) with the exposure and outcome are both less than E-value. One could then correctly conclude that if any uncontrolled confounder(s) had joint confounding associations, *above and beyond* any controlled confounders, that were weaker those of the known confounder(s), and hence the E-value, then such uncontrolled confounder(s) could not explain away the meta-analytic mean. In the context of studies with good existing confounding control, such a finding might bolster one's confidence that any uncontrolled confounder(s) would be unlikely to have such strong confounding associations.

Point #2: Verbeek et al. (2021) characterize evidence strength in a meta-analysis in terms of only the meta-analytic mean.

When we introduced E-value analogs for meta-analysis, we emphasized that evidence strength in meta-analyses of heterogeneous effects should not be characterized only in terms of the meta-analytic mean (Mathur and VanderWeele, 2020a,b). We discussed reporting the E-value for the meta-analytic mean, but our central focus was on metrics of evidence strength that better characterize the distribution of potentially heterogeneous effects across studies. We had suggested estimating the proportion of causal population effects that are meaningfully strong, defined as effects above a threshold that the meta-analyst has chosen to represent a meaningfully strong causal effect in the scientific context (e.g., $RR = 1.1$ or some other threshold) (Mathur and VanderWeele, 2020a,b). Additionally, we suggested estimating the proportion of effects below a second, possibly symmetric, threshold in the opposite direction from the meta-analytic mean. Unlike the meta-analytic mean alone, these proportion metrics can help identify if: (1) few meaningfully strong effects exist despite a “statistically significant” meta-analytic mean; (2) some large effects also exist despite an apparently null point estimate; or (3) strong effects in the direction opposite of the meta-analytic mean also regularly occur (Mathur and VanderWeele, 2019b). These metrics can also sometimes help adjudicate apparent conflicts between multiple meta-analyses (Mathur and VanderWeele, 2019a; Lewis et al., 2020).

Our proposed modification regarding Point #2.

We would recommend reporting not only the E-value for meta-analytic mean and its confidence interval, but also metrics that represent the severity of uncontrolled confounding that would be required to reduce to less than some value r (e.g., 0.15) the proportion of causal population effects stronger than a threshold (q) chosen to represent a meaningfully strong effect size (Mathur and VanderWeele, 2020a, 2022). These metrics are straightforward to calculate using the R package EValue or the website www.evalue-calculator.com/meta (Mathur et al., 2018). We recently provided a detailed tutorial for using these tools and interpreting their results (Mathur and VanderWeele, 2022).

Point #3: Verbeek et al. (2021)’s proposal does not characterize confounding bias that may be heterogeneous across studies.

In many meta-analyses, studies differ in their susceptibility to uncontrolled confounding due, for example, to differences in design and in the set of confounders that were measured and controlled (Mathur and VanderWeele, 2022). When confounding bias is heterogeneous across studies, the E-value can indeed be applied as Verbeek et al. (2021) suggest, but it is important to note that this E-value represents the *average* strengths of association *across studies* that uncontrolled confounder(s) would need to have with studies’ exposures and outcomes in order to shift the meta-analytic mean to the null (Mathur and VanderWeele, 2020a; VanderWeele and Ding, 2017). When confounding bias is thought to differ substantially across studies, it may be difficult to interpret sensitivity analyses that

characterize the severity of uncontrolled confounding in terms of only the average strengths of confounding associations across studies.

Our proposed modification regarding Point #3.

We would again recommend reporting not only the E-value for the meta-analytic mean and its confidence interval, but also the aforementioned metrics that characterize heterogeneity across studies in both causal effects and in confounding bias (Mathur and VanderWeele, 2020a,b, 2022). Specifically, those metrics allow consideration of bias factors that are log-normal across studies; in some cases, the metrics can also provide an estimated lower bound on the proportion of meaningfully strong effects if the bias might be heterogeneous, and one does not want to make assumptions about its distribution (Mathur and VanderWeele, 2020a,b, 2022).

1. Conclusion

We agree with Verbeek et al. (2021) that E-value analogs for meta-analyses could be an informative addition to future GRADE guidelines. However, if Verbeek et al. (2021)'s proposal is to be included in future GRADE guidelines, we would also propose supplementing their proposal with the modifications we suggest above. We recently gave our own recommendations and tutorial for reporting E-value analogs in meta-analyses (Mathur and VanderWeele, 2022), along with recommendations for defining study eligibility criteria that help limit confounding bias at the outset. We hope that our suggestions regarding Verbeek et al. (2021)'s approach, as well as our own recommendations, will be taken into consideration in potential updates to the GRADE guidelines.

Funding

This research was supported by (1) National Institutes of Health (NIH) grants R01 LM013866R01 and R01 CA222147; (2) the NIH-funded Biostatistics, Epidemiology and Research Design (BERD) Shared Resource of Stanford University's Clinical and Translational Education and Research (UL1TR003142); (3) the Biostatistics Shared Resource (BSR) of the NIH-funded Stanford Cancer Institute (P30CA124435); and (4) the Quantitative Sciences Unit through the Stanford Diabetes Research Center (P30DK116074). The funders had no role in the design, conduct, or reporting.

References

- Ding P and VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology*, 27 (3):368, 2016. [PubMed: 26841057]
- Lewis M, Mathur MB, VanderWeele TJ, and Frank MC. The puzzling relationship between multi-lab replications and meta-analyses of the published literature. 2020. Preprint: <https://psyarxiv.com/pbrdk/>.
- Mathur MB and VanderWeele TJ. Finding common ground in meta-analysis “wars” on violent video games. *Perspectives on Psychological Science*, 14(4):705–708, 2019a. [PubMed: 31188714]
- Mathur MB and VanderWeele TJ. New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, 38(8):1336–1342, 2019b. [PubMed: 30513552]
- Mathur MB and VanderWeele TJ. Sensitivity analysis for unmeasured confounding in meta-analyses. *Journal of the American Statistical Association*, 115(529):163–172, 2020a. [PubMed: 32981992]
- Mathur MB and VanderWeele TJ. Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. *Epidemiology*, 31(3):356–358, 2020b. [PubMed: 32141922]

- Mathur MB and VanderWeele TJ. Methods to address confounding and other biases in meta-analyses: Review and recommendations. Annual Review of Public Health, 2022. Preprint retrieved from <https://osf.io/v7dtq/>.
- Mathur MB, Ding P, Riddell CA, and VanderWeele TJ. Website and R package for computing E-values. Epidemiology, 29(5):e45, 2018. [PubMed: 29912013]
- Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, Morgan RL, Gartlehner G, Kunz R, Katikireddi SV, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. Journal of Clinical Epidemiology, 111:105–114, 2019. [PubMed: 29432858]
- VanderWeele T and Ding P. Sensitivity analysis in observational research: Introducing the E-value. Annals of Internal Medicine, pages doi: 10.7326/M16-2607, 2017.
- VanderWeele T, Mathur M, and Ding P. Correcting misinterpretations of the E-value. Annals of Internal Medicine, 2019a.
- VanderWeele TJ. Are Greenland, Ioannidis, and Poole opposed to the Cornfield conditions? A defense of the E-value. under review.
- VanderWeele TJ and Mathur MB. Commentary: developing best-practice guidelines for the reporting of e-values. International Journal of Epidemiology, 49(5):1495–1497, 2020. [PubMed: 32743656]
- VanderWeele TJ, Ding P, and Mathur M. Technical considerations in the use of the E-value. Journal of Causal Inference, 7(2), 2019b.
- Verbeek JH, Whaley P, Morgan RL, Taylor KW, Rooney AA, Schwingshackl L, Hoving JL, Katikireddi SV, Shea B, Mustafa RA, et al. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: A GRADE concept paper. Environment International, 157:106868, 2021. [PubMed: 34530289]