

CONTROVERSY & DEBATE SERIES

Controversy and debate on credibility ceilings. Paper 1: Fundamental problems with the “credibility ceiling” method for meta-analyses

Maya B. Mathur^{a,*}, Tyler J. VanderWeele^b

^aQuantitative Sciences Unit, Stanford University, Stanford, CA, USA

^bDepartment of Epidemiology, Harvard University, Boston, MA, USA

Accepted 4 May 2020; Published online 23 May 2020

Abstract

Background and Objectives: The “credibility ceiling” method was proposed to conduct sensitivity analysis for unmeasured confounding and other forms of bias in meta-analyses and has been used in umbrella reviews to grade evidence strength. However, we explain that the method has fundamental statistical flaws.

Methods: We use statistical reasoning to assess the method’s validity, providing intuition for our findings by presenting simple applied examples in which the method yields clearly incorrect conclusions.

Results: The credibility ceiling is not a valid bias correction, as we show mathematically and illustrate using examples in which, for example, the method incorrectly “adjusts” the meta-analytic point estimate in the wrong direction. Although the originators describe the method as limiting the credibility of any given observational study to a fixed ceiling, we show why this interpretation in fact bears little relation to what the method actually does.

Conclusion: Given the fundamental problems with the credibility ceiling method and its demonstrated potential for misleading conclusions, we recommend against its use. © 2020 Elsevier Inc. All rights reserved.

Keywords: Meta-analysis; Sensitivity analysis; Bias; Observational studies; Confounding

The “credibility ceiling” method was proposed to conduct sensitivity analysis for unmeasured confounding and other forms of bias in meta-analyses [1–3] and has been used in umbrella reviews to grade evidence strength across different meta-analyses (e.g., [4,5]). We argue that this method has substantial statistical shortcomings, illustrate the misleading consequences that can arise, and consequently recommend against its use.

The credibility ceiling method proceeds as follows. First, one assumes that, for any given observational study, the probability that its point estimate is in the wrong direction (i.e., that it disagrees in sign with the underlying effect) must be at least as large as some prespecified value, c , because one can almost never be sure of conclusions from observational studies, which are subject to potentially numerous biases. The method proposes to “correct” the meta-analytic point estimate as follows. For each meta-analyzed study, one calculates the probability that a normal variable, u_i , whose mean and variance are equal to the study’s point estimate and

variance estimate, respectively, would disagree in sign with the study’s point estimate itself. If this probability is less than c , the study’s variance is inflated until the probability such a normal random variable would disagree in sign is exactly c . Finally, as a sensitivity analysis, all studies are meta-analyzed using these potentially inflated variances, rather than their actual sample variances.

The method’s originators provided no mathematical derivation nor statistical justification when proposing the credibility ceiling [1], and unfortunately the method is subject to several substantial statistical shortcomings. Fundamentally, the method is not a valid bias correction. Unmeasured confounding causes bias in studies’ point estimates in the sense that the point estimates systematically differ from the underlying causal effects they are intended to estimate. This bias within the individual studies propagates to the meta-analytic estimate. When conducting bias corrections or sensitivity analyses, we are often concerned with confounding that has inflated the point estimates away from the null, and that is the setting we will discuss here. A statistically valid bias correction or sensitivity analysis method would then need to counteract this bias by shifting either individual studies’ estimates or the meta-analytic

* Corresponding author. Quantitative Sciences Unit, Stanford University, CA 94304, USA. Tel.: 650-863-2384.

E-mail address: mmathur@stanford.edu (M.B. Mathur).

Key findings

- The credibility ceiling method for meta-analyses is fundamentally flawed and can lead to highly misleading conclusions. Therefore, it should not be used.

What this adds to what was known?

- No statistical justification of the credibility ceiling method was provided by its originators.
- We provide our own mathematical derivation of the method, which reveals its fundamental problems.
- The method's originators have substantially misinterpreted what the method does.

What is the implication and what should change now?

- The credibility ceiling method should not be used.
- We hope that the method's originators will revise the method based on our critiques, or else directly rebut our mathematical demonstration of the method's flaws.

estimate appropriately toward the null. The credibility ceiling method leaves the studies' point estimates unchanged and instead affects the meta-analytic point estimate only indirectly through the increased variances. That is, the method essentially penalizes certain studies by increasing their variances; because the meta-analytic estimate is calculated by giving more weight to studies with small variances, the penalized studies' influence on the meta-analytic estimate is reduced. Thus, the credibility ceiling method affects the meta-analytic estimate only by moving it closer to the point estimates in the studies that were not penalized through the variance inflation.

We first provide some intuition for the severely misleading conclusions that can result from this approach and then explain why the method comes to these incorrect conclusions. As a simple thought experiment, suppose that all meta-analyzed studies have the same point estimate of 1, so the naïve meta-analytic estimate before any correction for unmeasured confounding is also 1. Now suppose we apply the credibility ceiling method. Regardless of the studies' sample variances, the meta-analytic point estimate will still be exactly 1 after applying the credibility ceiling simply because all studies have the same point estimate, so it does not matter how studies are weighted in the meta-analysis via their variances. In fact, the meta-analytic estimate will remain unchanged in this case regardless of the severity of unmeasured confounding or other

biases we choose to consider (i.e., regardless of the choice of c), which would seem to suggest incorrectly that the meta-analytic point estimate itself is entirely unaffected by bias, although its confidence interval and P -value will become less precise. Again, this incorrect conclusion arises because the credibility ceiling was not derived as a valid bias correction, but rather was constructed as an ad hoc re-weighting of studies in the meta-analysis.

For the same reason, the credibility ceiling can even shift the meta-analytic estimate in the wrong direction. Suppose we have a meta-analysis of ten studies, all of whose estimates are biased away from the null: five whose point estimates and variance estimates are all equal to 1, and five whose point estimates are all equal to 1.5 and whose variance estimates are all equal to 5. The naïve meta-analytic estimate before any correction for unmeasured confounding is 1.08. If we apply the credibility ceiling with, for example, $c = 0.25$, only the first five studies are penalized via variance inflation, and the remaining five retain their original variance estimates. This leads to a "corrected" meta-analytic estimate of 1.15, which in fact larger than the naïve estimate! This would seem to suggest, again incorrectly, that accounting for bias would somehow increase rather than decrease the meta-analytic estimate. (The confidence interval and P -value in this case will indicate greater uncertainty after applying the credibility ceiling correction, but critically, the correction to the meta-analytic estimate is nevertheless in the wrong direction and therefore clearly does not represent a valid bias correction.)

These problems reflect statistical misunderstandings embedded in the credibility ceiling's framework. When determining which studies to penalize, the normal random variable u_i used to calculate the probability that each study's point estimate would be in the "wrong" direction. But what is meant by the "wrong direction"? What is relevant to causal inference, and what should be meant by the "wrong" direction, is disagreement between the study's confounded (biased) point estimate and its underlying, causal effect. That is, the point estimate is intended to estimate a causal effect, and discrepancies between the two are the logical target of bias correction or sensitivity analysis methods for unmeasured confounding. (For one possible approach to conduct such a sensitivity analysis, see our previous work [6].) However, the mechanics of the credibility ceiling method in no way consider disagreement between the study's point estimate and the underlying causal effect. Rather than playing the role of the causal effect against which disagreement in direction should be assessed, the variable u_i is instead, by its construction, another random variable with the same degree of bias as the study's point estimate itself. It is therefore not a representation of the causal effect. Thus, in the credibility ceiling method, the probability that is fixed to c cannot be interpreted, as the originators intended, as the probability that the study's point estimate is in the "wrong" direction in a meaningful sense.

As the originators themselves observe, the credibility ceiling specifically penalizes, via variance inflation, those studies with large, positive point estimates or small variances. This selective penalty arises directly from assuming a credibility ceiling that is unrelated to a study's point estimate or variance. Yet in reality, for a given degree of bias, credibility is in fact highest for studies with large point estimates and small variances—exactly the studies that the credibility ceiling penalizes. (We justify this point mathematically in the Supplement.) That is, the more severe unmeasured confounding or other bias becomes, the larger and more precise a study's estimate must be to provide credible evidence for causality; this is not an assumption but rather is a direct mathematical consequence of sampling theory. Furthermore, it is not reasonable to assume that the maximum credibility of any observational study can never surpass a fixed value because, in fact, the actual credibility is not bounded below any fixed “ceiling” regardless of the severity of bias (as we justify mathematically in the Supplement). Given these fundamental problems with the credibility ceiling method and the resulting potential for misleading conclusions, we strongly caution against its use.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2020.05.006>.

References

- [1] Salanti G, Ioannidis JP. Synthesis of observational studies should consider credibility ceilings. *J Clin Epidemiol* 2009;62:115–22.
- [2] Ioannidis JP. Commentary: adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. *Int J Epidemiol* 2011;40:777–9.
- [3] Papatheodorou SI, Tsilidis KK, Evangelou E, Ioannidis JP. Application of credibility ceilings probes the robustness of meta-analyses of biomarkers and cancer risk. *J Clin Epidemiol* 2015;68:163–74.
- [4] de Rezende LF, de Sá TH, Markozannes G, Rey-López JP, Lee IM, Tsilidis KK, et al. Physical activity and cancer: an umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. *Br J Sports Med* 2018;52:826–33.
- [5] Köhler CA, Evangelou E, Stubbs B, Solmi M, Veronese N, Belbasis L, et al. Mapping risk factors for depression across the lifespan: an umbrella review of evidence from meta-analyses and mendelian randomization studies. *J Psychiatr Res* 2018;103:189–207.
- [6] Mathur MB, VanderWeele TJ. Sensitivity analysis for unmeasured confounding in meta-analyses. *J Am Stat Assoc* 2019;115:163–72.