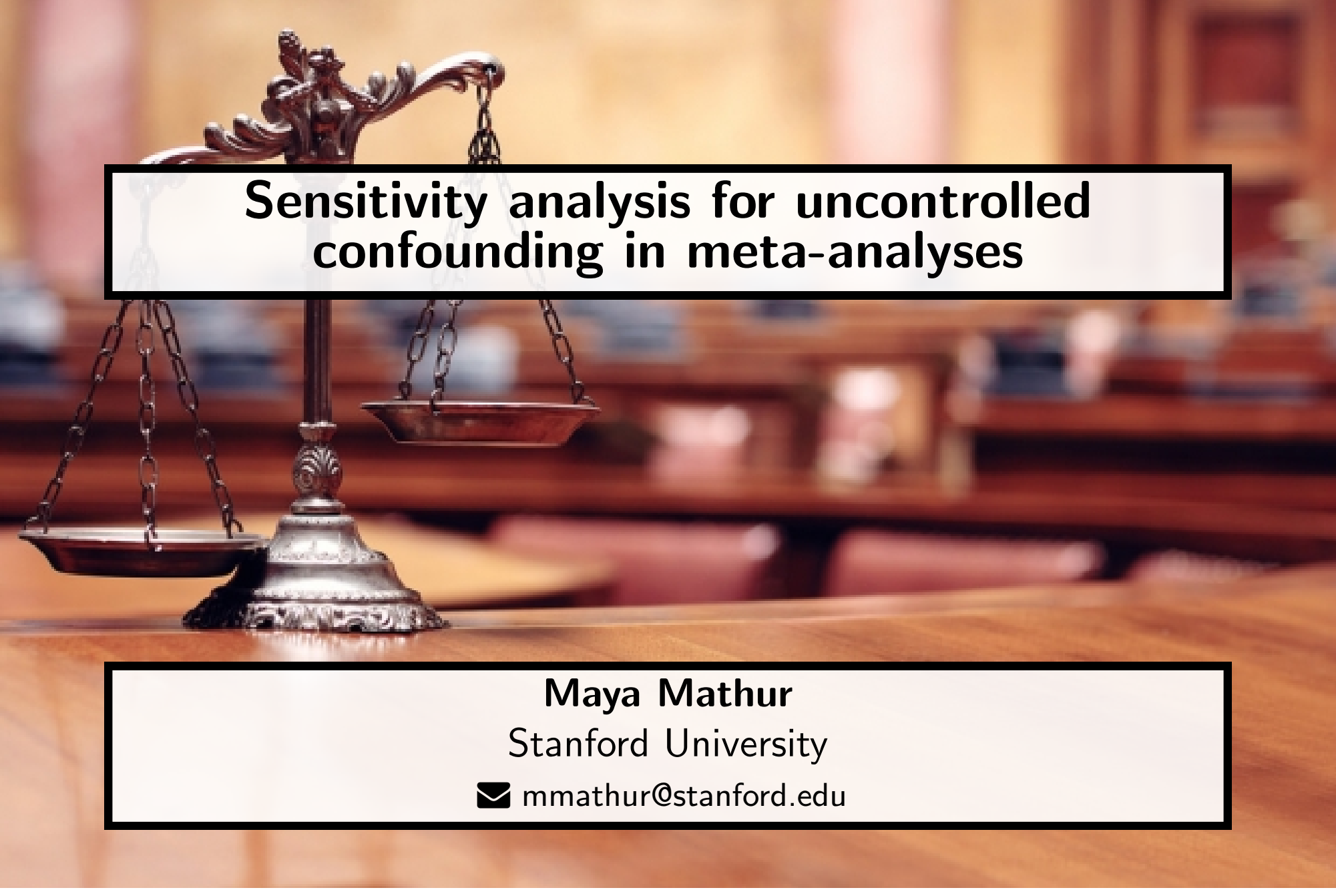


Sensitivity analysis for uncontrolled confounding in meta-analyses

Maya Mathur
Stanford University
mmathur@stanford.edu

Causal inference and confounding

Our focus today is on meta-analyses that address **causal questions** (not questions of association).

Studies that randomize the exposure usually provide the strongest evidence for causation. But often we cannot randomize.

In nonrandomized studies, any association we observe between the exposure and outcome might not be causal. It could be biased by uncontrolled confounders.

Causal inference and confounding

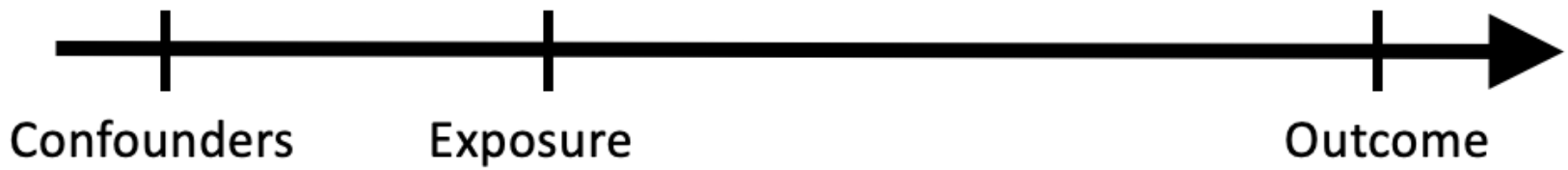
Confounders: Variables that affect both the exposure and the outcome, potentially producing a spurious association.

In individual studies, one way to eliminate bias from confounding is to measure confounders before the exposure occurred and statistically control for them:

- Regression adjustment
- Restricting analysis to subsets with the same values of confounders
- Propensity score methods
- Etc.

Causal inference and confounding

Also need a clear **temporal ordering** of confounders, exposure, and outcome.



Confounding in meta-analyses

Meta-analysts can try to reduce bias due to uncontrolled confounding when designing inclusion criteria:[1]

- Cross-sectional studies provide very weak evidence for causation unless temporal ordering of exposure and outcome is clear.
- Default starting point: Include only randomized studies plus longitudinal nonrandomized studies that (i) measure exposure before outcome and (ii) control for baseline confounders and baseline outcome.
- Can also stipulate that studies with weaker designs will be included only in secondary analyses.

[1] Mathur & VanderWeele (2022b); *Ann Rev Public Health*

Confounding in meta-analyses

Even with well-chosen inclusion criteria, meta-analyses of nonrandomized studies are often still at some risk of bias due to uncontrolled confounding.

Goal

Assess the strength of evidence for causality in meta-analyses of nonrandomized studies that are potentially subject to uncontrolled confounding.

How severe would uncontrolled confounding have to be to "explain away" the results of the meta-analyses?

We'll consider methods address this question in a way that doesn't require knowledge about the uncontrolled confounder(s).

Sensitivity analysis for a single study

VanderWeele TJ & Ding P (2017). Sensitivity analysis in nonrandomized research: Introducing the E-value. *Annals of Internal Medicine*.

Ding P & VanderWeele TJ (2016). Sensitivity analysis without assumptions. *Epidemiology*.

VanderWeele TJ & Mathur MB (2020). Commentary: Developing best-practice guidelines for the reporting of E-values. *International Journal of Epidemiology*.

Example: Benzodiazepines and dementia

Does long-term use of benzodiazepines increase the risk of dementia?

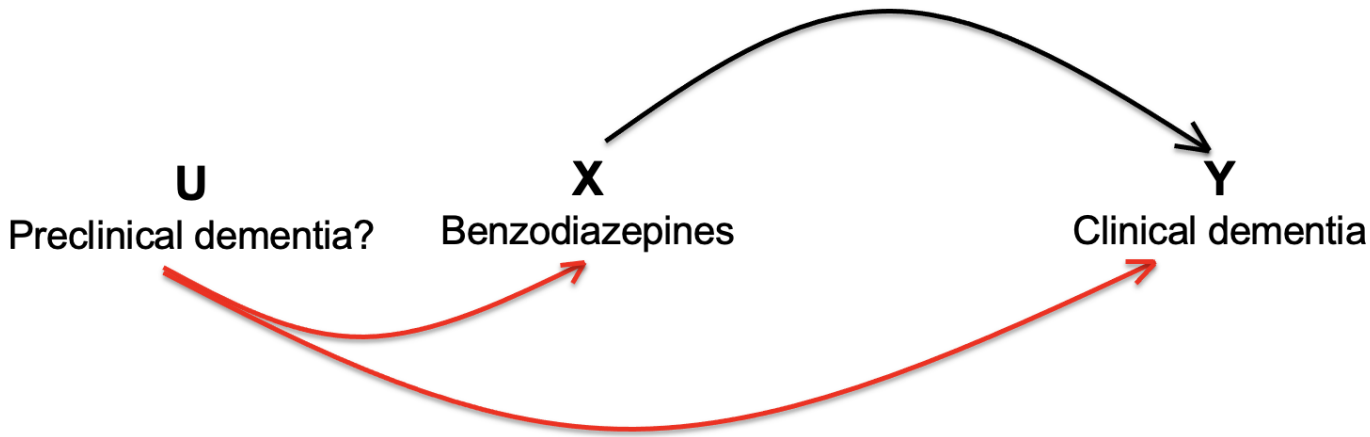
This question has been contentious, in part because of the potential for uncontrolled confounding.[2]

[2] Salzman (2020)

Example: Benzodiazepines and dementia

Preclinical dementia and related symptoms (e.g., sleep disturbances) could cause a patient to take benzodiazepines.

Also, preclinical dementia could later progress to full-scale dementia.



So preclinical dementia is a **confounder**. If not statistically controlled in a nonrandomized study, it might produce a spurious association between benzodiazepines and clinical dementia.

The E-value for the point estimate

Confounding RRs: The strength of association (RR scale) that uncontrolled confounder(s) have jointly with the exposure and/or outcome (conditional on measured confounders).

E-value: The minimum confounding RRs that uncontrolled confounder(s) would need to have with both exposure and outcome to fully explain away the observed effect (i.e., to have no causal effect).

$$\textbf{E-value} = RR + \sqrt{RR \cdot (RR - 1)},\ RR > 1$$

(For $RR < 1$, first take its inverse.)

Interpreting the E-value

Large E-value $\Rightarrow$ Only **severe** uncontrolled confounding could explain away the effect $\Rightarrow$ **robust** to uncontrolled confounding

Small E-value $\Rightarrow$ **Weak** uncontrolled confounding could potentially explain away the effect $\Rightarrow$ **not robust** to uncontrolled confounding

The E-value for the confidence interval

It's also good practice to report the E-value for the confidence interval.

That E-value represents the minimum strengths of confounding RRs that uncontrolled confounder(s) would need to have jointly with both exposure and outcome **to shift the confidence interval to include the null**.

In practice, the E-value for the CI can be obtained by applying the E-value formula $(RR + \sqrt{RR(RR-1)})$ to the CI limit that is closer to the null.

Example: Benzodiazepines and dementia

de Gage et al. (2012) conducted a longitudinal study of older adults, adjusting for numerous pre-exposure confounders:

$RR \approx 1.38$, 95% CI: $[1.05, 1.81]$

RRs are converted from the hazard ratio scale because the outcome was not rare.

Example: Benzodiazepines and dementia

E-value for point estimate: $1.38 + \sqrt{1.38 \cdot (1.38 - 1)} = 2.11$

E-value for CI limit: $1.05 + \sqrt{1.05 \cdot (1.05 - 1)} = 1.30$

"With an observed risk ratio of 1.38, uncontrolled confounder(s) that were associated with both benzodiazepine use and dementia by a risk ratio of 2.11-fold each, above and beyond the measured confounders, could **explain away the estimate**, but jointly weaker confounding associations could not. Uncontrolled confounder(s) that were associated with both benzodiazepine use and dementia by a risk ratio of 1.30-fold each, above and beyond the measured confounders, could **shift the CI to include the null**, but weaker confounding could not."

Example: Benzodiazepines and dementia

Now we have to think about these E-values **in scientific context and given the study design**.

Given the quality of existing confounding control (e.g., the study designs and confounders that were measured and controlled), is it **plausible or not plausible** that there actually were uncontrolled confounder(s) with confounding RRs of $RR = 1.30$ to 2.11?

Sensitivity analysis for meta-analyses

Mathur MB & VanderWeele TJ (2020). Sensitivity analysis for uncontrolled confounding in meta-analyses. *Journal of the American Statistical Association.*

Mathur MB & VanderWeele TJ (2020). Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. *Epidemiology.*

Mathur MB & VanderWeele TJ (2022). Methods to address confounding and other biases in meta-analyses: Review and recommendations. *Annual Review of Public Health.*

Is being overweight protective or detrimental?

Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories

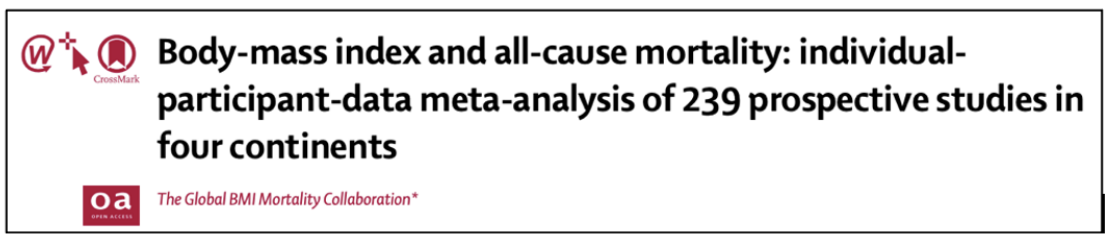
A Systematic Review and Meta-analysis

Katherine M. Flegal, PhD
Brian K. Kit, MD
Heather Orpana, PhD
Barry I. Graubard, PhD

Importance Estimates of the relative mortality risks associated with normal weight, overweight, and obesity may help to inform decision making in the clinical setting.

Objective To perform a systematic review of reported hazard ratios (HRs) of all-cause mortality for overweight and obesity relative to normal weight in the general population.

Beneficial effect of overweight:
HR = 0.94; 95% CI: [0.91, 0.96]



Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents

oa

*The Global BMI Mortality Collaboration**

Detrimental effect of overweight:
HR = 1.11; 95% CI: [1.10, 1.11]

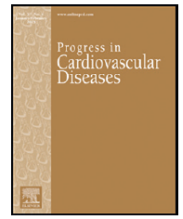
Flegal et al. (2013); Global BMI Mortality Collaboration (2016)

Is being overweight protective or detrimental?

Contents lists available at ScienceDirect

Progress in Cardiovascular Diseases

journal homepage: www.onlinepcd.com



The obesity wars and the education of a researcher: A personal account

Katherine M. Flegal *

Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

ARTICLE INFO

Available online xxxx

Keywords:
Obesity
Mortality
Controversy
Guidelines

ABSTRACT

A naïve researcher published a scientific article in a respectable journal. She thought her article was straightforward and defensible. It used only publicly available data, and her findings were consistent with much of the literature on the topic. Her coauthors included two distinguished statisticians. To her surprise her publication was met with unusual attacks from some unexpected sources within the research community. These attacks were by and large not pursued through normal channels of scientific discussion. Her research became the target of an aggressive campaign that included insults, errors, misinformation, social media posts, behind-the-scenes gossip and maneuvers, and complaints to her employer. The goal appeared to be to undermine and discredit her work. The controversy was something deliberately manufactured, and the attacks primarily consisted of repeated assertions of preconceived opinions. She learned first-hand the antagonism that could be provoked by inconvenient scientific findings. Guidelines and recommendations should be based on objective and unbiased data. Development of public health policy and clinical recommendations is complex and needs to be evidence-based rather than belief-based. This can be challenging when a hot-button topic is involved.

Is being overweight protective or detrimental?

In both meta-analyses, most studies (Flegal) or all studies (GBMC) **did not adjust** for probable confounders:

- Socioeconomic status
- Physical activity
- Diet
- Baseline body mass index (BMI)
- Etc.

How might we characterize sensitivity to uncontrolled confounding in these meta-analyses?

Flegal et al. (2013); Global BMI Mortality Collaboration (2016)

The GRADE approach

Current standard practice (e.g., required in Cochrane Collaboration reviews) is the GRADE approach:

1. Heuristically gauge "proportion of information" in meta-analysis contributed by studies at low vs. high risk of bias
2. Use this to decide whether to downgrade the overall certainty rating for meta-analysis from the default "high certainty" to "moderate" / "low" / "very low"
3. Can choose to upgrade rating again if, e.g., pooled estimate is large (GRADE suggests criterion of $RR > 2$ or $RR < 0.5$)

Higgins et al. (2019); Schünemann et al. (2019)

Current standard practice

This approach has limitations:

- Hard to intuit how much "information" each study contributes to the meta-analysis when studies' estimates and precisions differ.
- Deciding how to downgrade/upgrade overall certainty rating can be highly subjective.
- You ultimately get a qualitative rating of overall "certainty", rather than a quantitative summary of how numerical estimates might have been affected by bias.

To help address these limitations, let's look at **E-value analogs for meta-analyses.**

Sensitivity analysis for $\widehat{\mu}$

We can directly apply the E-value to the pooled estimate ($\widehat{\mu}$) in a meta-analysis.

This E-value represents the **average** confounding RRs that uncontrolled confounder(s) would need to have with studies' exposures and outcomes in order to shift $\widehat{\mu}$ to the null.

Mathur & VanderWeele (2020b), *JASA*; Mathur & VanderWeele (2022b); *Ann Rev Public Health*

Sensitivity analysis for $\widehat{\mu}$

Overweight and mortality

Strength of confounding RRs required to shift point estimates to null:[3]

- Flegal: $RR = 1.36$
- GBMC: $RR = 1.43$

And to shift each CI to include null:

- Flegal: $RR = 1.25$
- GBMC: $RR = 1.36$

Given studies' limited control of confounding, these are pretty small...

[3] Based on point estimates upon our own re-analysis, which differed negligibly from authors' own estimates.

Mathur & VanderWeele (2022a); *JAMA Network Open*

Sensitivity analysis for $\widehat{\mu}$

This E-value does not require assumptions on distribution of the population causal effects or the distribution of bias across studies provided that:[4]

- Any distributional assumptions of the meta-analysis model are met. If population *confounded* effects are non-normal, the meta-analysis method must accommodate this.[5]
- The bias in each study is independent of its population causal effect and its standard error.

[4] Mathur & VanderWeele (2020b), *JASA*; Mathur & VanderWeele (2022b); *Ann Rev Public Health*

[5] Hedges et al. (2010); Pustejovsky & Tipton (2021)

Limitations

Limitation of this simple E-value analog: Characterizes evidence strength in a meta-analysis only in terms of $\widehat{\mu}$ and its CI.

In addition to $\widehat{\mu}$, it is good to more holistically describe the potentially heterogeneous distribution of effects.

We can then conduct sensitivity analyses that describe the heterogeneous distribution of effects.

The proportion of meaningfully strong effects

One approach:

1. Choose a threshold representing a **meaningfully strong effect size** in scientific context (e.g., $HR > 1.1$)
2. Estimate $\widehat{P}_{>q}$, the proportion of population **effects above that threshold**

Can also estimate and report the proportion, $\widehat{P}_{<q^*}$, of population effects below a second (e.g., symmetrical) threshold on the other side of the null (e.g., $HR < 0.90$).

Mathur & VanderWeele (2019), *Stats in Med*; Mathur & VanderWeele (2020a), *Epidemiology*; R package MetaUtility

The proportion of meaningfully strong effects

These metrics can help identify if:

- There are few meaningfully strong effects despite a "significant" meta-analytic mean; or
- There are some strong effects despite a null meta-analytic mean; or
- Strong effects in the direction opposite the meta-analytic mean also regularly occur

---

Mathur & VanderWeele (2019), *Stats in Med*; Mathur & VanderWeele (2020a), *Epidemiology*; R package MetaUtility

The proportion of meaningfully strong effects

These proportions can be estimated using robust nonparametric procedures (based on shrinking studies' estimates toward $\widehat{\mu}$)[6] or using simple parametric estimates and inference:[7]

$$\widehat{P}_{>q} = 1 - \Phi\left(\frac{q - \widehat{\mu}}{\widehat{\tau}}\right), \widehat{\tau} > 0$$

$$\widehat{P}_{<q^*} = \Phi\left(\frac{q^* - \widehat{\mu}}{\widehat{\tau}}\right), \widehat{\tau} > 0$$

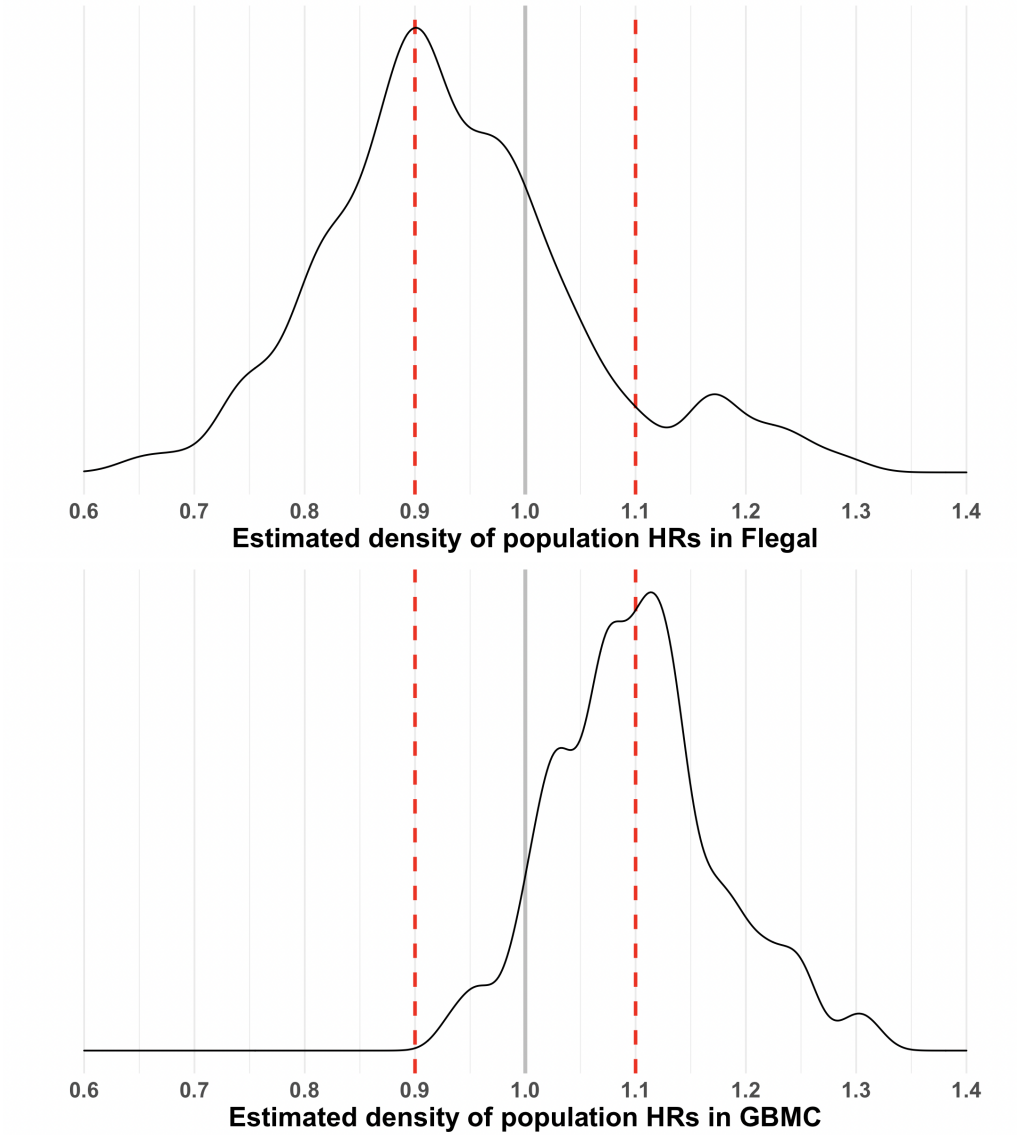
where $\widehat{\mu}$ is the meta-analytic mean, $\widehat{\tau}$ is the heterogeneity estimate (i.e., standard deviation of population effects), and q and q^* are chosen effect-size thresholds.

---

[6] Mathur & VanderWeele (2020a)

[7] Mathur & VanderWeele (2019)
R package `MetaUtility::prop_stronger`

The proportion of meaningfully strong effects



The proportion of meaningfully strong effects

Before considering confounding

Percentage of studies with meaningfully strong *protective* effects ($HR < 0.9$):

- Flegal: 40% [28%, 51%]
- GBMC: 0%

Percentage of studies with meaningfully strong *detrimental* effects ($HR > 1.1$):

- Flegal: 9% [4%, 15%]
- GBMC: 50% [34%, 63%]

Mathur & VanderWeele (2022a), *JAMA Network Open*

Sensitivity analysis for $\widehat{P}_{>q}$

Homogeneous bias

Let's now do a sensitivity analysis regarding $\widehat{P}_{>q}$ instead of $\widehat{\mu}$.

Rationale: When effects are heterogeneous, we might define "explaining away" the results of the meta-analysis in terms of substantially reducing the proportion of meaningfully strong effects.

$\widehat{\boldsymbol{G}}(\boldsymbol{r}, \boldsymbol{q})$: The minimum confounding RRs that uncontrolled confounder(s) would need to have with both the exposure and the outcome to reduce to less than r (e.g., 0.15) the proportion of studies with **causal** population effects stronger than q.

Sensitivity analysis for $\widehat{P}_{>q}$

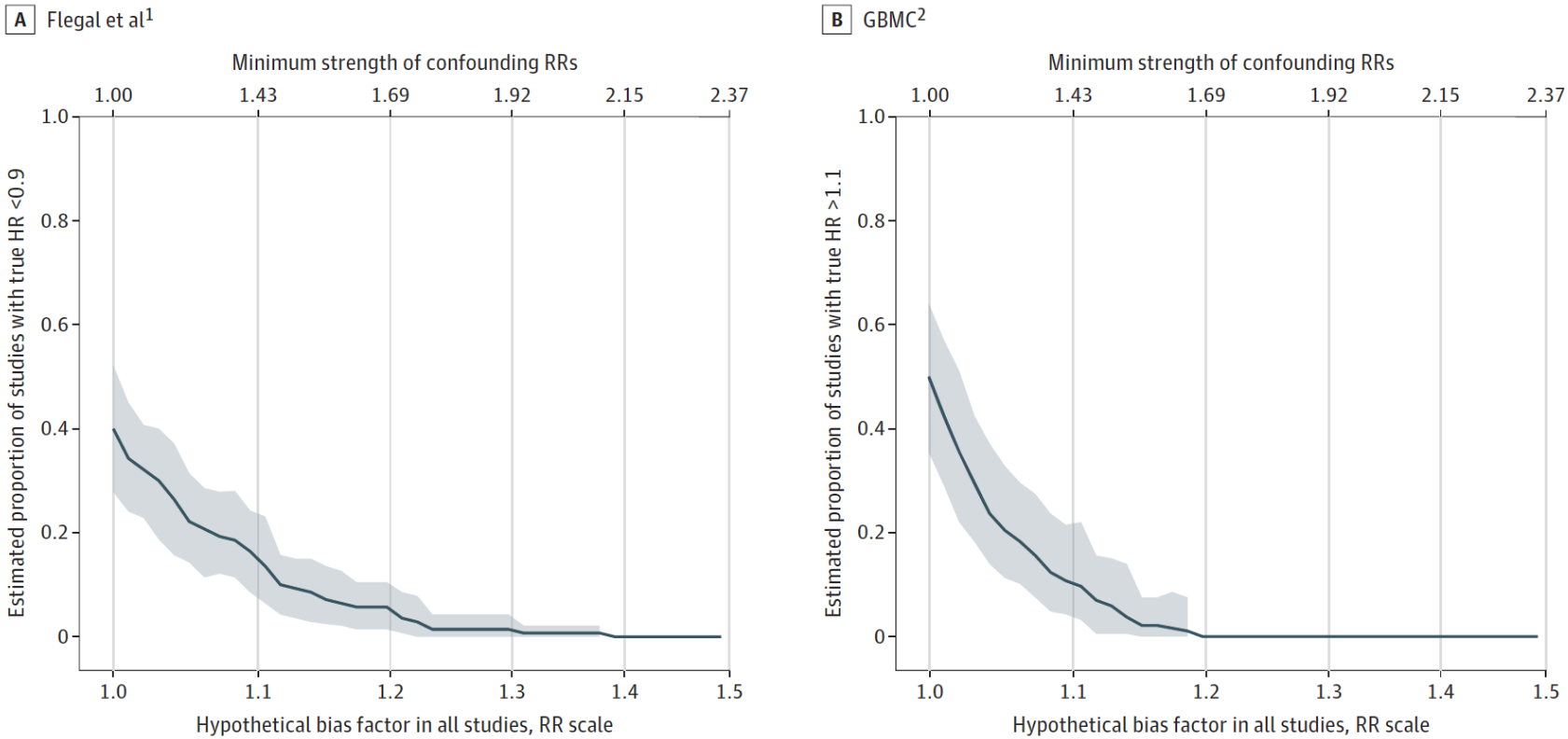
Homogeneous bias

Confounding RRs in each study required to reduce percentages of meaningfully strong protective HRs (Flegal) or detrimental HRs (GBMC) to only 15%:

- Flegal: 1.43
- GBMC: 1.10

Sensitivity analysis for $\widehat{P}_{>q}$

Homogeneous bias



Mathur & VanderWeele (2022a), *JAMA Network Open*

Sensitivity analysis for $\widehat{P}_{>q}$

Heterogeneous bias

Those analyses considered bias of homogeneous strength across studies. In certain settings,[8] this is statistically conservative.[9]

We can additionally consider bias that is heterogeneous across studies.

For these particular meta-analyses, considering bias that accounted for 80% of the estimated total between-study variance ($\widehat{\tau}^2$) yielded similar conclusions.[10]

[8] Depends on the choice of q relative to $\widehat{\mu}$

[9] Mathur & VanderWeele (2020b, 2022b)

[10] Mathur & VanderWeele (2022a), *JAMA Network Open*

Conclusions on overweight and mortality

For these 2 meta-analyses, confounding RRs of only 1.25 to 1.43 in each study could shift $\widehat{\mu}$ or its CI to null, or could substantially reduce $\widehat{P}_{>q}$.

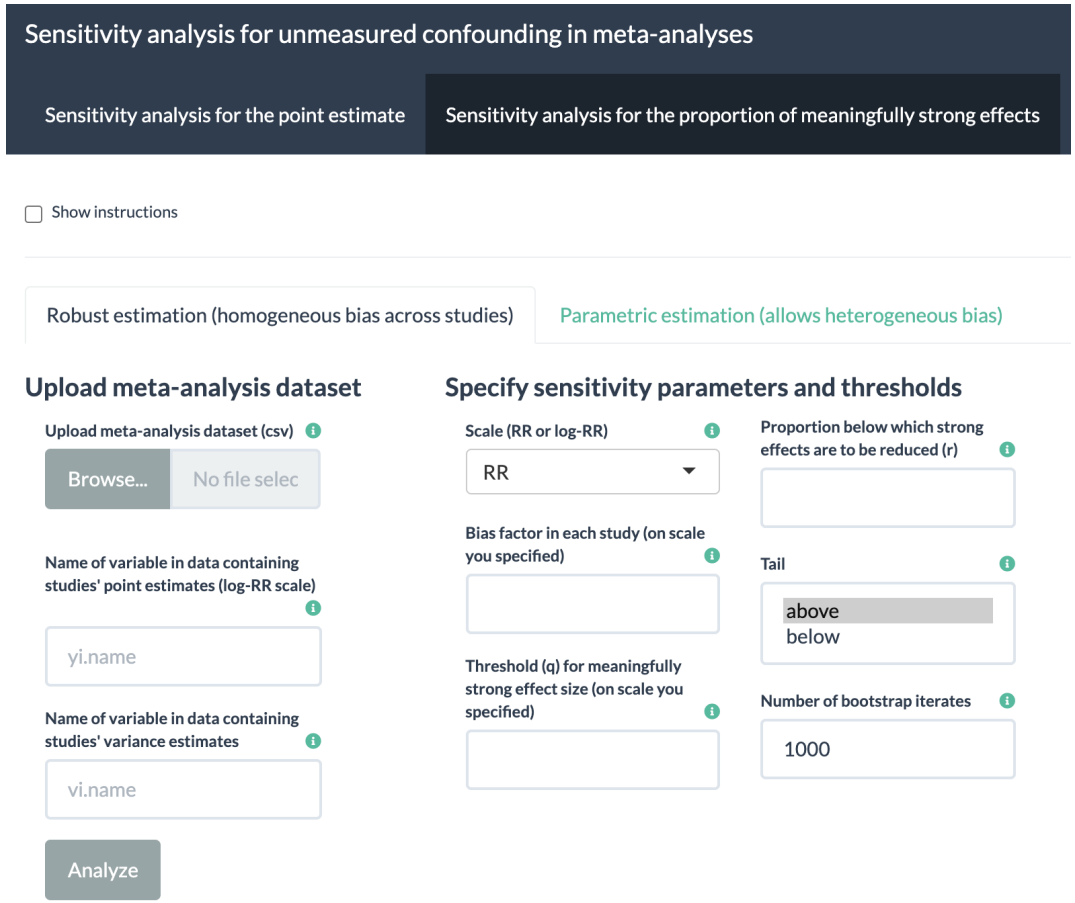
Given studies' limited control of confounding by physical, social, behavioral, and psychological factors, these sensitivity analyses suggest **neither meta-analysis provided robust evidence** for effects in either direction.

Establishing potentially small effects of being overweight on mortality would require improved study designs for primary studies and meta-analyses alike.

Mathur & VanderWeele (2022a), *JAMA Network Open*

Software

R: Function `confounded_meta` in package `EValue`
Online calculator, including plots: `evalue-calculator.com/meta`



Mathur et al. (2018), *Epidemiology*
Detailed tutorial in Mathur & VanderWeele (2022b), *Ann Rev Public Health*

Concluding remarks

Further guidance on within-study biases

In a review paper, we cover other sensitivity analysis methods and give overall recommendations:[11]

1. Pre-specify inclusion criteria that reduce risks of bias. Meta-analyses addressing causal questions should usually exclude cross-sectional studies.
2. Pre-specify which study designs will be included in primary vs. secondary analyses. Stratify on designs that provide substantially differing levels of evidence.
3. Qualitatively characterize risks of bias using existing risk-of-bias rating tools[12] and other summaries we suggest.
4. Quantitatively assess sensitivity to residual biases and interpret the results in light of the qualitative risk-of-bias assessments.

[11] Mathur & VanderWeele (2022b); *Ann Rev Public Health*

[12] Sterne et al. (2016)

Summary

Confounding bias in synthesized studies can propagate to meta-analysis estimates.

In addition to limiting bias through inclusion criteria and using existing tools to rate each study's risks of bias, it is helpful to quantitatively characterize the robustness of meta-analysis results to possible confounding.

We saw methods that are analogous to the E-value for individual studies and are straightforward to apply in practice.

Sometimes these analyses suggest that uncontrolled confounding is a serious threat to the credibility of meta-analyses (as in the meta-analyses on being overweight); other times, they suggest the opposite.

Slide deck (including reference list)

https://osf.io/68yp3/

Contact me

mmathur@stanford.edu

References

de Gage, S. B., Bégaud, B., Bazin, F., Verdoux, H., Dartigues, J.-F., Pérès, K., ... Pariente, A. (2012). Benzodiazepine use and risk of dementia: prospective population based study. *BMJ*, *345*.

Ding, P., & VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology*, *27*(3), 368.

Flegal, K. M., Kit, B. K., Orpana, H., & Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *Journal of the American Medical Association*, *309*(1), 71–82.

Global BMI Mortality Collaboration. (2016). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *The Lancet*, *388*(10046), 776–786.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Mathur, M. B., Ding, P., Riddell, C. A., & VanderWeele, T. J. (2018). Website and R package for computing E-values. *Epidemiology*, *29*(5), e45.

Mathur, M. B., & VanderWeele, T. J. (2019). New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, *38*(8), 1336–1342.

Mathur, M. B., & VanderWeele, T. J. (2020a). Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. *Epidemiology*, *31*(3), 356–358.

Mathur, M. B., & VanderWeele, T. J. (2020b). Sensitivity analysis for unmeasured confounding in meta-analyses. *Journal of the American Statistical Association*, *115*(529), 163–172.

Mathur, M. B., & VanderWeele, T. J. (2022a). Assessing uncontrolled confounding in associations of being overweight with all-cause mortality. *JAMA Network Open*, *5*(3), e222614–e222614.

Mathur, M. B., & VanderWeele, T. J. (2022b). Methods to address confounding and other biases in meta-analyses: Review and recommendations. *Annual Review of Public Health*, *43*.

Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, 1–14.

Salzman, C. (2020). *Do benzodiazepines cause Alzheimers disease?* Am Psychiatric Assoc.

Schünemann, H. J., Cuello, C., Akl, E. A., Mustafa, R. A., Meerpohl, J. J., Thayer, K., ... others (2019). 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.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... others (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, *167*(4), 268–274.

VanderWeele, T. J., & Mathur, M. B. (2020). Developing best-practice guidelines for the reporting of the E-value. *International Journal of Epidemiology*.