

Real and Illusory Precision Gains in Meta-Analysis to Speed Action on Carcinogens

Alexander P. Keil



ABSTRACT

One should avoid benzene exposure, all other things being equal. Risk assessment can help inform human health outcomes when all other things are not equal, as when competing legal or economic interests arise. In sparse literatures where exposures may be highly deleterious and yet understudied, there is a dire need for evidence synthesis, such as meta-analysis, to maximally inform risk assessment. Here, using the analysis and approach of Scholten and colleagues from the current issue as a touch point,

I describe how meta-analysis could ideally meet this aim and how it often fails to do so. Some of the current literature on transportability of causal effects is illustrative, and I describe how some of the lessons from this literature could be applied within the innovative framework of Scholten and colleagues to leverage meta-analysis within the broader decision-making framework of risk-assessment.

See related article by Scholten et al., p. 751

Meta-Analysis and Decisions

Given the choice between of whether or not to be exposed to compounds with known toxicity and no known health benefits, one should avoid exposure, all other things being equal. All other things are never equal, so the onus is often on toxicologists, epidemiologists, and statisticians to precisely estimate human health effects of exposures, and these estimates are considered along competing interests.

In this issue, Scholten and colleagues consider such an exposure with the explicit goal of estimating a “precise exposure–response curve (ERC)” relating benzene and acute myeloid leukemia (AML; ref. 1). The authors apply a clever Bayesian approach to evidence synthesis across animal and human domains, a necessary exercise due to the paucity of human outcomes data on benzene and AML.

Relative to informal ways of combining data across human cancer outcome, human biomarker intermediate, and animal domains, the approach of Scholten and colleagues injects needed rigor into evidence synthesis and summary. Here, I discuss their approach as a form of evidence synthesis and review recent challenges to the use of traditional meta-analytic estimates as “the” summary estimate to pass on to risk analysis. The ideas described presumably apply to any sparse literature, such as those from emerging hazards.

Whose Curve?

Risk assessment is fundamentally a problem of making causal inferences. That is, we have (i) some target population in mind, and (ii) target comparisons of cancer outcomes we would observe in that population, had they been unexposed, with the same population had they been exposed at some level/duration/etc. The focus of meta-analysis is often on part (ii), or the “target parameter,” which in the case

of Scholten and colleagues is a flexible ERC. However, this target parameter crucially depends on the “target population” (2). If, say, we ran a randomized trial of benzene exposure and measured subsequent AML, we would generally expect very different results if that trial were conducted in 90-year-olds versus 40-year-olds, or, say, smokers versus nonsmokers.

Accordingly, Scholten and colleagues explicitly define a target population in the “impact assessment” section of the paper: the general (human) population of the Netherlands, unexposed until age 20. However, if the meta-analyzed parameter itself differs across populations according to (say) the age distribution of the target population, then it is crucial to define the target population for which the meta-analyzed result applies. For example, in risk analysis we would like to know whether we might expect more- or less-severe effects for local worker populations that differ by average age (due to local labor conditions) or local worker populations with distinct genetic lineages that impact baseline disease rates. If we ignore these factors during meta-analysis, then we lose the ability to make impact assessments that account for differences in these factors across populations. Consequently, the rate ratios that inform such impact assessments may not be appropriate for the specified target population.

What is a Summary Effect?

Somewhat more formally, a fixed-effect meta-analysis seeks to combine various estimates of an “invariant” parameter, or a parameter that we truly do believe will be constant, except for random error, across different studies. Such estimates are “transportable” in that we believe, for example, that the underlying stratum-specific ERC (for say, ages 40–50 among employed males) will be identical across all studies, even if the estimated conditional ERCs vary across studies due to random error (3). In such scenarios, heterogeneity of “overall” ERCs across studies may be attributed to different distributions of age, gender, and employment status, rather than variations of the invariant, stratum-specific ERCs. The simple tool of direct standardization readily allows generation of an overall exposure response curve for a given target population with unique age/gender/employment distributions, provided that we have such stratum specific estimates (4). Thus a summary ERC for a given target population could be derived via standardization with meta-analyzed stratum-specific ERCs. In those ideal circumstances, precision gains from meta-analysis represent

Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Corresponding Author: Alexander P. Keil, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, McGavran-Greenberg Hall, Chapel Hill, NC 27516. Phone: 275-997-435; E-mail: akeil@unc.edu

Cancer Epidemiol Biomarkers Prev 2022;31:695–7

doi: 10.1158/1055-9965.EPI-21-1286

©2022 American Association for Cancer Research

genuine reductions in random error, which seems to be the point of meta-analysis.

The “ideal circumstances” caveat of interpreting meta-analyzed parameters is large, especially in observational studies. There seems to be an implicit hope that the systematic biases of observational studies (e.g., confounding) will cancel out, even though some biases have predictable directions, like healthy worker survivor bias (5, 6) or sparse-data bias (7). Given heterogeneity of overall ERCs across studies, there is also seemingly a hope that one can use a random-effects meta-analysis to target a kind of “average-study” type estimate (8). Such a summary effect is interpreted as the expected result of a new study performed with some unknown mish-mash of study characteristics that lead to heterogeneity in the first place. Scholten and colleagues rightly use prediction intervals to summarize their meta-regression estimates. However, such an approach treats design and analysis issues as random error that inflate prediction intervals, rather than as factors that can potentially be controlled within meta-regression (9).

In a random effects meta-analysis, heterogeneity is modeled from available studies, so the implied target population depends on which studies are included. In the case of Scholten and colleagues’ analysis, the summary effect is the overall ERC taken from a randomly drawn study population, which may not be fully human. If a goal of meta-analysis is to maximally leverage data for accurate human cancer risk analysis, that goal is at risk if the ERC varies by species/endpoint. The results of Scholten and colleagues suggest that it does (Table 3). Gains in precision from including animal/biomarker data are pyrrhic if they make the subsequent predictions less applicable to human cancer outcomes.

The Double-Edged Bayesian Sword

Bayesian meta-analysis, especially under a random-effects model, is appealing because of the coherence between Bayesian modeling and the multilevel nature of random-effects models. It may also be appealing in the case in which animal or biomarker studies operate as sources of priors, rather than as data to calculate a likelihood. In a loose sense, for a Bayesian meta-analysis there is little difference between likelihoods and priors, such that the approach of Scholten and colleagues seems a clever way to best represent prior evidence from animal and biomarker studies.

The necessity of prior specification and justification, including random effects covariances, dulls the luster of Bayesian analysis. The sensitivity analysis in the appendix shows how wrought this exercise can be. Specifically, changing the Cauchy scale prior for the random effect variance from the preferred (i.e., main analysis) 1.0 to a less informative 5.0 nearly doubled (1.8×) the prediction interval ratio, indicating the prior is highly influential on the estimated level of heterogeneity in the meta-analysis. The preferred prior forces the animal and human studies to appear more similar than the data suggest they are, running counter to some standard advice (10). When seemingly all relevant studies on a topic are included as data, precision gains from informative priors seem illusory—where does this prior knowledge come from, after all?

Lessons from Meta-Transportability

It is a truism in the causal inference methodology literature that the hardest work of causal inference is not in the methods, but in the background learning and study design/implementation (11). In this

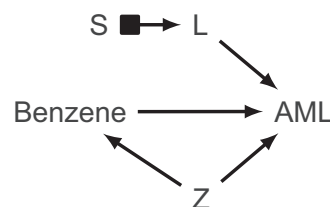


Figure 1.

A “selection diagram” demonstrating potential confounding by some factors “Z” that affect Benzene exposure and AML as well as potential heterogeneity sources “L” that vary according to study “S” and may be useful in Bayesian meta-regression for explaining among-literature heterogeneity.

vein, Scholten and colleagues painstakingly harmonize their data in terms of what is estimated and measured. Some harmonization steps (for example: equating disparate statistical parameters) are inconsequential under reasonable assumptions, whereas others (combining estimates from acute doses with those from chronic, longitudinal exposures) are on shakier ground and may contribute to apparent heterogeneity (9).

Looking ahead, the literature on transportability and meta-transportability suggests some best practices for applying innovative methods like those of Scholten and colleagues for the purpose of providing summary ERCs. Tools such as transport diagrams (12) are crucial in this endeavor (Fig. 1). These tools force us to be explicit about how biases such as confounding might operate in our studies (“Z” in the diagram: factors that influence benzene exposure and AML), but they also force us to be explicit about sources of heterogeneity vary across studies (the arrow between “S”, or study, and “L”, which could represent, for example, AML-related genetic disorders that vary across studies but are not otherwise related to exposure). One potentially obvious “L” to include in Scholten and colleagues’ analysis is “animal” vs. “human” vs. “biomarker,” based on differences in summary estimates across different sets of studies (their Table 3). Higher-dimensional L are good candidates for random-effects modeling (13).

The ease with which Bayesian modeling approaches like that of Scholten and colleagues can handle complexity gives immense promise for allowing explicit inclusion of factors in the meta-regression that may crucially impact heterogeneity of the true underlying ERC (14). Explaining heterogeneity via measured factors, rather than assuming it is random (9), puts us on better theoretic grounds for estimating a summary effect for risk analysis (2). Scholten and colleagues’ approach seems to be a clear step on the pathway toward better evidence synthesis in sparse literatures, but we are not there yet. Future implementations could benefit, both in precision and applicability, by explaining heterogeneity rather than letting it come out in the meta-analytic wash.

Author’s Disclosures

A.P. Keil reports grants from NIEHS during the conduct of the study.

Acknowledgments

This work was funded by NIH/NIEHS (grant no. R01 ES 029531, to A.P. Keil).

Received November 10, 2021; revised February 3, 2022; accepted February 8, 2022; published first April 4, 2022.

References

- Scholten B, Portengen L, Pronk A, Stierum R, Downward GS, Vlaanderen J, et al. Estimation of the exposure response relation between benzene and acute myeloid leukemia by combining epidemiologic, human biomarker, and animal data. *Cancer Epidemiol Biomarkers Prev* 2022;21:751–7.
- Dahabreh IJ, Petito LC, Robertson SE, Hernán MA, Steingrimsson JA. Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiology* 2020;31:334–44.
- Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Statistical Science* 2014;29:579–95.
- Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;60:578–86.
- Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology* 1994;5:189–96.
- Buckley JP, Keil AP, McGrath LJ, Edwards JK. Evolving methods for inference in the presence of healthy worker survivor bias. *Epidemiology* 2015;26:204–12.
- Richardson DB, Cole SR, Ross RK, Poole C, Chu H, Keil AP. Meta-analysis and sparse-data bias. *Am J Epidemiol* 2021;190:336–40.
- Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J Roy Stat Soc Ser A* 2009;172:137–59.
- Savitz DA, Forastiere F. Do pooled estimates from meta-analyses of observational epidemiology studies contribute to causal inference? *Occup Environ Med* 2021;78:621–2.
- Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 2006;1:515–34.
- Rubin DB. For objective causal inference, design trumps analysis. *Annals Appl Stat* 2008;2:808–40.
- Pearl J, Bareinboim E. Transportability of causal and statistical relations: a formal approach. In: *Proceedings of the 25th AAAI Conference on Artificial Intelligence*; 2011 Aug 7–11.
- Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics* 2000;56:915–21.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis*. Third edition. Boca Raton, FL: CRC Press; 2014.