

Reply to: Concerns about the Burden of Proof studies

Received: 25 January 2023

Accepted: 8 March 2023

Published online: 14 April 2023

 Check for updates

Aleksandr Y. Aravkin^{1,2,3}, Susan A. McLaughlin¹, Peng Zheng^{1,2}, Haley Lescinsky¹, Michael Brauer^{1,2,4}, Simon I. Hay^{1,2} & Christopher J. L. Murray^{1,2}✉

REPLYING TO A. J. Glenn et al. *Nature Medicine* <https://doi.org/10.1038/s41591-023-02294-8> (2023)

We, together with the Global Burden of Disease Study and Burden of Proof (BoP) collaborators, have published a BoP Capstone Methods paper and suite of associated meta-analyses that introduce and apply a new framework for synthesizing evidence to evaluate relationships between selected risk factors and health outcomes^{1–5}. The BoP approach was designed to address key problems in current analytical frameworks. To provide helpful information to users making decisions around risk exposure, the approach systematically estimates a flexible mean risk–outcome function, avoiding making strong assumptions such as log-linearity of the relationship. Complementing the mean risk–outcome relationship, BoP analysis provides a BoP risk function (BoPRF) that represents the lowest estimate of excess harmful risk associated with a risk factor, incorporating unexplained between-study heterogeneity after accounting for known variation in study design characteristics. From the BoPRF, we calculate summary risk–outcome scores and star-rating measures that present conservative estimates of the risk–outcome relationship and enable comparisons across different risk–outcome pairs.

In their Matters Arising on the BoP papers, Glenn et al.⁶ argue that our inclusion of between-study heterogeneity improperly inflates uncertainty intervals and that our star-rating summary measure is too simplistic. They further suggest that we wrongly assume risk–outcome relationships to be non-log-linear. We welcome this opportunity to engage in discourse and clarify misunderstandings. We all agree on the importance of using methods that summarize the available risk–outcome evidence in a way that most accurately captures the health risks people experience because of risk factor exposures.

With respect to the critique by Glenn et al. that our method improperly inflates uncertainty intervals, we note that in classic meta-analysis methods, between-study heterogeneity affects posterior uncertainty only marginally by enlarging the reported standard errors. As studies accumulate in the literature, even if each study contradicts the previous one, classic posterior uncertainty around effect size estimates continues to shrink, making the basis for health risk guidance ever more certain despite apparent disagreement between studies.

The phenomena of strong recommendations in spite of highly variable conclusions of emerging studies lead, justifiably, to broad critiques of the entire field⁷. The status quo in meta-analytic research—where the inclusion of more studies creates increased precision in the estimate even if they are divergent—is counterintuitive. Classic confidence intervals by themselves often fail to capture what matters to scientists and the public because they largely ignore variation in estimates of effects between studies. We believe the BoP approach addresses this problem by providing a complementary summary measure that incorporates between-study heterogeneity to obtain a prediction interval at the study level¹. This interval answers the question: ‘If a large criterion standard study was conducted today, how varied might its findings be, based on all available information?’ In our reports^{1–5} and visualizations (<https://vizhub.healthdata.org/burden-of-proof/>), we present both classically reported uncertainty intervals and the BoP uncertainty interval. The BoP approach provides new insight, allowing users to differentiate between research fields where results are consistent from those where they are not.

Specific study results may vary in part due to study design, length of follow-up and other topic-specific observable variables that correlate exactly to categories that Grading of Recommendations, Assessment, Development, and Evaluations^{8,9} and the Cochrane collaboration¹⁰ currently use to evaluate study quality. The BoP approach provides a mechanism to account for differences between study-level characteristics, to the extent that such differences are known and can be encoded to explain between-study heterogeneity. Any remaining unexplained between-study heterogeneity then contributes to the overall rating of effect size and evidence strength. Unexplained variation in results that remains after accounting for study-level covariates can (and should) contribute to a lower risk score and lower star ratings. We believe that users of the results will appreciate this transparency.

Glenn et al. incorrectly conclude that we assume that relationships between red meat consumption and chronic disease are not log-linear. A central principle in the BoP approach is to forgo the assumption that risks are log-linear; however, this is very different

¹Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA. ²Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle, WA, USA. ³Department of Applied Mathematics, University of Washington, Seattle, WA, USA. ⁴The University of British Columbia, Vancouver, British Columbia, Canada. ✉e-mail: cjlm@uw.edu

from assuming risks are ‘not log-linear’. As simulation studies in the Capstone paper¹ clearly show, when the dose–response relationship is log-linear, we recover that shape. A major contribution of the BoP method is that it allows more nuanced assessment of the underlying risk, letting the data inform the shape of the relationship rather than the analyst imposing it. Applying the BoP approach to analyze the risk–outcome pairs included in the Global Burden of Disease Study, we see that most relationships are not in fact log-linear, but some are, including systolic blood pressure versus heart disease (at least, for systolic blood pressure between 120 and 170 mmHg)⁴. We suggest that indiscriminately assuming log-linearity is a generic problem in many canonical systematic reviews^{11–15}, which may assume a log-linear relationship for computational convenience. Shortcomings of previous methodological attempts to account for non-log-linear relationships are discussed in the methods appendix of the Capstone paper¹ and illustrated using simulation.

Because of the inherent limitations around correcting for study-level bias in meta-analyses and, in terms of dietary risk analyses, including studies that do not carefully consider alternative foods or conduct substitution analyses, Glenn et al. suggest that a better approach would be to obtain primary, individual-level data for all original studies and use standard methods to define variables and adjust for covariates. Pooled studies can be helpful for exploring bias and answering myriad nuanced questions that can be addressed with primary detailed data shared among investigators. However, depending on how pooled studies are conducted, the results may also mask a large amount of heterogeneity in findings across studies that deserves explanation. We share Glenn et al.’s concern about the underlying literature and note it as a limitation¹. The critique stated here applies to studies that do not consider diet substitution (that is, most available studies) and to any meta-analyses that use such studies. Unfortunately, detailed individual-level information is rarely available, although ongoing efforts to synchronize and collaborate between investigators are commendable and promising. Nonetheless, most work in meta-analysis relies on extracting information across studies where only results were reported and individual-level data are unavailable.

We respect that these investigators feel strongly about the health risks of red meat consumption and recognize that they are looking for an explanation as to why the results of the BoP analysis do not agree with their previous beliefs. Hypotheses stand only until they are refuted; therein lies the dynamic of scientific debate and progress. There are certainly groups that feel differently from Glenn et al.^{16,17}, not to mention the disagreements surrounding myriad other risk–outcome pairs^{11–17}. The BoP approach incorporates divergent study results and study designs, accounts for explainable differences and allows comparison between strength of evidence across different scientific domains.

References

- Zheng, P. et al. The Burden of Proof studies: assessing the evidence of risk. *Nat. Med.* **28**, 2038–2044 (2022).
- Lescinsky, H. et al. Health effects associated with consumption of unprocessed red meat: a Burden of Proof study. *Nat. Med.* **28**, 2075–2082 (2022).
- Dai, X. et al. Health effects associated with smoking: a Burden of Proof study. *Nat. Med.* **28**, 2045–2055 (2022).
- Razo, C. et al. Effects of elevated systolic blood pressure on ischemic heart disease: a Burden of Proof study. *Nat. Med.* **28**, 2056–2065 (2022).
- Stanaway, J. D. et al. Health effects associated with vegetable consumption: a Burden of Proof study. *Nat. Med.* **28**, 2066–2074 (2022).
- Glenn, A. J., Gu, X., Hu, F. B., Wang, M. & Willett, W. C. Concerns about the Burden of Proof studies. *Nat. Med.* <https://doi.org/10.1038/s41591-023-02294-8> (2023).
- Ioannidis, J. P. A. The challenge of reforming nutritional epidemiologic research. *JAMA* **320**, 969–970 (2018).
- Siemieniuk, R. & Guyatt, G. What is GRADE? *BMJ Best Practice* <https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/> (2021).
- Schunemann, H., Brozek, J., Guyatt, G. & Oxman, A. *GRADE Handbook* (GRADE Working Group, 2013).
- Higgins, J. P. et al. *Cochrane Handbook for Systematic Reviews of Interventions* (John Wiley & Sons, 2019).
- Larsson, S. C. & Orsini, N. Red meat and processed meat consumption and all-cause mortality: a meta-analysis. *Am. J. Epidemiol.* **179**, 282–289 (2014).
- Abete, I., Romaguera, D., Vieira, A. R., Lopez de Munain, A. & Norat, T. Association between total, processed, red and white meat consumption and all-cause, CVD and IHD mortality: a meta-analysis of cohort studies. *Br. J. Nutr.* **112**, 762–775 (2014).
- Farvid, M. S. et al. Consumption of red meat and processed meat and cancer incidence: a systematic review and meta-analysis of prospective studies. *Eur. J. Epidemiol.* **36**, 937–951 (2021).
- Kazemi, A. et al. Intake of various food groups and risk of breast cancer: a systematic review and dose–response meta-analysis of prospective studies. *Adv. Nutr.* **12**, 809–849 (2021).
- Zhang, R., Fu, J., Moore, J. B., Stoner, L. & Li, R. Processed and unprocessed red meat consumption and risk for type 2 diabetes mellitus: an updated meta-analysis of cohort studies. *Int. J. Environ. Res. Public Health* **18**, 10788 (2021).
- Kolata, G. Eat less red meat scientists said. Now some believe that was bad advice. *The New York Times* (30 September 2019).
- Zeraatkar, D. et al. Red and processed meat consumption and risk for all-cause mortality and cardiometabolic outcomes: a systematic review and meta-analysis of cohort studies. *Ann. Intern. Med.* **171**, 703–710 (2019).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Author contributions

A.Y.A. and S.A.M. managed the estimation or publications process. A.Y.A. and S.I.H. wrote the first draft of the manuscript. H.L. had primary responsibility for applying the analytical methods to produce estimates. H.L. had primary responsibility for seeking, cataloging, extracting or cleaning data and designing or coding figures and tables. M.B. provided data or critical feedback on data sources. A.Y.A., P.Z. and M.B. developed the methods or the computational machinery. A.Y.A., P.Z., M.B. and C.J.L.M. provided critical feedback on the methods or results. A.Y.A., S.A.M., H.L., M.B., S.I.H. and C.J.L.M. drafted the work or revised it critically for important

intellectual content. A.Y.A. and S.I.H. managed the overall research process.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Christopher J. L. Murray.

Reprints and permissions information is available at www.nature.com/reprints.