A Study Protocol to Evaluate the Reliability of Scientific Citations Provided to Courts by Experts about Adverse Health Effects of Air Quality

S. Stanley Young and Warren B. Kindzierski

Problem statement

What is the size of the analysis search space of randomly selected studies cited in Expert testimony to courts about air quality−adverse health effects?

Narrative

In the epidemiological analysis of large, complex data sets, many questions of interest can arise, and each question can be examined by a statistical analysis (e.g., Gelman and Loken 2019, Young and Kindzierski 2019a,b).

P-hacking, a practice of manipulating data or analysis to find a publishable result, poses a significant threat in the context of air quality-adverse health effect research. With numerous questions asked of a data set, there is a high risk of p-hacking (Stefan and Schönbrodt 2023). If there is no statistical correction for multiple testing and multiple modeling, then bias can result. -This can lead to a multitude of false positive results being taken as true, which in turn can lead to a false claim being canonized (Smaldino and McElreath 2016, Nissen et al. 2016).

If a random sampling of air quality−adverse health effect scientific papers (observation studies) provided to courts by an Expert reveals a significant, statistical search space and there is no correction for multiple testing and multiple modeling bias, then the entire observation study literature cited by the Expert could be viewed as a part of a potentially false canonization of air quality−adverse health effect research claims.

For example, an expert provides the court with a list of 500 scientific papers claiming adverse health effects of air quality. A 5–20% sample from a population whose characteristics are known is considered acceptable for most research purposes as it allows for generalizations about the population (Creswell 2003).

It is of utmost importance to ensure that the 500 papers had sufficiently known characteristics by the Expert that are relevant to the court case (i.e., they are about some aspect of air quality−adverse health effects). Why else would these papers be offered to the court by the Expert? Given this assumption and the large number of papers (500), a random sampling and evaluation of ~5% (24) papers should be adequate to be able to describe a feature of interest common to all 500 scientific papers.

This feature is the statistical reliability of claims that poor air quality is associated with poor health conditions. A large analysis search space and the absence of correction for multiple testing and multiple modeling bias in the papers in the random sample implies irreproducibility (falseness) of claims in the scientific papers provided to the court.

Methods

The number of research questions asked (i.e., hypotheses test performed) will be counted out for representative papers to assess the reliability of claims that poor particulate matter (PM2.5) air quality is associated with adverse health.

1. Secure list of papers provided to the court.
2. Randomly select ~5% of the papers with the aim of choosing 24 suitable papers for review. If a paper is unacceptable (i.e., not a published paper, not dealing with air quality and human health effects, not accessible for review), the next paper on the list will be reviewed. We purposely excluded chamber, panel, and genetic studies as less relevant than environmental epidemiology studies.
3. Features to document during the review:
    a. 1st author.
    b. Email address of the corresponding author.
    c. Year of publication.
    d. Title.
    e. Funding: gov, industry, other.
    f. Note if authors make their data public.
    g. Note if any statistical analysis was undertaken to correct for multiple testing and multiple modeling bias.
4. Counts to collect (following Young and Kindzierski 2019b, 2022; Kindzierski et al. 2023):
    a. # of outcomes, O
    b. # of predictors, P
    c. # of covariates, C
5. Report the analysis search space (number of statistical hypotheses tests performed) = $O \times P \times 2^C$
6. Note if any of these selected key papers with negative findings about particulate matter−adverse health effects were cited (i.e., Styer 1995, Chay 2003, Enstrom 2005, Young 2017).

References

Chay, K., Dobkin, C., Greenstone, M. 2003. The Clean Air Act of 1970 and adult mortality. *Journal of Risk and Uncertainty,* 27, 279–300. https:// doi.org/10.1023/A:1025897327639

Creswell, J. 2003. *Research Design-Qualitative, Quantitative and Mixed Methods Approaches, 2nd ed.* Thousand Oaks, CA: Sage Publications.

Enstrom, J. E. 2005. Fine particulate air pollution and total mortality among elderly Californians, 1973–2002. *Inhalation Toxicology*, 17(14), 803–816. https://doi.org/10.1080/08958370500240413

Gelman, A., Loken, E. 2019. The garden of forking paths: why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking' and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/gelman/research/unpublished/p_hacking.pdf.

Kindzierski, W.B., Young, S.S., Dunn, J.D. 2023. Reliability of meta-analysis research claims for gas stove cooking−childhood respiratory health associations. *International Journal of Statistics and Probability*, 12(3), 40−57. https://doi.org/10.5539/ijsp.v12n3p40

Nissen, S.B., Magidson, T., Gross, K., Bergstrom, C.T. 2016. Publication bias and the canonization of false facts. *eLife*, 5, e21451. https://doi.org/10.7554/eLife.21451.001

Smaldino, P.E., McElreath, R. 2016. The natural selection of bad science. *Royal Society Open Science*, 3, 160384. http://dx.doi.org/10.1098/rsos.160384

Stefan, A.M., Schönbrodt, F.D. 2023. Big little lies: a compendium and simulation of p-hacking strategies. *Royal Society Open Science*, 10, 220346. https://doi.org/10.1098/rsos.220346

Styer, P., McMillan, N., Gao, F., Davis, J., Sacks, J. 1995. Effect of outdoor airborne particulate matter on daily death counts. *Environmental Health Perspectives*, 103, 5: 490–497. https://doi.org/10.1289/ehp.95103490

Young, S.S., Smith, R.L., Lopiano, K.K. 2017. Air quality and acute deaths in California, 2000-2012. *Regulatory Toxicology and Pharmacology*, 88, 173–184. https://doi.org/10.1016/j.yrtph.2017.06.003

Young, S.S., Kindzierski, W.B. 2019a. Combined background information for meta-analysis evaluation. https://arxiv.org/abs/1808.04408

Young, S.S., Kindzierski, W.B. 2019b. Evaluation of a meta-analysis of air quality and heart attacks, a case study. *Critical Reviews in Toxicology*, 49(1), 85–94. https://doi.org/10.1080/10408444.2019.1576587

Young, S.S., Cheng, K.-C., Chen, J.H., Chen, S.-C., Kindzierski, W.B. 2022. Reliability of a meta-analysis of air quality−asthma cohort studies. *International Journal of Statistics and Probability*, 11(2), 61–76. https://doi.org/10.5539/ijspv11n2p61