Statistical Science
2020, Vol. 35, No. 3, 476–478
https://doi.org/10.1214/20-STS776
Main article: https://doi.org/10.1214/19-STS728

A New Template for Empirical Studies: From positivity to Positivity

Rhian Daniel

1. GENERAL COMMENTS

The article by VanderWeele et al. [3] is a *tour de force* of innovation, education and pragmatism, and is a must-read for all students and researchers in statistical science and associated disciplines. As well as suggesting a new outcome-wide approach to the analysis of empirical studies, the article gathers together numerous other new or recent methodological and practical advances that are useful even outside the proposed framework. These include the modified disjunctive cause criterion for confounder selection, the E-value for sensitivity analysis, valuable new insights on well-known correction methods for multiple testing and a novel metric for the expected number of false positive findings. In addition, the paper comprehensively summarises the vast literature on confounder-adjusted analyses in a succinct, accessible and educational manner oozing with practical advice, even on how to report results in space-limited journals. As if consciously practising what they preach, the authors include in a single paper an exploration of almost all the associated issues, caveats, extensions, modifications and comparisons: a bells-and-whistles-wide methodological contribution that other authors may have split into a dozen papers or more.

When it comes to issues such as the so-called replication/reproducibility crisis and formal causal aspects of the analysis of observational studies, the awareness of problems and potential pitfalls are of course essential in engendering appropriate caution and humility. However, a *can't do* attitude ("$p < 0.05$ doth not a finding make", "correlation is not causation") is less likely to improve matters than a clear, concrete and implementable alternative approach, such as this one by VanderWeele et al.

Their central suggestion in a nutshell is that researchers concerned with a particular exposure should study (and report) its effect, not on a single outcome of interest to them, but rather on as wide a range of outcomes as is feasible. They discuss the advantages of doing so, many of which relate to research efficiency, reproducibility and lessening publication bias (by ensuring that more null results are published). An obvious but compelling advantage arises when an exposure (e.g., HRT) has a harmful effect on one outcome (e.g., cancer) and a protective effect on another (e.g., heart disease). Such results reported separately contribute to the adage that "today's poison is tomorrow's wonder drug". Many less obvious advantages, for example that some outcomes may plausibly serve as negative controls for others, are also compelling.

The article is (cautiously) positive about the extent to which well-conducted observational studies can offer evidence on cause–effect relationships, with many enlightening and nuanced discussions on the likely magnitude of biases arising from various sources in different contexts. In the remainder of this commentary, I will first mention a few additional minor notes of caution that came to mind whilst attempting to read the article under a commentator's hat. One of these concerns potential violations of the *positivity* assumption (the large-P Positivity in the title), which is discussed in more detail in Section 3. My overwhelming feeling towards the paper, however, is positive (the small-p positivity in the title).

2. A FEW MINOR CAUTIONARY NOTES

In several parts of the article, the authors describe situations in which an analyst will be faced with two suboptimal options. For example, in Section 2.5, when studying the effect of physical activity on cardiovascular disease, BMI is plausibly both a confounder and a mediator. If repeated measurements of BMI are available, one option is to adjust for BMI at a wave previous to the exposure measurement, risking residual confounding, and another is to adjust for a more recent measurement, risking adjusting for a partial mediator of the effect of interest. The authors suggest doing both in a sensitivity analysis. In this example, failure to fully adjust for the confounding through BMI will likely lead to *over*estimating the beneficial effect of physical activity, whereas adjustment for a measurement of BMI on the causal pathway from physical activity to cardiovascular disease will likely lead to *under*estimating the beneficial effect. The two estimates could thus plausibly be viewed as bounds, with the true effect lying somewhere between the two. It is worth mentioning as an additional caution, however, that in many other settings both analyses could be biased in the *same* direction. This will tend to happen whenever the covariate's effect on the exposure is in the opposite direction from the exposure's effect on the covariate (e.g., high

Rhian Daniel is a Reader in Medical Statistics, Division of Population Medicine, Cardiff University, Cardiff, United Kingdom (e-mail: danielr8@cardiff.ac.uk).

LDL cholesterol increases probability of statin use, which in turn reduces subsequent LDL cholesterol). When stuck between a rock and a hard place, we should not be lulled into thinking that the truth always lies somewhere between the two.

In Section 2.6 on adjusting for prior exposure, there is again a danger that we are lulled into a false sense of security. The authors write that "control for prior exposure can also help further rule out other forms of unmeasured confounding. This is so because, if control is made for prior exposure then, for an unmeasured confounder U to explain away an observed exposure-outcome association, the unmeasured confounder would have to be associated with both the outcome and the baseline exposure, independent of prior level of exposure". They point out too, of course, that whether or not one adjusts for prior exposure changes the question being addressed. The argument is compelling, but must be offset against the fact that the true effect of a single instance of the exposure on the outcome (controlling for prior exposure) will likely be considerably smaller in magnitude than the total effect of all instances of the exposure, and thus the reduction in unmeasured confounding achieved by adjusting for prior exposure may be comparable to the reduction in the targeted effect.

In Section 3.3, the authors recommend the use of multiple imputation for dealing with (relatively small amounts of) missing data in exposure and confounders. A well-documented pitfall with such an approach occurs if the researcher fails to include the outcome of interest in the imputation model [2], leading to a dilution of conditional associations between exposure/confounders and outcome. It should thus presumably be recommended in an outcome-wide approach to include *all* outcomes simultaneously in the imputation model for the exposure and confounders. There is a risk that, when dealing with the greater complexity of the outcome-wide approach, this important consideration may be overlooked.

As a final cautionary note, if the proposed new template were widely adopted in the health sciences, is there a danger that research from specific disease registries (e.g., cystic fibrosis, kidney transplant) would be overlooked due to their typically narrow (disease-specific) range of outcomes, despite their many strengths in other aspects? They typically include a very high percentage of patients with the relevant condition, almost all of whom are seen at regular and frequent (e.g., yearly) intervals, with accurate data on a rich set of relevant confounders and very little missing data. This tension will perhaps be removed in future as the data from such registries are increasingly incorporated into record-linked national data banks, from which a range of other outcomes could be obtained.

3. A CAUTIONARY NOTE ON POSITIVITY

In Section 3.2, the authors make explicit an argument to which others informally appeal, when comparing the plausibility of the no unmeasured confounding assumption across different settings. They write:

> "In many economic contexts it is assumed that agents have some degree of information about their own potential outcomes that is not available in the data for which measurements are available, and that the agents use this information to select into the treatment or exposure groups. ... In contrast, in a number of biomedical settings, the patient or participant may not have analogous information; it may be that the patients physician is the principal decision-maker concerning which treatment may be best, and that the information available to the physician is in fact roughly the same information available in the data to the researcher".

This is highly compelling, and for a number of biomedical settings, it may be entirely reasonable. However, in some clinical settings, the manner in which the clinician bases her decision on the measured confounder(s) is (at least in theory) deterministic, according to some clinical guidelines, such as "if age exceeds 50 and serum total cholesterol level exceeds 5.5mmol/l, prescribe statin therapy". Such a guideline, if strictly adhered to, leads to violations of the so-called *positivity* assumption [1], where, on the basis of their measured confounders, the probability of being exposed is either 0 or 1 for some/all patients. Causal inference (on the effects of statins) would be rendered impossible by such strong confounding, at least for the subgroups of patients for whom the positivity violations take place, even when the confounders are all observed and perfectly measured. In clinical settings, therefore, causal inference, which requires overlap between exposure groups in their levels of the measured confounders, relies on the fact that clinicians deviate from such guidelines; crucially, to avoid unmeasured confounding, these deviations need to be haphazard, as opposed to being driven by some degree of information about the patient's potential outcomes. There is perhaps often likely to be a trade-off in practice, with settings with relatively fewer problems of unmeasured confounding suffering from a greater problem with positivity violations, and *vice versa*.

To illustrate this (see Figure 1), where we have simulated toy data to characterise what might be expected in economic and clinical settings, respectively. In both settings, there is a continuous measured confounder C, a continuous unmeasured confounder U, a binary exposure A and a continuous outcome Y. The effect of C, U and A on Y are the same (all nonnull) in the two settings.

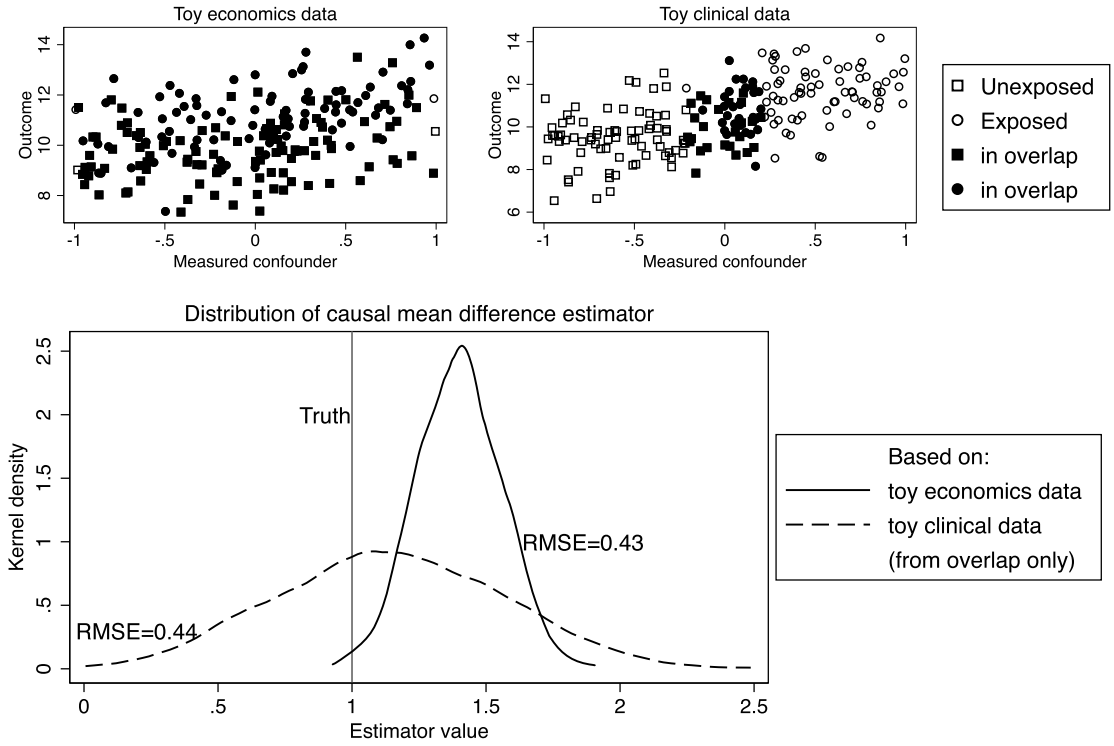

FIG. 1. *An illustration of a possible trade-off between unmeasured confounding and positivity violations when comparing typical data from economics and clinical medicine.*

In the economic setting, the effect of U on A is greater than the effect of C on A: unmeasured confounding is thus a significant problem. In the clinical setting, the effect of U on A is negligible, but the effect of C on A is very strong: positivity violation is thus a problem. At the top of Figure 1, example datasets, with the overlap illustrated, are shown. At the bottom of Figure 1, the distribution of the causal mean difference estimators over 1000 simulated datasets are compared, where all analyses are restricted to the region of overlap. The greater restriction in the clinical data leads to a larger variance, and also to an 'exaggeration' of the otherwise small bias due to unmeasured confounding. In these toy illustrative examples, the root mean squared error is similar in the two settings.

Such positivity violations will presumably be made more likely by the outcome-wide framework, given its policy to adjust for more confounders than is strictly necessary.

REFERENCES

[1] PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21** 31–54. MR2867537 https://doi.org/10.1177/0962280210386207

[2] STERNE, J. A. C., WHITE, I. R., CARLIN, J. B., SPRATT, M., ROYSTON, P., KENWARD, M. G., WOOD, A. M. and CARPENTER, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Br. Med. J.* **338** b2393.

[3] VANDERWEELE, T. J., MATHUR, M. B. and CHEN, Y. (2020). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. *Statist. Sci.* **35** 437–466.