Statistical Science
2020, Vol. 35, No. 3, 467–471
https://doi.org/10.1214/20-STS769
Main article: https://doi.org/10.1214/19-STS728

Comment: On the Potential for Misuse of Outcome-Wide Study Designs, and Ways to Prevent It

Stijn Vansteelandt and Oliver Dukes

We congratulate the authors, VanderWeele, Mathur and Chen (2020) (hereafter referred to as VMC), for making an interesting and important proposal, and thank the editor for the opportunity to comment on it. We agree with VMC that outcome-wide epidemiology has the potential to overcome many of the weaknesses of the traditional epidemiological approach. Scientific reports that express the effects of an exposure on a variety of different outcomes provide a more complete view of the exposure impact, while lessening the risk of selective analysis and reporting. We see much value in it, though caution is warranted. In this commentary, we highlight a number of key limitations, which will in turn suggest preferred analysis strategies that we find important to consider in addition to (or instead of) those described by VMC.

1. BIAS INFLATION

With the analysis of multiple outcomes comes a growing of risk of bias in the effect of the exposure on (at least one of) those outcomes. Such inflated risk of bias may be the result of the more elaborate need for modelling (e.g., modelling each outcome separately) and the ensuing risk of model misspecification, the increased risk of (informative) missing data in those outcomes, a potentially reduced lack of care in collecting data on risk factors for all these outcomes (see Section 3) or in modelling the outcomes' dependence on measured risk factors, ... This expresses itself in particular into an inflated risk of Type I errors. Such inflation is not acknowledged by multiplicity adjustments such as the Bonferroni correction, which assume the absence of bias.

Stijn Vansteelandt is Professor, Department of Applied Mathematics, Computer Sciences and Statistics, Ghent University, Krijgslaan 281, S9, 9000 Gent, Belgium (e-mail: stijn.vansteelandt@ugent.be). Stijn Vansteelandt is also affiliated as a part-time professor in the Department of Medical Statistics of the London School of Hygiene and Tropical Medicine, London, United Kingdom. Oliver Dukes is Postdoctoral Researcher, Department of Applied Mathematics, Computer Sciences and Statistics, Ghent University, Krijgslaan 281, S9, 9000 Gent, Belgium (e-mail: oliver.dukes@ugent.be).

To appreciate this, let $\hat{\theta}_j$ express the estimated effect of exposure on the jth outcome ($j = 1, \ldots, k$). Suppose that $\hat{\theta}_j$ is normally distributed around θ_j with standard deviation $\sigma/\sqrt{n}$, where n is the sample size. Suppose further that the exposure has no effect on any of the outcomes, but that θ_j is nonetheless normally distributed with mean θ and standard deviation τ, which may both differ from zero as a result of bias. Under the above settings, the probability to find the exposure being associated with at least one of k mutually independent outcomes at the $\alpha 100\%$ significance level, when Bonferroni correction is used, equals

$$1 - \left[\Phi\left\{\frac{-\Phi^{-1}(\alpha/2k) - \theta\sqrt{n}/\sigma}{\sqrt{1+\tau^2 n/\sigma^2}}\right\} - \Phi\left\{\frac{\Phi^{-1}(\alpha/2k) - \theta\sqrt{n}/\sigma}{\sqrt{1+\tau^2 n/\sigma^2}}\right\}\right]^k.$$

Figure 1 displays this for $n = 100$, $\sigma = 1$, $\alpha = 0.05$ and $\theta = 0$, $\tau = 0.1$ (left), amounting to bias up to 2 standard errors away from zero for most outcomes, $\theta = 0.1$, $\tau = 0$ (middle), amounting to bias of 1 standard error for all outcomes, and $\theta = 0.1$, $\tau = 0.1$ (right), amounting to bias between -1 and 3 standard errors away from zero for most outcomes. These figures visualise the growing risk of false detections that may result from an accumulated risk of bias across all outcomes.

In view of these concerns, it is essential in our opinion that outcome-wide epidemiologic analyses be based on propensity scores. Since the same propensity score model can be used across all analyses, analyses that solely rely on correct specification of a propensity score model (see Sections 2 and 3 for specific proposals) do not suffer an increasing risk of model misspecification bias as more outcomes are being considered. In particular, their risk of bias due to model misspecification is the same as in the traditional epidemiologic design, in which one primary outcome is carefully studied. Further support for a propensity score analysis comes when drawing a parallel with outcome-wide randomised experiments; here, the propensity score is known by design, rendering an analysis that solely relies on correct specification of a propensity score (model) arguably the method of choice.