Invited Commentary

Invited Commentary: Beware the Test-Negative Design

Daniel Westreich* and Michael G. Hudgens

* Correspondence to Dr. Daniel Westreich, Department of Epidemiology, CB #7435, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (e-mail: djw@unc.edu).

Initially submitted February 26, 2016; accepted for publication April 8, 2016.

In this issue of the Journal, Sullivan et al. (Am J Epidemiol. 2016;184(5):345–353) carefully examine the theoretical justification for use of the test-negative design, a common observational study design, in assessing the effectiveness of influenza vaccination. Using modern causal inference methods (in particular, directed acyclic graphs), they describe different threats to the validity of inferences drawn about the effect of vaccination from test-negative design studies. These threats include confounding, selection bias, and measurement error in either the exposure or the outcome. While confounding and measurement error are common in observational studies, the potential for selection bias inherent in the test-negative design brings into question the validity of inferences drawn from such studies.

Confounding; epidemiologic methods; influenza vaccine; selection bias; test-negative study design

In this issue of the Journal, Sullivan et al. (1) provide a theoretical justification for the test-negative study design, which is routinely used to assess the effectiveness of the annual influenza vaccine. Additionally, the test-negative design has been used or proposed for use in several contexts beyond the influenza vaccine. For example, the US Food and Drug Administration recently considered use of the test-negative design for evaluating the effectiveness of Ebola vaccines that received accelerated approval without demonstration of clinical benefit (2).

Causal inference methods provide a formal framework for quantifying the (causal) effects of an exposure or treatment. Using causal methods, particularly directed acyclic graphs, Sullivan et al. carefully examine the theoretical justification for the test-negative design. They describe several threats to the validity of inferences drawn from studies which employ the test-negative design. These threats include confounding, selection bias inherent in the test-negative design, and measurement error in either the exposure or the outcome. Perhaps the first two of these threats, confounding and selection bias, are the most central issues regarding the validity of the test-negative design.

SELECTION BIAS

The key premise of the test-negative design, as Sullivan et al. note, is to remove “bias due to confounding by health-care-seeking behavior” (1, p. 348). Individuals with a propensity to seek care when ill may be more likely to receive the annual influenza vaccine (V) and may be more likely to exhibit behavior that reduces the risk of influenza infection (I). Such health-care-seeking (HS) behavior could confound the association between vaccination and infection. Thus, a cohort study which fails to account for health-care-seeking behavior could lead to biased estimates of the effect of the annual influenza vaccine. This is depicted by the path V ← HS → I in Figure 3A of Sullivan et al.’s article (1, p. 348). If health-care-seeking behavior were measured, then the usual confounder adjustment techniques could be used to yield valid vaccine effect estimates from a cohort study. However, the propensity of an individual to seek health care is difficult, if not impossible, to measure directly. Thus, health-care-seeking behavior will, in general, be an unmeasured confounder, and cohort studies may not yield valid estimates.

Investigators hope that by conditioning on patients presenting to health-care providers with influenza-like illness, the test-negative design will avoid (confounder) bias due to (unmeasured) health-care-seeking behavior. Unfortunately, as is demonstrated in Figure 4A of Sullivan et al. (1, p. 349), conditioning on testing for influenza-like illness (T = 1) will not in general eliminate this bias. Because T is a collider on path V ← HS → T ← I, conditioning on T = 1 fails to eliminate the noncausal association between vaccination and influenza infection. That is, despite conventional
wisdom, there is no formal justification for the claim that the test-negative design eliminates bias due to health-care-seeking behavior. Were HS measured, then by conditioning on HS one could block the path V ← HS → T ← I and the test-negative design could give rise to valid inferences about the effect of vaccine under certain assumptions. But if HS were measured, then confounding due to HS could also be controlled for in a cohort study. Moreover, it is plausible there are other unmeasured variables, say U, which influence V and T such that there may be other noncausal pathways of the form V ← U → T ← I. In this case, the test-negative design could yield invalid inferences due to collider-bias, whereas a cohort study (which does not restrict to T=1) would be free from such bias.

LACK OF GENERALIZABILITY

Sullivan et al. and others (e.g., Jackson and Nelson (3)) have pointed out that the test-negative design is not a case-control study design, in that there is no formal sampling of controls in order to estimate the distribution of exposure in the population that gave rise to the cases. Nonetheless, the test-negative design resembles a case-control study in that individuals are informatively selected based on factors related to their outcome status. This raises issues of generalizability (external validity). Because data are collected only on persons who experience symptoms and seek care for influenza-like illness (T = 1), the extent to which the results from the study generalize to the general population (where T = 0 or T = 1) is not clear. Jackson and Nelson state that because “the test-negative study includes only persons who have sought care for an ARI [acute respiratory illness], the study population is restricted to persons who would seek care if they developed an ARI” (3, p. 2166). If the impact of the vaccine differs by health-care-seeking behavior or associated factors, then the results of a test-negative study may not be generalizable to persons with T = 0.

UNCOLLAPSIBLE ODDS RATIO

Vaccine effectiveness (VE) is typically estimated in a test-negative study using a logistic regression model adjusting for potential confounders. Sullivan et al. indicate that under certain assumptions this will provide a consistent estimate of the causal odds ratio (1). However, the odds ratio is not collapsible, such that the conditional causal odds ratio will not in general equal the marginal causal odds ratio (4). This complicates interpretation of VE estimates across different studies, since the conditional causal odds ratio will generally depend on which confounders are being conditioned upon. Furthermore, under certain conditions, the conditional causal odds ratio will necessarily be farther from the null value of 1; thus, logistic regression-adjusted effect estimates of an exposure or treatment will be biased for the marginal causal odds ratio. For this reason, in the context of prospective cohort studies, inverse probability-weighted estimators and other alternative estimators have been recommended instead of logistic regression-adjusted estimators for drawing inference about the marginal causal odds ratio (5–8).

To demonstrate the bias of logistic regression-adjusted estimates of VE in a test-negative design due to collapsibility, suppose the directed acyclic graph in Sullivan et al.’s Figure 4A holds with the exception that, for simplicity, the arrow from C to HS is removed. Suppose C equals 1 with probability 0.9 and equals −1 otherwise and that HS is a uniform random variable on [0, 1], and assume that HS is measured. Suppose V (given C and HS) is Bernoulli-distributed with mean 1/(1 + exp(3C + HS)). Suppose the probability an individual becomes infected (given C and HS) is 1/(1 + exp(3C + HS)) if not vaccinated and 1/(1 + exp(3C + HS + 1)) if vaccinated, such that the true marginal causal odds ratio equals approximately 0.74. Suppose the probability that an individual gets tested (T = 1), given HS and I, equals 1/(1 + exp(1 + HS + T)). We simulated 2,000 cohorts, each containing 10,000 individuals, according to this model. On average, approximately 11% of individuals became infected and 1,700 persons in the cohort were included in the test-negative design (i.e., T = 1). Fitting a logistic regression model adjusted for C and HS to the subset of individuals with T = 1, the average estimated odds ratio equaled 0.39. In other words, the true VE equaled (1 − odds ratio) × 100% = 26%, yet the test-negative design on average estimated VE to be 61%. Even if data were observed on the whole cohort, the situation was not improved: The average logistic regression model-adjusted estimate of VE was 63%. The conditional odds ratio estimates from both the test-negative design and the cohort study were substantially biased for the marginal odds ratio even in the absence of unmeasured confounders or collider-bias due to noncollapsibility of the odds ratio.

CONCLUSIONS

Sullivan et al. (1) have made an important contribution to epidemiology by providing a theoretical basis for assessing the validity of the test-negative study design. Moreover, their results indicate that caution should be exercised in relying upon causal inferences drawn from test-negative designs without strong justification. In scenarios where such justification is plausible, questions about generalizability remain, and estimators alternative to those currently employed in the analysis of test-negative studies may be preferred.

A surprising aspect of this contribution by Sullivan et al. is that a theoretical justification (or lack thereof) for use of the test-negative design for drawing inference about the effect of an exposure or treatment had not (to our knowledge) been provided previously, despite the test-negative design’s being routinely used by researchers globally to make inferences about the effectiveness of influenza vaccination. In the future, should a new epidemiologic study design be proposed, causal inference methods should be used at the onset to assess whether the proposed new design can provide valid inference about the effect of an exposure or treatment and, if so, under what assumptions.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina.

Carolina at Chapel Hill, Chapel Hill, North Carolina (Daniel Westreich); and Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Michael G. Hudgens).

Funding was provided by grants DP2 HD084070 and R01 AI085073 from the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES


