On the Use of Imperfect Negative Control Exposures in Epidemiologic Studies

Marc G. Weisskopf,[a,b] *Eric J. Tchetgen Tchetgen,*[b,c] *and Raanan Raz*[a,d]

The use of negative controls to detect suspected and unsuspected threats to causal inference from confounding in observational epidemiology studies was recently described by Lipsitch et al.[1] In the context of negative control exposures, the basic principle relies on a variable (B) that is "U-comparable" to the exposure (A)–outcome (Y) relationship of interest (Fig. A). By U-comparable, it is meant that the unobserved common causes of B and Y are the same as the unobserved common causes of A and Y. (Thus, on a directed acyclic graph [DAG; Fig. A] have the same incoming arrows from uncontrolled variables. Note that B–Y does not need to share common causes of A–Y that are controlled [L in the DAG], but in practice they often do.) An additional important condition is that B does not cause the outcome (Y) (and that Y does not cause B). In this ideal case, the presence of an association between B and Y when A is included in a regression model of Y on A and measured covariates implies residual confounding, while a null B–Y association implies no empirical evidence of residual confounding.

In an analysis of the association between daily ozone levels and daily emergency room visits for asthma, it was shown how ozone levels 1 day after (day +1) that on which asthma visits are assessed (day 0) can serve as an ideal negative control exposure.[2] Because ozone levels on day +1 cannot have caused an asthma attack on day 0, in a model with both ozone on the day before (day −1) and day +1, any association between ozone on day +1 and asthma on day 0 would indicate residual confounding of the association between ozone on day −1 and asthma. This approach of using future air pollution levels as a negative control exposure was also extended to spatial analyses by the same authors.[3] Exposures that occur after the outcome event in question, assuming U-comparability, can be ideal negative control exposures because they cannot cause the outcome. Therefore, any association found with the outcome (in a model that includes both A and B of Fig. A) implies residual confounding.

But what if one does not know whether a variable (B′) is causally associated with the outcome Y (Fig. B)? Can such a variable still be used as a negative control exposure? This scenario is more problematic because if a B′–Y association is found in a model that includes A, then one cannot know whether the B′–Y association results from residual confounding or a true causal relation between B′ and Y. However, this does not necessarily mean that exposures that could potentially cause the outcome are never useful as negative controls. If such a variable (B′ in Fig. B) is still U-comparable with the exposure of interest (A), then a finding of no B′–Y association (in a model with A) would imply *both* no causal association between B′ and Y *and* no residual confounding of the A–Y association

From the [a]Department of Environmental Health, [b]Department of Epidemiology, [c]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; and [d]Faculty of Medicine, Braun School of Public Health & Community Medicine, Hebrew University-Hadassah, Jerusalem, Israel.

Some of the work on which this manuscript was based was funded by a grant to MW from the US Department of Defense (W81XWH-08-1-0499). Additional funding was received from Mr. James Crystal in honor of Mrs. Lillian Yaros.

The authors report no conflicts of interest.

Correspondence: Marc G. Weisskopf, Departments of Environmental Health and Epidemiology, Harvard T.H. Chan School of Public Health, Landmark Center, 401 Park Dr., P.O. Box 15677, Boston, MA 02215. E-mail: mweissko@hsph.harvard.edu.

ISSN: 1044-3983/16/2703-0365
DOI: 10.1097/EDE.0000000000000454

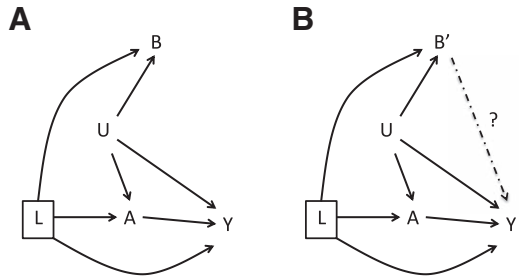

FIGURE. A, Causal diagram showing an ideal negative control exposure B for use in evaluating studies of the causal relationship between exposure A and outcome Y. B should ideally have the same incoming arrows as A. To the extent this criterion is met, B is called U-comparable with A. In theory, an ideal negative control exposure B does not need to share incoming arrows with controlled variables (represented as L with the box around it), but in practice an ideal B usually will. B, Causal diagram showing an imperfect negative control exposure B′, for which the association with the outcome Y is uncertain (represented as the *dashed arrow* between B′ and Y).

(under the faithfulness assumption, discussed further below). While it is the case that a finding of a null B′–Y association is an estimate and therefore subject to some uncertainty, there is similar uncertainty about a null association between an ideal negative control exposure B and Y (Fig. A) and the corresponding implication that there is no residual confounding of A–Y.

An additional important consideration is measurement error. If the imperfect negative control exposure (B′) is not measured as accurately as the exposure of interest (A), then B′–Y could look null and A–Y not, despite similar causal effects on Y. However, this would presumably be less of an issue if B′ is measured at least as accurately as A, in which case bias to the null from (nondifferential) measurement error by case status, which is a *sine qua non*, could not result in a larger effect estimate for A than B′. However, this is not an extra assumption of imperfect negative controls as the same concern holds for ideal negative control exposures (B). If B is not measured as accurately as A, then a null B–Y association may result from measurement error and not truly imply no residual confounding of A–Y.

One additional assumption beyond those described generally for negative control exposures is needed when one is uncertain about the causal relation between the potential negative control exposure variable and the outcome. One must assume that there is not residual confounding bias between B′ and Y that may be of equal magnitude, but opposite direction to the true causal association between the two. If that were the case, the estimated effect of B′ would appear null despite a true B′–Y causal association. Such pathological situations can be ruled out on the basis of faithfulness arguments which will often be reasonable in health-related applications.[4–6] Indeed, a negative control exposure may be most useful when one has a strong prior as to the direction of possible confounding bias: if the direction of confounding bias were away from the null causal association, then a null B′–Y association would imply no residual confounding of A–Y.

Examples of the use of imperfect negative control exposures such as B′ exist in epidemiology. In the analysis of the association between maternal smoking during pregnancy and offspring birth weight, paternal smoking has been used as a negative control exposure.[7,8] Paternal smoking during pregnancy could in theory also have an effect on offspring birth weight. When modeled alone, paternal smoking was associated with birth weight, but this would be expected even absent a causal association if maternal smoking is associated with low birth weight because of the association between maternal and paternal smoking. However, in mutually adjusted models, the association between paternal smoking and birth weight was null while the association between maternal smoking and birth weight (lower birth weight among children of mothers who smoke) persisted.[7,8] The null association with paternal smoking could occur if residual confounding of the paternal smoking–birth weight association ($B' \leftarrow U \rightarrow Y$) is equal and opposite to a true causal association ($B' \rightarrow Y$). However, we believe such a pathological scenario to be rare as dictated by faithfulness, and our prior knowledge suggests that this is the case. Such a scenario would either mean that paternal smoking caused greater birth weight ($B' \rightarrow Y$) or that the direction of the $U \rightarrow$ paternal smoking (B′) effect was opposite to that of $U \rightarrow$ maternal smoking (A). Thus, the interpretation that there truly is no causal association between paternal smoking (B′) and Y seems more probable, which would imply no residual confounding of the maternal smoking–birth weight association (assuming paternal smoking is U-comparable with the maternal smoking-birth weight association).

A notable caveat is that if B′ (paternal smoking) causes A (maternal smoking)—a plausible situation that is not depicted in the DAG of Figure B (it would mean an arrow from B′ to A)—then the path $B' \rightarrow A \leftarrow U \rightarrow Y$ would be open in a model that conditions on A. If the magnitude of this association is equal and opposite to the direct effect of $B' \rightarrow Y$ then a null empirical association between B′ and Y would not necessarily reflect a lack of residual confounding of $A \rightarrow Y$. However, once again we believe such a pathological scenario to be rare as dictated by faithfulness. In general, this concern is one that could also apply to ideal negative control exposures if the ideal negative control exposure B causes A. If that is true (which it clearly would *not* be for ozone on a day after the day of interest as in the example given above[2]), then it is possible that the path $B \rightarrow A \leftarrow U \rightarrow Y$ could exactly balance the path $B \leftarrow U \rightarrow Y$, which would mean a null $B \rightarrow Y$ association would not imply no residual confounding of $A \rightarrow Y$.

Additional examples of negative control exposures include two studies that have examined the association between maternal exposure to particulate matter (PM) during different trimesters of pregnancy and autism spectrum disorder (ASD) in her child in models that included more than one trimester's

exposure.[9,10] In both cases, prior to conducting the study, it was unclear whether exposure during one trimester would be causally related to ASD while exposure during the other trimesters would not. In both cases, an association was observed with exposure during the third trimester (higher odds of ASD among mothers with higher PM exposure), but null associations with exposure during other trimesters were observed when exposure for the different trimesters were modeled simultaneously. Similar to the example above, it is possible that a true causal association exists between, for example, first trimester PM exposure and ASD, but that residual confounding perfectly negates that effect to produce an overall null association. But again as in the example above, the faithfulness argument would rule this out and, furthermore, for confounding to balance out a true causal effect of first trimester PM would imply either that higher first trimester PM exposure reduces the odds of ASD or that the residual confounding has opposite effects on first and third trimester PM, both unlikely scenarios. Therefore, the most plausible explanation of the null first trimester PM–ASD association is that there is in fact no causal association between them, with, thus, the implication that there is no residual confounding of the association seen with third trimester PM (assuming first trimester PM is U-comparable with the third trimester PM–ASD association).

PM levels on one day could include actual PM from the day before, which would mean a B′ → A arrow (B′ being PM on the earlier day, A being PM on the next day) as discussed for paternal smoking above. However, as time between exposure periods increases this gets weaker and weaker, and after a 3-month interval would be completely gone.[11] Thus, expert knowledge on characteristics of the exposures allow us to rule out the concern that conditioning on A could introduce any bias to mask a true causal effect of B′ on Y. In general with air pollutants, however, this concern can often be completely ruled out. If the imperfect negative control is an exposure level in a time period after the exposure window where an association is seen then the imperfect negative control exposure clearly could not cause the earlier exposure no matter what the time interval between the exposure windows considered, as in the example above of ideal negative control exposures.[2] In one of the prior papers on autism, PM during the pregnancy, 9 months before, *and* the 9 months after pregnancy were modeled together and the associations with exposure during both the period before and after were found to be null while exposure during pregnancy was associated with elevated odds of autism.[10] In this case, exposure during the 9 months after pregnancy clearly could not affect exposure during the pregnancy and so the concern is entirely eliminated.

Greater use of negative controls in epidemiological studies could bring great strength to many studies. While negative control exposures that absolutely could not cause the outcome are ideal, this is not an absolute requirement. Potential negative control exposure variables with a more uncertain causal influence on the outcome should still be considered. Their utility will depend more heavily on the results obtained, but in some scenarios, they can also be a powerful tool for causal inference.

ABOUT THE AUTHORS

MARC G. WEISSKOPF is an Associate Professor of Environmental and Occupational Epidemiology at the Harvard T.H. Chan School of Public Health. His research focuses on environmental risk factors for neurological disorders and diseases, with an interest in epidemiologic methods issues of particular relevance for environmental health studies. ERIC J. TCHETGEN TCHETGEN is a Professor of Biostatistics and Epidemiologic Methods at the Harvard T.H. Chan School of Public Health with an interest in methodology to minimize the possibility of bias in the design and analysis of observational studies and imperfect experiments. RAANAN RAZ is an assistant professor with the Braun School of Public Health and Community Medicine at the Hebrew University of Jerusalem–Hadassah. He holds a PhD in neuroscience from Bar Ilan University and an MSc in epidemiology from Tel Aviv University. His research is focused on environmental factors in brain development.

REFERENCES

1. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21:383–388.
2. Flanders WD, Klein M, Darrow LA, et al. A method for detection of residual confounding in time-series and other observational studies. *Epidemiology*. 2011;22:59–67.
3. Flanders WD, Klein M, Darrow LA, et al. A method to detect residual confounding in spatial and other observational studies. *Epidemiology*. 2011;22:823–826.
4. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press; 2001.
5. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press; 2000.
6. Hernan M, Robins J. *Causal Inference*. Abingdon, UK: Chapman & Hall/CRC; 2015.
7. Smith GD. Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings? *Basic Clin Pharmacol Toxicol*. 2008;102:245–256.
8. Smith GD. Negative control exposures in epidemiologic studies. *Epidemiology*. 2012;23:350–351; author reply 351.
9. Kalkbrenner AE, Windham GC, Serre ML, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology*. 2015;26:30–42.
10. Raz R, Roberts AL, Lyall K, et al. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: a nested case-control analysis within the Nurses' Health Study II Cohort. *Environ Health Perspect*. 2015;123:264–270.
11. Seinfeld JH, Pandis SN. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, NJ: John Wiley & Sons, Inc; 2006.