



Invited Commentary

Invited Commentary: Decomposing with a Lot of Supposing

Jay S. Kaufman*

* Correspondence to Dr. Jay S. Kaufman, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec, Canada H3A 1A2 (e-mail: jay.kaufman@mcgill.ca).

Initially submitted July 28, 2010; accepted for publication July 30, 2010.

In this issue of the *Journal*, VanderWeele and Vansteelandt (*Am J Epidemiol.* 2010;172(12):1339–1348) provide simple formulae for estimation of direct and indirect effects using standard logistic regression when the exposure and outcome are binary, the mediator is continuous, and the odds ratio is the chosen effect measure. They also provide concisely stated lists of assumptions necessary for estimation of these effects, including various conditional independencies and homogeneity of exposure and mediator effects over covariate strata. They further suggest that this will allow effect decomposition in case-control studies if the sampling fractions and population outcome prevalence are known with certainty. In this invited commentary, the author argues that, in a well-designed case-control study in which the sampling fraction is known, it should not be necessary to rely on the odds ratio. The odds ratio has well-known deficiencies as a causal parameter, and its use severely complicates evaluation of confounding and effect homogeneity. Although VanderWeele and Vansteelandt propose that a rare disease assumption is not necessary for estimation of controlled direct effects using their approach, collapsibility concerns suggest otherwise when the goal is causal inference rather than merely measuring association. Moreover, their clear statement of assumptions necessary for the estimation of natural/pure effects suggests that these quantities will rarely be viable estimands in observational epidemiology.

causal inference; conditional independence; confounding; decomposition; estimation; interaction; logistic regression; odds ratio

Abbreviation: DAG, directed acyclic graph.

The story of effect decomposition analysis in epidemiology is one of the more curious in our field's methodological history. It is exceedingly common to encounter statements about mediation in journal articles throughout epidemiology, public health, and biomedicine, and yet almost nothing is said about these techniques in our methods textbooks, nor (until recently) has there been much guidance from the biostatistics literature. The few brief mentions of this issue in epidemiology textbooks generally recommend the traditional approach described and critiqued in the current paper by VanderWeele and Vansteelandt (1): Estimate an exposure effect with adjustment for confounders and compare this with the same parameter estimated in a second model that also controls for a putative intermediate. Express the difference between these 2 estimates as a proportion of the first estimate, and

the resulting number is the proportion of the effect of exposure that is mediated by the additional variable. A common interpretation in the applied literature is that blocking the selected intermediate from occurring would prevent this fraction of the exposure effect. For example, if the confounder-adjusted exposure effect on the ratio scale is 2.0, and the addition to the model of a putative intermediate changes the exposure effect estimate to 1.8, then 20% of the effect is relayed through this intermediate, and 20% of the cases attributable to exposure could be prevented if this pathway were interrupted by a public health intervention on the intermediate (2, p. 160). As described with persuasive clarity by these authors, this analytical approach is rarely defensible and very prone to provide substantially misleading inference. It is high time for us, as a discipline, to do better.

Until the mid-1980s, there was very little attention in the epidemiologic literature to the issue of adjusting for factors affected by exposure. A few authors, citing the sociologic tradition, recommended the simple and poorly justified approach described above. In a series of articles in 1986 and 1987, however, Robins (3) laid out a novel and comprehensive approach to longitudinal data in epidemiologic studies, including consideration of covariates affected by a dose of exposure at time 1 that might then confound or modify the next dose of exposure at time 2. This was the first formal foundation articulated for effect decomposition in epidemiology, a theme that was further clarified in a tutorial article by Robins and Greenland in 1992 (4). The 1992 article began with a devastating counterexample to the simple strategy described above, and yet the standard practice in epidemiology journals changed little. A second revolution followed in the mid-1990s when Pearl (5) introduced his directed acyclic graphs (DAGs) into biostatistics, which led to their rapid incorporation into epidemiologic methods by the end of the decade (6). The use of DAGs allowed for a clearer representation of an analyst's beliefs about the structure of variables in relation to one another and, therefore, to the implications of adjusting for factors affected by exposure and the existence of direct and indirect effect pathways. DAGs also helped to clarify and popularize the earlier insights of Robins and Greenland's 1992 tutorial (7), and since 2005 or 2006 there has been a veritable explosion of methodological developments surrounding estimation of direct and indirect effects in epidemiology and biostatistics. The majority of the papers published on this topic have appeared in just the last few years, including many important contributions by both VanderWeele and Vansteelandt.

As noted by these authors, almost all of the formal work on direct and indirect effects in epidemiology to date has focused on the difference scale. The contribution of the current paper is to extend many of these novel developments and insights to the more commonly used effect measure, the odds ratio. The authors are clearly in command of deep understanding of all of the relevant causal and statistical issues involved here, and yet this goal—extending recent causal insights to the odds ratio—strikes me as akin to retrofitting a modern jet fighter to run on coal. The authors suggest that the odds ratio might be necessary because it is the parameter available from case-control studies, but this is a viewpoint that became outdated over 30 years ago (8). Indeed, it ought no longer to be necessary to estimate or publish an odds ratio from a well-designed epidemiologic study (9, 10).

VanderWeele and Vansteelandt (1) provide a carefully formulated causal definition of effect decomposition in epidemiologic studies, but they apply this to an effect measure that has no interpretation as an average individual causal effect unless it approximates the risk or rate ratio by virtue of study design or a rare disease assumption (11). The authors note that the rare disease assumption will be necessary if there are interactions between confounders C and either exposure A or mediator M , but they state that no such assumption is necessary for controlled direct effects adjusted

for confounders C in the absence of these interactions. I would argue, however, that the rare disease assumption will always be necessary in the presence of covariates C , because of an inherent limitation of the odds ratio. If the odds ratio does not approximate the relative risk, then the addition of another covariate C could change the effect estimate, even if this additional variable is not a confounder or a surrogate for a confounder (12). Moreover, when the odds ratio does not approximate the relative risk, the causal estimate can differ depending on the method of adjustment for C . For example, standard outcome regression could provide a different effect measure than inverse probability of treatment weighting (13). Finally, since the adjusted odds ratio is not any kind of summary of the average individual causal effects, decomposition analysis cannot be a reliable tool for either inferring mechanism or predicting the effects of population interventions when using this measure in a traditional case-control design with a common outcome (11).

The authors are meticulous in laying out, in admirably simple notation, the assumptions required for the estimation of controlled direct and pure/natural direct and indirect effects. Although they recommend greater attention to these assumptions by practitioners, they do not comment on the plausible attainment of these assumptions in a real analysis. I believe, however, that some honest appraisal of these necessary conditions would render the pure/natural species of effects largely moot. Like valid instrumental variables, the Higgs boson, and Santa Claus, these estimators may be best viewed as belonging to the class of things that are really nifty ideas but which unfortunately have not yet been observed in the real world. As an illustration of the difficult estimation of pure/natural effects, consider the authors' applied example involving exposure to dampness and mold, mediated by perception of control, with respect to the outcome of depression. Valid estimation of pure/natural direct and indirect effects requires that there be no unmeasured common causes of exposure and mediator, such as poverty and residential instability. Moreover, it requires that no consequence of exposure, measured or unmeasured, confound the association between sense of control and depression. But if mold causes any physical illness, such as mycosis or allergic reaction, this would likely affect both sense of control and risk of depression (i.e., variable L in VanderWeele and Vansteelandt's Figure 2). If these health effects were measured, one could estimate controlled direct effects but not pure/natural direct and indirect effects (14). Moreover, estimation by inverse probability of treatment weighting would yield the marginal odds ratio with respect to L , rather than the conditional odds ratio produced by standard logistic regression analysis. Once again, the rare disease assumption would be necessary to rescue the unambiguous causal interpretation of this odds ratio.

VanderWeele and Vansteelandt (1) caution analysts to check logistic regression models carefully for heterogeneity of exposure odds ratios across strata of intermediate variable M before reporting direct and indirect effects. This is surely wise advice, but it must be noted that there are numerous improprieties that will generate apparent heterogeneity across strata of M , including sampling variability,

differential confounding effects across strata (2, p. 213), and differential or nondifferential measurement error (15). This raises some doubts about what it means to specify the logistic model correctly, especially given the collapsibility concerns described above. For example, the authors caution that omission of the θ_{3am} term in model 7 could result in an exposure coefficient θ_1 that is close to null because of averaging of effects in opposite directions across strata of M . On the other hand, the stratum-specific odds ratios will generally be further from the null than the average odds ratio, even if there is no mediation at all because A has no effect on M .

The 2 main pitfalls in decomposition analysis that are identified by VanderWeele and Vansteelandt (1) are confounding and effect measure modification. Detecting and addressing these problems in epidemiology are sufficiently difficult tasks under any circumstances but become substantially more difficult when the odds ratio is the chosen effect measure. A tremendous strength of the recent literature on effect decomposition, including the current article, is the use of an explicitly causal framework for understanding and interpreting direct and indirect effects, and for clearly defining the conditions that allow these effects to be estimated. When the disease outcome is common, however, the odds ratio is a particularly inopportune choice of measure for this framework. In that situation, it is the use of the odds ratio that should be rare.

ACKNOWLEDGMENTS

The author is supported by funding from the Canada Research Chairs program.

Conflict of interest: none declared.

REFERENCES

1. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010; 172(12):1339–1348.
2. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 2nd ed. Boston, MA: Jones and Bartlett Publishers; 2007.
3. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.
4. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3(2):143–155.
5. Pearl J. Causal diagrams for experimental research. *Biometrika*. 1995;82:669–710.
6. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48.
7. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol*. 2002;31(1):163–165.
8. Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol*. 1976;103(2):226–235.
9. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol*. 2005; 162(3):199–200.
10. Langholz B. Case-control studies = odds ratios: blame the retrospective model. *Epidemiology*. 2010;21(1):10–12.
11. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*. 1987;125(5): 761–768.
12. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189–212.
13. Kaufman JS. Marginalia: comparing adjusted effect measures. *Epidemiology*. 2010;21(4):490–493.
14. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20(1): 18–26.
15. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med*. 1998;55(10):651–656.