Published in final edited form as:

Epidemiology. 2012 January ; 23(1): 10–12. doi:10.1097/EDE.0b013e31823b5492.

The Wizard of Odds

Richard F. MacLehose^{1,2} and Jay S. Kaufman³

¹Division of Biostatistics, University of Minnesota

²Division of Epidemiology, University of Minnesota

³Department of Epidemiology, Biostatistics and Occupational Health, McGill University

In the 1939 Metro-Goldwyn-Mayer film, *The Wizard of Oz*, the benevolent Wizard is asked to solve the existential dilemmas of the Scarecrow, Tin Man and Cowardly Lion. Seemingly unable to provide them the intangible things that they seek, he instead gives them three mundane items that he happens to have in his bag. The munificence of the Wizard is not cast into doubt by this sleight of hand, however, because the characters discover that the traits they sought were things they had possessed all along. The article by VanderWeele and colleagues¹ responds to a crisis in reproductive epidemiology over how to handle measures of maturity and development. Variables such as gestational age and birth weight are ubiquitous in perinatal research, but problematic for all the reasons described in their paper. Analysts want to condition on some measure of maturation to put neonates on an equal developmental footing, but exposures of interest may affect the progress of the pregnancy, making conditional estimates biased as measures of total effect. At the same time, direct effects are generally not identifiable because of unmeasured common causes of the maturity indicator and the outcome, leading to the birth-weight paradox described by the authors. Reproductive epidemiologists therefore look eagerly to methods wizards for a magical solution to their problem: for the brains, heart and courage necessary for their daily travails. What have the methodologists got in their bag?

The first method proposed by VanderWeele et al.¹ is to estimate the total effect of an exposure such as smoking on neonatal mortality within strata defined by the estimated risk of a developmental intermediate such as low birthweight. The authors predict the risk of low birthweight through a regression model, and create categories of high and low risk. This is a straightforward extension of the typical practice of estimating total effects within strata of baseline covariates. In this case, the strata are defined as some more complicated function of the covariates, but the approaches are similar; they estimate the total effect of smoking among sub-categories of the population. The ability of this method to resolve the birth-weight paradox depends, however, on the discriminatory power of the risk score and the prevalence of the intermediate. The authors estimate the effect of smoking on neonatal mortality as $OR_{high\ risk}=1.6$ and $OR_{lowrisk}=1.3$, among those at high and low risk. Rather than implying a resolution of the paradox, however, these estimates are entirely compatible with a state of nature in which smoking actually has a protective effect among low-birth-weight infants.

Imagine we know $p_{am} = \Pr(Death|SET[Smoke = a], SET[LBW = m])$, where $smoke=1$ indicates maternal smoking and $LBW=1$ indicates infant low birthweight. If we posit values of p_{am} we can ascertain how well the authors' stratification appears to resolve the paradox. Suppose, for example, that the birth-weight paradox is not actually a paradox, and that

*Address for Correspondence: Richard F. MacLehose, Division of Epidemiology and Community Health, University of Minnesota, 1300 S 2nd St, Suite 300, Minneapolis, MN 55454, Phone: (612) 624-1932, Fax: (612) 624-0315, mac10029@umn.edu.

smoking is actually beneficial in one stratum and harmful in the other. If we take as parameters $p_{11}=0.04$, $p_{01}=0.06$, $p_{10}=0.05$, and $p_{00}=0.03$, then in the absence of confounding, the true controlled direct effect of smoking among normal-weight infants would be $OR_{(m=0)} = 1.7$ while among low-birth-weight infants it would be $OR_{(m=1)} = 0.7$. Suppose we apply the stratification proposed by the authors¹ and our classification scheme has some associated set of positive and negative predictive values ($ppv_{(a=1)}$, $ppv_{(a=0)}$, $npv_{(a=1)}$ and $npv_{(a=0)}$). The stratified total effects obtained by the proposed method are given by the formulae:

$$OR_{high\ risk} = \frac{[p_{11} * ppv_{(a=1)} + p_{10}(1 - ppv_{(a=1)})] / [1 - \{p_{11} * ppv_{(a=1)} + p_{10}(1 - ppv_{(a=1)})\}]}{[p_{01} * ppv_{(a=0)} + p_{00}(1 - ppv_{(a=0)})] / [1 - \{p_{01} * ppv_{(a=0)} + p_{00}(1 - ppv_{(a=0)})\}]} \\ OR_{low\ risk} = \frac{[p_{10} * npv_{(a=1)} + p_{11}(1 - npv_{(a=1)})] / [1 - \{p_{10} * npv_{(a=1)} + p_{11}(1 - npv_{(a=1)})\}]}{[p_{00} * npv_{(a=0)} + p_{01}(1 - npv_{(a=0)})] / [1 - \{p_{00} * npv_{(a=0)} + p_{01}(1 - npv_{(a=0)})\}]} \quad (1)$$

For example, if the risk classification model has $ppv_{(x=1)} = ppv_{(x=0)} = 0.9$ and $npv_{(x=1)} = npv_{(x=0)} = 0.9$, then equation (1) yields $OR_{low\ risk} = 1.5$ and $OR_{high\ risk} = 0.7$, which are similar to the controlled direct effects. However, an intermediate with a low prevalence such as low birthweight will have poor ppv unless the specificity is nearly 1.0. If the risk classification model has $ppv_{(x=1)} = ppv_{(x=0)} = 0.3$ and $npv_{(x=1)} = npv_{(x=0)} = 0.9$ then equation (1) yields $OR_{low\ risk} = 1.5$ and $OR_{high\ risk} = 1.2$. A practitioner could be tempted to believe these results imply that the qualitative effect modification referred to as a “paradox” does not really exist when, in fact, it does. In the absence of extremely good predictive models for low birthweight, it is therefore hard to have much faith in the ability of this method to detect a “paradox” were it actually to exist. The authors’ finding of harmful effects of smoking in both risk strata cannot be taken as evidence of having “resolved” the paradox; in the absence of an excellent prediction tool for the intermediate, it is the answer one would expect regardless of the truth.

Jumping to the third method, the authors¹ propose estimation of the principal-stratum direct effect. Resolution of the problem by means of restriction to a stratum in which exposure has no effect on the intermediate predates the cited article of Frangakis and Rubin² --for example, in the work of Joffe et al.^{3,4} But Joffe and colleagues identified this stratification on substantive grounds, whereas the authors’ approach here treats it as a latent factor that must be estimated with a model. The proposed sensitivity analysis parameter is an honest way of dealing with the inherent uncertainty of this estimation, but one that will often be difficult in practice. There is little substantive guidance available about the likely magnitude of this parameter, only its sign. Even if one happens to get this value right in the analysis, however, the effect estimate must be applied in public-health terms to a sub-population whose membership and prevalence are unknown.⁵ The authors are quite frank about all of these dilemmas, which leaves them not entirely sanguine about the prospects for this approach.

This leaves approach 2, which has been applied elegantly in previous papers,⁶ and which we agree will be of greatest value. This approach follows a long line of methodological development for effect estimation in the presence of unmeasured covariates.⁷ By specifying the prevalence of an unmeasured confounder within strata of the intermediate, as well as the magnitude of its effect on the outcome, the authors¹ can correct the observed direct-effect estimates. As they note, one need not specify the parameters perfectly; it can be illuminating merely to estimate the extent of bias under alternative guesses at the bias parameters. But let’s go one step further by rephrasing the question: what values of the parameters given in the author’s expression (3) could resolve the birthweight paradox? The resolution of the paradox can be viewed in different ways; however, because the greatest concern seems to be that smoking could appear to be protective in some stratum, one might wonder what

parameter values lead to bias equal to the observed OR, and therefore to an adjusted OR=1.0. To determine this, one can rearrange expression (3) given by the authors as:

$$\gamma = \frac{\pi_{1m} + B - B * \pi_{om} - 1}{\pi_{1m} - B * \pi_{om}} \quad (2)$$

where π_{am} is the prevalence of the unmeasured confounder among those with smoking status a in low-birth-weight stratum m , B is the amount of bias necessary to render the observed association null, and is the unmeasured relationship between confounder and disease. In this case, a value of $B=0.76$ would lead to a perfectly null bias-adjusted OR and a resolution of the paradox. Rather than examine a few token examples, we plot the values of γ necessary at all combinations of π_{om} and π_{1m} in Figure 1. Notably, for some prevalence combinations there is no value of γ that can explain the paradox according to this formula. Moreover, prevalence combinations that approach $\pi_{om} = B * \pi_{vm}$ require increasingly large confounder-outcome associations to explain the paradox, while prevalence combinations further from that line require less substantial effects. In this setting, with limited substantive knowledge, it may be more useful to examine a graph such as this rather than two or three discrete points.

Approach 2 involves estimation of controlled direct effects, which have an established place in the epidemiologic toolkit.⁸ The authors¹ use a sensitivity parameter for unmeasured confounding of the intermediate, which is a pragmatic and effective elaboration. This seems to be the superior strategy of the three considered in this paper, as the authors themselves note. Yet it is important to add that a crucial consistency assumption is not met for an effect premised on intervention when no such intervention is realistic.⁹ We presumably have the technology, if not the malice, to fix all births as low birthweight by induction of premature labor. We have no mechanism for controlling pregnancies to normal birthweight, and, more importantly, these effect estimates have no pretense of corresponding to any real-world interventions. The authors therefore caution that the estimates should be granted no causal interpretation. Unfortunately, epidemiologists face this dilemma exactly because we need causal estimates. A more satisfying solution to this problem awaits improved knowledge of the unmeasured common causes of low birthweight and neonatal mortality. And so, like Dorothy at the end of the film, epidemiologists may find that the magic needed to resolve their problem actually lies in the substantive and methodological knowledge in their own backyard.

Citations

1. VanderWeele TJ, Mumford S, Schisterman E. Conditioning on intermediates in perinatal epidemiology. *Epidemiology*. 2012; 23:xxx-xxx.
2. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21-29. [PubMed: 11890317]
3. Joffe MM, Colditz GA. Restriction as a method for reducing bias in the estimation of direct effects. *Stat Med*. 1998 Oct 15; 17(19):2233-49. [PubMed: 9802181]
4. Joffe MM, Byrne C, Colditz GA. Postmenopausal hormone use, screening, and breast cancer: characterization and control of a bias. *Epidemiology*. 2001 Jul; 12(4):429-38. [PubMed: 11416781]
5. Robins JM, Greenland S. Comment on Angrist, Imbens and Rubin: Estimation of the global average treatment effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 91:456-458.
6. VanderWeele TJ, Hernández-Díaz S. Is there a direct effect of pre-eclampsia on cerebral palsy not through preterm birth? *Paediatr Perinat Epidemiol*. 2011 Mar; 25(2):111-5. [PubMed: 21281323]

7. Greenland, S.; Lash, TL. Bias Analysis in Modern Epidemiology, third edition. Rothman, KJ.; Greenland, S.; Lash, TL., editors. Lippincott Williams and Wilkins; Philadelphia: 2008.
8. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology*. 2006; 17(3):276–84. [PubMed: 16617276]
9. Hernán MA. Invited commentary: hypothetical interventions to define causal effects--afterthought or prerequisite? *Am J Epidemiol*. 2005 Oct 1; 162(7):618–20. [PubMed: 16120710]

Figure 1. Magnitude of confounder-disease association necessary to result in an adjusted $OR_{m=1}=1.0$ ($B=0.76$) for values of π_{0m} and π_{1m} . Prevalence values contained between the vertical planes are those for which no possible value of γ can result in an adjusted $OR_{(m=1)}$ ($B=0.76$).