Social Science & Medicine 76 (2013) 8–11

Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Commentary

Some models just can't be fixed. A commentary on Mortensen

Jay S. Kaufman*

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Ave West, Montreal, QC H3A 1A2, Canada

ARTICLE INFO

Article history:
Available online 5 November 2012

Keywords:
Endogeneity
Fixed effects
Life course
Pregnancy
Regression
Social epidemiology
Socioeconomic indicators
Selection bias

Social epidemiology aims to estimate the causal effects of exposures such as education and income on disease outcomes in human populations (Oakes & Kaufman, 2006). These causal effects are desired as a basis for predicting the health impacts of social policies, such as educational loans or tax credits. Observational studies produce estimated associations between the exposures and outcomes, but it is widely appreciated that these associations may fail to estimate causal effects due to a variety of potential biases, such as confounding. While the standard approach to dealing with confounding in observational epidemiology has long been to measure and adjust for the important common causes of exposure and outcome, this approach is falling out of favor in social epidemiology (Harper & Strumpf, 2012). Increasingly, it is reasoned that the determinants of social variables such as education and income are simply so diffuse and difficult to measure accurately that no set of obtained covariates would be sufficient to estimate a plausibly casual estimate.

Various proposed solutions to this crisis have emerged, including randomized trials of social interventions (Kaufman, Kaufman, & Poole, 2003), the fortuitous occurrence of so-called "natural experiments" (Cerdá et al., 2012), and various designs that rely on discovering determinants of exposure that are conditionally independent of outcome (Glymour, 2006). The field of econometrics has no doubt served as a role model in this endeavor, having long ago given up on any pretense of defeating omitted variable bias through traditional covariate modeling (Angrist & Pischke, 2008). Several crucial influences are notable in this regard, including the importance of control groups, as in the so-called "difference in differences" model (Harper, Strumpf, & Kaufman, 2012), the leverage afforded by exogenous variation, as in instrumental variables designs (Glymour, 2006), and the preference for conditional over marginal estimates in hierarchical and panel settings in order to eliminate confounding at the upper level of aggregation. For example, if studying adverse birth outcomes in various neighborhoods, a marginal or random effects model attempts to control for confounding at the neighborhood-level via measured neighborhood-level covariates. This is suspect if such measures are limited, and so an attractive alternative is to compare exposed and unexposed women within the same neighborhood, a so-called "fixed effect" estimate (Schempf & Kaufman, 2012). Because this fixes the cluster (e.g., the neighborhood) in all comparisons of exposed versus unexposed units, it completely removes all possibility of confounding at this level.

The promise of controlling for all cluster-level confounding is a powerfully seductive one, but must be balanced in practice against other concerns, such as precision and selection bias (Kaufman, 2008). If many clusters have only exposed or unexposed units, for example, then much of the data may be discarded, which has implications for both internal and external validity. Moreover, as seen in the intriguing new paper by Mortensen (Mortensen, 2013), when covariates are logically forced to be collinear with the exposure within cluster (but not across clusters), any advantage

* Tel.: +1 514 398 7341; fax: +1 514 398 4503.
E-mail address: jay.kaufman@mcgill.ca.

0277-9536/$ – see front matter
http://dx.doi.org/10.1016/j.socscimed.2012.10.013

sought from the fixed effects design is precluded. In such circumstances, the subject matter dictates an alternate design, no matter how clever the ambitions of the analyst.

Mortensen is interested in the causal relations between socioeconomic position and adverse birth outcomes, a worthy topic given that gradients in risk exist almost universally and account for substantial fractions of infant morbidity and mortality even in rich countries (Kramer, Séguin, Lydon, & Goulet, 2000). The exposure in this case is difficult to characterize, and even more difficult to measure, but the author has at his disposal some of the best population cohort data in the world, the nearly 800,000 Danish births that occurred from 1997 to 2007. After exclusions for multiples, stillbirths, nearly 100,000 with foreign-born grandparents and another 150,000 births with no siblings or maternal cousins in the cohort, the author still has left at his disposal nearly half a million live singleton births with extensive and high-quality linked socioeconomic data. This is an enviable resource, and Mortensen focuses on education and income in relation to the outcomes of preterm birth (i.e., gestational age less than 37 completed weeks) and small-for-gestational age births (i.e., the lowest 10th percentile of births at each gestational age).

In these comments I focus on the outcome of preterm birth, in part for simplicity, and in part because small-for-gestational age is problematic due to an intrinsic conditioning on gestational age in the definition (VanderWeele, Mumford, & Schisterman, 2012). Moreover, I focus on the socioeconomic measure of completed years of maternal education, again for simplicity, and also because the income results were largely null in all models. Mortensen pursued three different analytic designs in order to estimate the causal effect of maternal education on occurrence of preterm birth, and this in itself is laudable. Designs have innate strengths and weaknesses for a particular causal question, and the opportunity to compare across analyses can make a paper considerably more informative. He referred to these designs as cohort, sibling-control and case-crossover analyses, and modeled the exposure effect with a proportional hazards regression, using gestational age as the underlying time scale and censoring at 37 weeks.

The cohort study simply involved the standard epidemiologic approach of considering exposed and unexposed women in the population, and attempting to isolate the causal effect by conditioning on all of the common causes of the exposure and outcome. Since educational attainment is a complex behavior and preterm birth is a largely mysterious outcome, one can't invest much confidence in the hope that all important confounding has been addressed. I'll focus on university graduates as the "exposed" group (since the extreme stratum of women with doctoral degrees is quite sparse). Compared to women with the minimum compulsory educational level, those with university education had on average 64% the rate of delivery at each gestational age through 37 weeks (HR = 0.64, 95% CI: 0.60–0.67). This is in fact a species of fixed effect model, since the regression model compares women only in the same risk-set (indexed by gestational age). Women who are censored by events or loss to follow-up are therefore selected out, but it appears that this proportion is negligible in this context, since the outcome is relatively rare and there is no drop-out (since the data are restricted to live births).

In the second design, the sibling-control design, the author fixes the comparison to involve female siblings (i.e., women matched on having the same mother), with further stratification for parity and regression control for maternal age and other variables. Thus the contrast can be thought of as essentially between a woman and her sister(s) of the same age, giving birth to their first child (or some subsequent matched parity). Clearly there are in fact few sisters of the same age, but maternal age was categorized here in 5-year intervals, and regression models are quite happy to extrapolate anyway. This analysis has the impressive characteristic that all shared traits of the mother's family are held fixed in the comparison, thus preventing any confounding at this level. For example, if there is geographic confounding such that some regions have high average educational attainment and low preterm birth risk, this negative association would appear to be causal protection of higher education in the cohort analysis, but will be controlled in the sibling-control design. Compared to sisters with the minimum compulsory educational level, those with university education had on average 90% the rate of delivery at each gestational age through 37 weeks (HR = 0.90, 95% CI: 0.78–1.04).

While the confounding of this estimate can be predicted to be much more modest than the confounding of the cohort estimate, this does not imply that it is closer to the true causal effect. Aside from residual confounding at the individual level, there are two other sources of error that contaminate this estimate to a greater extent than the cohort estimate: sampling variability and selection bias. Greater sampling variability arises because nearly 270,000 births (57%) are excluded from the analysis because the mother had no female sibling in the cohort. Another undisclosed quantity of observations contribute nothing to the likelihood because they are concordant in their exposure status (e.g., two sisters who have both attained a university education when they have their first child). The 95% CI for the regression coefficient in the sibling-control design is therefore about 2.5 times wider, and this implies that the variance is about 5 times greater. Since mean square error (MSE) is the sum of variance and squared bias, one would have to expect a big bias reduction to make this precision loss worthwhile (Kaufman, 2008).

Unfortunately, dropping more than two-thirds of the births may also incur a dramatic selection bias, since the observations are not omitted randomly (Hernán, Hernández-Díaz, & Robins, 2004). Consider the structure of the data in Fig. 1, with each maternal grandmother having 2 daughters, who each have 1 birth. GM is the educational attainment of the grandmother, M1 and M2 the educational attainments of the two sibling mothers, and B1 and B2 the outcomes of their two births. There is also an unmeasured personality trait for each mother (P1 and P2) that causes higher educational achievement and healthier pregnancy. Therefore the educational status of each mother is some function of the grandmother's education and her unmeasured personality trait. The observed association between maternal education and birth outcome is therefore confounded by P, and the absence of directed arrows from M to B represents the null hypothesis that maternal education has no causal effect on preterm birth. To the extent that P is imbalanced across sisters, this individual level confounding variable is present in both cohort and sibling-control analyses. Now what happens if we do as Mortensen has essentially done and remove all of the sibling pairs in which M1 = M2? If a mother achieved her higher education status through the effect of GM, then she is more likely to be concordant with her sister. If she did not achieve her higher education status through the effect of GM, however, then she is more likely to have achieved it through the effect of P. Among the subset of the population in which M1 ≠ M2,

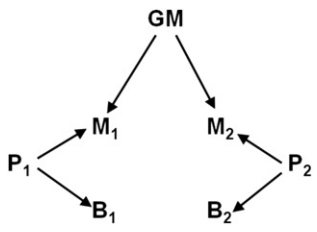

Fig. 1. Sibling-control study.