Reading Response

Susanna Makela

February 22, 2017

This is a longer response since it combines this week's and last week's responses.

The first part is for last week, though I'm not going to write about the readings. Instead, I'll write about a paper on estimation of neighborhood-level causal effects in epidemiology and a response paper. The first paper, by Oakes (2003), is titled "The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology" and published in *Social Science & Medicine*. I came across this paper in looking for background literature on estimating the causal effect of neighborhood-level variables on health outcomes, which is what my project is about. Oakes argues that estimating the these neighborhood-level effects in a way that endows them with a causal interpretation is impossible to do with observational data. Specifically, he casts doubt on the ability of current statistical models to facilitate inference about the "independent causal effect of neighborhood contexts health" and argues that the causal utility of such effects are meaningless for practical social epidemiology.

Oakes' argument about the ability of statistical models to enable inference about neighborhood effects comes from two problems. First, he claims that if we adequately control for the variables that determine "selection" of individuals into neighborhoods (i.e., perfectly model the assignment mechanism), we will have effectively "explained away" all of the variability between neighborhoods and thus have no treatment effect left to model. Second, he argues that neighborhood effects emerge as the result of interactions between residents and are thus indistinguishable and unidentifiable because "there is no exogenous intervention causing them."

Oakes also presents two problems that he claim make the utility of "neighborhood effects" meaningless in a practical context. One problem is that many causal effects are based on extrapolation. For example, moving an individual of a fixed low socioeconomic status (SES) from a low-SES neighborhood to a high-SES one presumably involves at least some degree of extrapolation because there is likely little to no data on low SES individuals in high SES neighborhoods.

Another problem is that moving large numbers of (say) low SES individuals to a high SES neighborhood would itself change the composition of the high SES neighborhood and thus violate SUTVA.

Given the problems he sees in the study of neighborhood-level causal effects, Oakes advocates incorporating qualitative methods with quantitative ones and focusing on community trials where investigators randomize entire neighborhoods into treatment and control groups, similarly to what has been done in health economics. He does not see observational studies that use multilevel statistical models to estimate neighborhood-level causal effects as a useful tool in understanding the relationship between health and neighborhood contexts.

Diez Roux (2004) is a response to Oakes' questioning of neighborhood studies. She responds to each of the four problems Oakes points out and highlights areas of agreement and disagreement. First, in response to Oakes' statement that adequately accounting for "selection" of individuals into neighborhoods will remove all between-neighborhood variability, Diez Roux points out that this is "an empirical question that can be examined in the data." With enough overlap between relevant individual and neighborhood characteristics, the confounding Oakes is worried about may not be an issue. Diez Roux also questions the claim that the characteristics of a neighborhood are entirely determined by the individuals who live there. She asserts that features like availability of healthy foods, while potentially partially a function of residents' characteristics, are also affected by "exogeneous factors related to policies and macroeconomic forces."

Oakes' worries about extrapolation were, in my mind, the weakest of his arguments, and Diez Roux points out that this issue arises in many contexts and can again be empirically determined from the data. If there is sufficient overlap in (for example) neighborhood and individual level SES, then the causal effect of neighborhood SES does not require extrapolation to the extent that Oakes claims it does. Finally, while Oakes is correct in noting that moving a large number of low SES individuals from one neighborhood to another will change the characteristics of that neighborhood, Diez Roux states that most neighborhood studies are not interested in such effects; rather, they are aiming to understand the relationship between neighborhood-level characteristics and health.

Diez Roux largely agree with Oakes' recommendations for incorporating qualitative methods into neighborhood studies, but disagrees that randomized community trials are the only solution. She emphasizes the need for direct measurements of neighborhood characteristics of interest, as opposed to proxies formed by aggregating individual-level values, and better definitions of relevant areas. She concludes by pointing out that the fundamental goal of neighborhood studies is to help formulate better policy, and so understanding how much

evidence is needed in order to conclude that neighborhood-level interventions should be a component of health policy, particularly when (I would add) that evidence cannot definitively be interpreted as causal.

All of this leaves me wondering how to move forward with my own project. A question I was dealing with before I read these papers was how to incorporate causal variables at multiple levels into a DAG, which I still haven't quite figured out. In my case, the relevant neighborhood characteristic, the proportion of people who practice open defecation, is by definition formed by aggregating the characteristics of individual residents (though strictly speaking, my data are for households, not individuals). Does this mean I can't consistently estimate the effects I'm after? What about simultaneously estimating the effect of household open defecation and its interaction with neighborhood-level open defecation? So many questions...

This week's reading was Chapter 8 of Imbens and Rubin. I liked seeing the full mathematical derivation of how to get $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W})$ and from there the estimate of the causal quantity of interest, but there is enough notation and equations that it may be a bit overwhelming for less mathematically sophisticated readers. I think the explanation in Chapter 8 of BDA (Gelman et al, 2016) conveys the same ideas but in a perhaps more digestible form. On the other hand, Imbens and Rubin explicitly work through the details of each step, which is sometimes glossed over in BDA. I also appreciated the section on frequentist approaches that described some of the advantages and disadvantages over the Bayesian approach.