Published in final edited form as:

J Am Stat Assoc. 2015 December ; 110(512): 1345–1347. doi:10.1080/01621459.2015.1033058.

Comment

Michael G. Hudgens¹ [Associate Professor]

¹Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599
(mhudgens@bios.unc.edu).

Congratulations to Professors Angrist and Rokkanen (AR) for this outstanding contribution to *Applications and Case Studies*. Their paper exemplifies an analysis of a real data set which is statistically innovative as well as scientifically and practically relevant. Below are a few comments on the paper by AR.

1 Context and AR Contribution

Causal inference about the effect of a treatment or exposure in an observational setting is a risky undertaking. Typically one relies on some form of a conditional independence assumption which asserts that potential outcomes are independent of treatment status conditional on some set of covariates. Specifically, if we let Y_0 and Y_1 denote potential outcomes when treated or not, D denote the actual treatment received, and X some set of pre-treatment covariates, then we might assume $Y_d \perp\!\!\!\perp D | X$ for $d = 0, 1$ where $\perp\!\!\!\perp$ denotes independence. This assumption is sometimes referred to as conditional exchangeability or no unmeasured confounding. A weaker version of this assumption is conditional mean independence

$$E(Y_d | D, X) = E(Y_d | X) \text{ for } d=0, 1. \quad (1)$$

Conditional mean independence (or exchangeability) is not subject to empirical test because the observed data provide no information about the distribution of $Y_d | D = d, X$. Therefore subject matter expertise must be relied upon to assess the assumption's plausibility. Ultimately, many statisticians feel uneasy about drawing inferences based on assumptions which cannot be tested empirically, which likely explains the reticence of many to draw causal inferences outside of the setting of randomized trials. While sensitivity analyses might be employed to assess robustness of causal inferences in the observational setting to departures from the no unmeasured confounding assumption, the extreme version of such analyses will yield treatment effect bounds that may be wide and largely uninformative (Richardson et al. 2014).

The regression discontinuity design (RDD) provides an alternative approach. RDD has a long history (Lee and Lemieux 2010), is well studied in econometrics (Imbens and Lemieux 2008) and has recently begun to receive greater attention in epidemiology and other areas of biomedical research (Bor et al. 2014). The popularity of the RDD owes to the ability to draw causal inferences from observational data without relying on the conditional exchangeability assumption. Rather, within a neighborhood of the running variable cutoff, inferences about the effect of treatment can be drawn with no such assumption.

The downside to RDD is that the resulting inference is local in that it applies only to individuals whose running variable is near the cutoff. In the uncommon situation where it is reasonable to assume the treatment effect is homogenous across the population, the local and global average treatment effects will be equal. Otherwise, the Achilles heel of the traditional RDD approach is inability to draw inference about treatment effects for individuals far from the cutoff. AR propose a solution to this problem by making the conditional independence assumption (CIA)

$$E(Y_d|R, X) = E(Y_d|X) \text{ for } d=0, 1, \quad (2)$$

which seems not too different from assumption (1). In fact, for the sharp RDD where $D = I[R > c]$ for cutoff c , (2) is stronger, i.e., implies (1). This then begs the question: are we essentially back where we started, or has progress been made?

The answer is the latter, because unlike the usual observational setting, (2) has testable implications in the RDD setting. In particular, denote the observed response by $Y = DY_1 + (1 - D)Y_0$. Then AR show in §3.1 that if (2) holds, then for $r > c$

$$E(Y|X, R=r, D=1) = E(Y|X, D=1) \quad (3)$$

Both expectations in (3) are identifiable from the observable data; thus we can test an implication of (2) by testing whether Y is conditionally mean-independent of R given X and $D = 1$ for individuals where $R > c$. An analogous argument applies for individuals with $D = 0$ and $R < c$.

AR suggest fitting separate linear regressions of Y on X and R using data to the left and right of c . Perhaps in some scenarios a more powerful test could be constructed by conducting a single regression based test using data from the left and right of c simultaneously. Employing a more powerful test of (2) might be important in scenarios such as the 9th grade BLS applicant data where the sample size is limited and failure to reject (2) may be due to low power rather than the assumption holding. For example, suppose (based on Figure 1b discussed in the next section) we have independent normal random variables $X \sim N(0, 1)$, $U \sim N(0, .5)$, $\epsilon_r \sim N(0, 1)$, $\epsilon_{y1} \sim N(0, 1)$, and $\epsilon_{y0} \sim N(0, 1)$, and let $R = X + U + \epsilon_r$, $Y_1 = X + U + \epsilon_{y1}$, $Y_0 = X + U + 1 + \epsilon_{y0}$, and $D = I[R > c]$ where $c = 0$; then for $n = 400$ empirical results indicate the regression test (based on the likelihood ratio statistic) using only individuals with $R > c$ will have $\approx 53\%$ power, whereas fitting $E(Y) = \beta_0 I[R < c] + \beta_1 X I[R < c] + \beta_2 R I[R < c] + \beta_3 I[R > c] + \beta_4 X I[R > c] + \beta_5 R I[R > c]$ using all data and simultaneously testing $H_0 : \beta_2 = \beta_5 = 0$ yields $\approx 74\%$ power, a 40% increase in power. In their Appendix AR suggest another CIA test which utilizes data from the left and right of the cutoff by contrasting RD and CIA based effect estimates. Conducting a single CIA test also partially mitigates the challenge of assigning “significance” to multiple CIA test results as in Table 2 of AR.

2 DAGs and Covariate Selection

A substantial part of modern causal inference research utilizes directed acyclic graphs (DAGs) to determine sets of covariates which are sufficient for conditional exchangeability.

Causal DAGs are popular in areas such as epidemiology (e.g., Greenland et al. 1999) and sociology (e.g., Morgan and Winship 2007), and less so in econometrics. Proponents of DAGs such as Pearl (2009) argue DAGs offer a certain transparency about assumptions that can be obscured by mathematical equations otherwise, whereas critics of DAGs argue that important mathematical details are being obscured by simply focusing on graphs. One attribute put forth by the DAG camp is that graphs provide a means to communicate with subject matter experts (i.e., non-statisticians) in such a way that leads to a principled approach to selecting covariates sufficient for conditional exchangeability.

Figure 1 depicts three possible single world intervention graphs (SWIGs) (Richardson and Robins 2013), a particular type of DAG, for the RDD setting. Here we consider the sharp RDD where D is a deterministic function of the running variable R . Thus the only directed edge into D emanates from R . In the RDD setting AR note that failure to control for R “is the only source of omitted variable bias in estimates of the causal effect of” D . In the language of DAGs, all backdoor paths, if they exist, between treatment D and potential outcome Y_d will include the running variable R . Moreover, because R is not a collider on any such paths, conditioning on R will be sufficient to block all backdoor paths. Thus regardless of whether the SWIG is Figure 1a, 1b, or 1c, conditioning on R will imply conditional exchangeability $Y_d \perp\!\!\!\perp D | R$. This provides justification for the traditional RDD approach where inference is conditional on the subset of individuals with running variable R near the cutoff.

The approach of AR goes beyond such local inferences, but relies on the additional assumption (2). For which SWIGs is this assumption justified? Only Figure 1a, because in this case conditioning on X is sufficient to block all paths between R and Y_d . In contrast, in Figure 1b the path $R \leftarrow U \rightarrow Y_d$ remains open when conditioning only on X . Thus (2) will not hold in general. Figure 1c depicts the well-known M-bias example. Here the path $R \leftarrow U \rightarrow X \leftarrow V \rightarrow Y_d$ is blocked unconditionally, implying $Y_d \perp\!\!\!\perp R$. However, by conditioning on the collider X , this path is no longer blocked such that (2) will not hold in general. In other words, conditioning on X can actually makes matters worse, not better. Note here that X will in general be associated with, and therefore a predictor of, the outcome Y_d . Thus it cannot be universally recommended to always include dependent variable predictors in X when assuming (2).

Ordinarily when drawing causal inference from observational data, expert opinion and prior knowledge would be utilized to construct the DAG (or SWIG) for the problem under consideration. Covariates X sufficient to block all backdoor paths between D and Y_d would then be selected and inference would proceed based on the implication of the assumed DAG that $Y_d \perp\!\!\!\perp D | X$ and, therefore, (1) hold. For example, one might assume the SWIG in Figure 1a and then hope it is not the case that in truth Figure 1b or Figure 1c holds instead. The insight of AR is that, in the context of RDD, the slightly stronger assumption (2) is subject empirical test. Thus, if contrary to the assumed SWIG in Figure 1a in fact the correct SWIG is given by Figure 1b or 1c, then a CIA test (as discussed above) may indicate this to the analyst.

3 SUTVA

The Stable Unit Value Assumption (SUTVA) (Rubin 1980) is typically invoked in causal inference and implies that each individual has only two potential outcomes, one outcome if treatment is received and one outcome otherwise (denoted by Y_1 and Y_0 above). SUTVA is composed of two sub-assumptions: (i) no hidden versions of treatment, and (ii) no interference between individuals. Below we discuss (i) and (ii) in the context of AR and the Boston schooling example.

3.1 Wanna get away, but not too far?

Sub-assumption (i) that there are no hidden versions of treatment stipulates that for each treatment level d there is only one version of d . For example, if $d = 1$ denotes a particular drug, then (i) assumes there are not different versions of this drug (e.g., there is not a low dose of the drug and a high dose of the drug). Invariably there is almost always some variation in the version of treatment (e.g., the drug is taken at night or in the morning, with food or on an empty stomach, and so on). If there are multiple versions of a treatment level, it is typically assumed these versions are irrelevant in the sense that the potential outcome Y_d is the same regardless of which version of d is received by an individual. VanderWeele (2009) refers to this as the treatment-variation irrelevance assumption. In the context of the Boston schooling example, if treatment level $d = 1$ indicates a BLS admission offer, then there are multiple versions of treatment level $d = 0$ which could correspond to BLA admission, O'Bryant admission, or no admission to any of the three Boston exam schools. AR are unwilling to assume treatment-variation irrelevance for $d = 0$, in particular, that the "distinction between BLA and O'Bryant is of no consequence." They argue how incorrectly assuming treatment-variation irrelevance can lead to bias and therefore limit their analyses to certain windows around the running variable cutoff. For analysts who in the future employ AR's approach to draw inference about treatment effects away from the cutoff, it will be important to assess for the data at hand whether treatment-variation is irrelevant.

3.2 Interference

An entirely different thread of the causal inference literature has recently explored relaxing the oft made assumption that there is no interference between individuals, i.e., that the treatment assignment of one individual does not affect the outcome of any other individual. Educational studies are one setting where the no interference assumption may be dubious (Hong and Raudenbush 2006; VanderWeele et al. 2013). It would be interesting to learn AR's opinion as to whether the no interference assumption is reasonable in the context of the Boston exam schools analysis. If the assumption is not reasonable, what are the implications of possible interference on their inferences?

Acknowledgments

MGH was partially supported by NIH grant R01 AI085073. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH. The author thanks Editor Joseph Ibrahim for inviting this comment.

References

- [1]. Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*. 2014; 25(5):729–737. [PubMed: 25061922]
- [2]. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;37–48. [PubMed: 9888278]
- [3]. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*. 2006; 101(475):901–910.
- [4]. Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*. 2008; 142(2):615–635.
- [5]. Lee DS, Lemieux T. Regression discontinuity designs in economics. *Journal of Economic Literature*. 2010; 48(2):281–355.
- [6]. Morgan, SL.; Winship, C. *Counterfactuals and Causal Inference*. Cambridge University Press; 2007.
- [7]. Pearl, J. *Causality*. 2nd Edition. Cambridge University Press; 2009.
- [8]. Richardson A, Hudgens MG, Gilbert PB, Fine JP. Nonparametric bounds and sensitivity analysis of treatment effects. *Statistical Science*. 2014; 29(4):596–618. [PubMed: 25663743]
- [9]. Richardson, TS.; Robins, JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series; 2013. Technical Report 128
- [10]. Rubin DB. Discussion of “Randomization analysis of experimental data in the Fisher randomization test” by D. Basu. *Journal of the American Statistical Association*. 1980; 75:591–593.
- [11]. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009; 20(6):880–883. [PubMed: 19829187]
- [12]. Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. *Journal of the American Statistical Association*. 2013; 108(502):469–482. [PubMed: 23997375]

Figure 1.

Possible SWIGs, where U , V , X are baseline covariates, R is the running variable, D is the treatment indicator, and Y_d is the potential outcome.