Soc Psychiatry Psychiatr Epidemiol (2017) 52:139–142
DOI 10.1007/s00127-016-1334-0

COMMENTARY

Commentary: on the application of potential outcomes-based methods to questions in social psychiatry and psychiatric epidemiology

Sharon Schwartz[1]

Received: 18 November 2016 / Accepted: 24 December 2016 / Published online: 8 February 2017

Introduction

Two papers in this issue, "Disparities at the Intersection of Marginalized Groups" [1] and "Causal Inference and Longitudinal Data: A Case Study of Religion and Mental Health" [2], provide compelling arguments for the utility of methods grounded in a potential outcomes framework in social and psychiatric epidemiology. While these methods can sometimes seem complex, the authors provide very clear and accessible guidance for their use.

However, as with all methods transitioning from technical development to widespread accessibility, to application beyond their original purpose, one unintended consequence of this clarity may be their inappropriate use and interpretation in future applications. In this commentary, I will discuss two such potential risks: conflating realized causal effects with intervention effects and substituting methodological assumptions for theory.

Motivating examples

The first paper focuses on interaction, a central concept in theories of the intersectionality of social categories. The scale dependence of statistical measures is a key barrier to the fruitful application of quantitative methods to address questions about intersectionality [3]; Jackson, Williams, and VanderWeele [1] provide a clear rationale for choosing an additive model. Similarly, VanderWeele, Jackson, and Li [2] describe how marginal structural models can overcome obstacles in the use of longitudinal data to resolve the longstanding controversy about reverse causation in the relationship between religious service attendance and depression. In both instances, methods from the potential outcomes framework solve important technical problems in the estimation of causal effects in social and psychiatric epidemiology.

These applications of potential outcomes methods entailed the movement of both the potential outcomes approach and intersectionality concepts out of the realms in which they were developed. Within epidemiology, methods grounded in the potential outcomes framework were explicated largely in the context of estimating the effectiveness of treatments for HIV. The causal effects in this context can be easily conceptualized as intervention effects. How does this interpretation of "causal effects" change when applied to social and psychiatric constructs? Similarly, how do we interpret concepts like "intersectionality", when they move from the rich qualitative observations in which they were developed, into a quantitative sphere that is often seen as atheoretical [3]?

Intersectionality and religious service attendance as contested terrain for potential outcomes

In a potential outcomes approach, a causal effect is generally defined as the difference in outcome for the same individual under two different treatments; it represents what would happen if we performed a hypothetical intervention

This comment refers to the articles available at doi:10.1007/s00127-016-1281-9 and 10.1007/s00127-016-1276-6.

✉ Sharon Schwartz
sbs5@columbia.edu

[1] Columbia University, MSPH, New York, USA

to change the exposure in a specific way from one value to another. The utility and goal of estimating causal effects is often described as providing information about changing the world through interventions rather than as attempts to explain how the world works [4].

This framework therefore fits best with exposures that can be conceptualized as treatments in a hypothetical RCT. Some researchers reserve the term 'causal effect' for factors that can be manipulated (e.g., [5]); other researchers are not as strict and frame the requirement as a "well-defined intervention" (e.g., [6]). There is considerable controversy about the parameters of "well-defined", but an essential feature is that to describe a hypothetical intervention there should be a statement about how the exposure would be changed [6, 7]. This perspective seems unproblematic when applied to questions like the causal effect of AZT on HIV, where the exposure is readily conceptualized as a treatment that we would indeed like to apply and where we can easily imagine manipulating the exposure in an RCT. However, imagining the RCT-like intervention that would change the way in which individuals are socially categorized, or change their attendance at religious services, is not as easy.

In the paper Causal Inference and Longitudinal Data [2], the authors define the causal effect generally estimated by marginal structural models as: "… the counterfactual outcome … had there been interventions on the exposure at follow-up visits 1, 2 and 3 to fix these values." (p. 22). This definition aligns with the potential outcomes defintion we described above. In applying this method to religious service attendance, however, there is no description of how we would assign people to attend religious services; no intervention is specified. The type of intervention surely matters because there are many different "versions of treatment". The motivations for attending services—to atone for sins, to gain a sense of community, because of family pressure—could each have different, even opposing, effects on the development of depression. The interpretation of the causal effect estimated from the marginal structural model is therefore not obvious.

Similarly, in Disparities at the Intersection of Marginalized Groups, race, early SES and their intersection are also not well-defined interventions. Indeed, within epidemiology these variables, particularly race, are often considered "attributes", to which the potential outcomes definition of a causal effect does not directly apply [8, 9]. In both instances, then, the causal effects do not represent what would happen if we intervened in a particular way to change the exposure from one value to another. The exposures represent contested terrain since they do not meet the potential outcomes requirement for manipulability or well-defined interventions. So, exactly how should we interpret the effects estimated by the decomposition [1] and marginal structural models? Is a causal interpretation feasible?

The interpretation of the "causal effects" in these two papers

We think that a reasonable interpretation of this causal effect is the one expressed by Vanderweele [10] in his recent book which explicates potential outcomes approaches to the estimation of mediation and interaction. This is a counterfactual definition [11], but one that does not rely on the intervention specification that is required for interpreting causal effects within a potential outcomes framework. We think this represents a realized causal effect [12].

> "The [potential outcomes] framework provides a formal and technical notation to conceptualize causation. This is done principally by conceiving of what might have occurred had some action or state been otherwise than it was. If some outcome would have been different had some exposure or action been other than it was, then we would say the exposure or action causes or affects that outcome … What the counterfactual framework allows for principally is a set of definitions that provide either criteria or sufficient conditions indicating that some event or exposure was a cause of another" (p. 4).

This effect answers the crucial question—did the exposure cause the outcome? There are three things of note in VanderWeele's definition that distinguish it from the more common definitions of potential outcomes. First, it specifically allows "states" as legitimate exposures for estimating causal effects; second, there is no intervention invoked; and third, it is about what actually happened rather than a prediction of what would happen under some intervention. Here, the counterfactual is a thought experiment, akin to

[1] The authors of this paper seem somewhat ambivalent about the causal status of the effect estimated. On the one hand, there is a disclaimer that the paper only estimated an association, but on the other they interpret the results of the decomposition model as estimating the effect that was due to race alone, due to SES alone, and due to the intersection of race and SES. This certainly seems like a causal attribution. Indeed, the authors clearly state that the study results are applicable when "disparities are taken to represent causal effects of race and SES, when proper adjustment has been made for confounding variables". Therefore we are assuming that the interpretation was for an estimate where confounder control was in effect, and the association disclaimer is just a "technical" issue in this particular analysis.

Pearl's surgery on equations—a causal contrast without an articulation of how the exposure was removed [13].

A realized causal effect cannot be interpreted as an estimate of some potential intervention on the exposure. Nonetheless, such causal effects are quite useful when the research question is about understanding how causes work in the world as a prelude to designing useful interventions. This interpretation seems well suited to the causal effects estimated in these papers, but is a departure from more common interpretations of potential outcomes effects. We think it would be helpful to "name" this type of causal effect to avoid inappropriate interpretations and recognize differences in the assumptions necessary for valid estimation.

Starting from the theory

Just as the interpretation of causal effects change when a method is applied outside the domain in which it was developed, the meaning of a social construct can change and require interpretational attention when uprooted from its theoretical origins. Intersectionality is a prime example. The paper on intersectionality [1] rightly focuses on interaction as an important component for quantitative studies of intersectionality that could benefit from potential outcomes approaches. The authors chose a particular additive model, the joint disparity decomposition model, due to its intrinsic policy value in describing the gains that would be achieved were the disparity removed, and because "the decomposition sheds light on the importance of each social status category and its intersection" [2] (p. 15). While these certainly represent benefits of the decomposition model, are these benefits commensurate with the concept of intersectionality? Does this measure advance the intersectionality theory? Would another measure have been chosen if we started from the social concepts and their theoretical underpinnings, rather than the methodological toolbox?

In considering these questions, let us take as an example the direct policy relevance of the decomposition model in relation to intersectionality theory. Although intersectionality draws our attention to groups which are overlooked when only one dimension of disadvantage is examined, it is hard to imagine that the effects estimated describe the "absolute gains in the population outcome that would be achieved were the disparity removed" p. 7. While this interpretation fits with causal effects for specific interventions, it does not fit with effects for which no intervention is specified. Interventions to reduce disparity would likely have effects through mechanisms other than changing the disparity. For example, even if intersectionality were important in determining incarceration rates, the disparity may be addressed through changes in policing policies and sentencing laws that may affect the absolute incarceration rate in many ways other than the disparity removal. The effect of intersectionality in producing the outcome, what the study estimates, and the effect of removing the disparity may well differ. Indeed, this interpretational problem is well recognized in potential outcomes approaches that require well-defined interventions [6].

Consider, for example, how the second justification for this approach, the decomposition component, may also alter the meaning—and intended purpose—of intersectionality theory. The notion of decomposing the causal effect into the proportion that is due to race, due to early SES, and due to their interaction, is antithetical to an intersectionality perspective. As articulated in the introduction to the paper, "an intersectional perspective recognizes that social categories are mutually constitutive in that the experience of one social category may differ across other categories". While finding an interaction indicating that race and early SES have greater effects than would be expected based on the additive effects of each alone supports intersectionality in producing these outcomes, it is difficult to imagine that a "mutually constitutive social category" could be divided into its component parts. Intersectionality does not only imply that the racialized low SES group has a greater burden than would be expected based on the independent effects of race and SES, but that the very concepts of race and SES themselves are race and class specific. Because of this interdependence in defining the exposure, the effect for the white low SES group is not the effect of low SES alone; rather, it is the effect of low SES in the context of being white.

Indeed, in the discussion section of the paper, these processes are clearly described. Thus, the decomposition method alters the very meaning of intersectionality. Perhaps, a method that provides evidence of an interaction but does not allow for the decomposition of the effects, rejected by these authors, would have been more consistent with the meaning of intersectional effects.

To maintain the integrity of the constructs, it might be useful to start from intersectionality theory about the particular outcomes under investigation—that is some theory about how race and early life SES intersect to create disparities in employment, wages and incarceration. For example, if we started from the theories described in the discussion section of the paper, rather than invoking them to help interpret the results from the decomposition models, what tools from the potential outcomes treasure trove would we select?

If we start with intersectionality theory, a central notion is that the meaning of social statuses and identities are context dependent. The meaning of race depends on early SES and the meaning of early SES depends on race—they are inextricably linked. This concept would seem closest to the

sufficient cause model of interaction and suggest the use of methods for assessing this "mechanistic" interaction [10]. As Vanderweele notes, this type of interaction focuses on "why and how" particular exposures affect outcomes in addition to the "for whom" and "to what extent" of other additive interaction approaches. The rich theorizing and qualitative evidence from intersectionality's roots provide hypotheses about the processes through which social statuses and identities emerge and describe which intersections are relevant for particular outcomes. They are about the "whys and hows".

Potential outcomes approaches to mediation would also be useful to test the mechanisms through which intersectionality is hypothesized to affect particular outcomes. For example, in the discussion section it is hypothesized that race and low SES may intersect due to the poor quality of schools in segregated neighborhoods. This mediational process could be examined using potential outcomes-based approaches. The question about the appropriate scale to test intersectional interactions could be confronted, with arguments made for the general preference for additive models but with the particular type of interaction dependent on the specific question, hypotheses, and goals of the social theory.

Conclusion

These articles will certainly encourage the use of tools from the potential outcomes framework to address a broader array of causal questions. Strict interventionist interpretations of the requirements for the estimation of causal effects have sometimes discouraged the application of counterfactual frames to the types of causal questions that are common in social and psychiatric epidemiology (e.g., [14]). The two articles in this volume amply illustrate the utility of the potential outcomes frame even when the exposure is not "manipulable" in the sense of being assignable in an RCT. However, as these tools are applied to questions for which they were not developed, at least within epidemiology, we need to attend to the appropriate interpretation of the causal effects estimated, and the commensurability of the way in which the methods have been used in the past to their utility in new arenas. On the other hand, we need to ensure that the constructs on which the questions are based remain intact when scrutinized using potential outcomes approaches. In this work, we need to ensure that the theory and question drives the method rather than the reverse.

Acknowledgements The ideas in this commentary were developed with the SUN collaborative (Ulka Campbell and Nicolle Gatto). It has greatly benefitted from comments by Seth Prins, Ilan Meyer and Lisa Bates on earlier drafts.

Compliance with ethical standards

Conflict of interest There is no conflict of interest.

Ethical standards An approval by an ethics committee was not applicable.

References

1. Jackson JW, Williams DR, VanderWeele TJ (2016) Disparities at the intersection of marginalized groups. Soc Psychiatry Psychiatr Epidemiol 51(10):1349–1359. doi:10.1007/s00127-016-1276-6
2. VanderWeele TJ, Jackson JW, Li S (2016) Causal inference and longitudinal data: a case study of religion and mental health. Soc Psychiatry Psychiatric Epidemiol 51(11):1457–1466. doi:10.1007/s00127-016-1281-9
3. Bauer GR (2014) Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. Soc Sci Med 110:10–17. doi:10.1016/j.socscimed.2014.03.022
4. Hernán MA, VanderWeele TJ (2011) Compound treatments and transportability of causal inference. Epidemiology (Cambridge, Mass) 22(3):368–377. doi:10.1097/EDE.0b013e3182109296
5. Holland PW (1986) Statistics and Causal Inference. J Amer Statistical Assoc 81:945–960
6. Hernán MA, Robins JM (2017) Causal inference. Chapman & Hall/CRC, Boca Raton (forthcoming)
7. Little RJ, Rubin DB (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 21:121–145. doi:10.1146/annurev.publhealth.21.1.121
8. VanderWeele TJ, Robinson WR (2014) On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology (Cambridge, Mass) 25(4):473–484. doi:10.1097/ede.0000000000000105
9. Naimi AI, Kaufman JS (2015) Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. Curr Epidemiol Rep 2(1):52–60. doi:10.1007/s40471-014-0030-4
10. VanderWeele TJ (2015) Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press, New York
11. Glymour C, Glymour MR (2014) Commentary: race and sex are causes. Epidemiology (Cambridge, Mass) 25(4):488–490. doi:10.1097/ede.0000000000000122
12. Schwartz S, Gatto NM, Campbell UB (2016) Causal identification: a charge of epidemiology in danger of marginalization. Ann Epidemiol 26(10):669–673. doi:10.1016/j.annepidem.2016.03.013
13. Pearl J (2010) The Foundations of Causal Inference. Sociological Methodology 40(1):75–149. doi:10.1111/j.1467-9531.2010.01228.x
14. Hernán MA, Taubman SL (2008) Does obesity shorten life? The importance of well-defined interventions to answer causal questions. Int J Obesity (2005) 32 Suppl 3:S8–S14. doi:10.1038/ijo.2008.82