Mediating to Opportunity: The Challenges of Translating Mediation Estimands into Policy Recommendations

Louisa H. Smith,[a] *and Gabriel L. Schwartz*[b]

In 1994, the US Department of Housing and Urban Development undertook an ambitious policy experiment, dubbed Moving to Opportunity (MTO).[1] The study aimed to answer one of the most vexing questions in social science: does living in a neighborhood with more economic opportunity and lower poverty mean better well-being? And could government housing vouchers change people's neighborhood contexts enough to matter? More than 4600 families living in public housing were randomized to receive either business-as-usual public housing benefits or one of two types of housing vouchers, allowing them to move into private housing. The results, at least for children, were murky: although younger children and older girls whose families received vouchers both seemed to enjoy long-term economic, behavioral, and health benefits, older boys appear to have been hurt by the intervention, with higher levels of substance use and poorer mental health.[2–6] Although many hypotheses have been advanced to explain why older boys fared so poorly, consensus is a long way off.

Faced with such strong treatment effect heterogeneity and multiple competing hypotheses about that heterogeneity, what's an epidemiologist to do? Mediation analysis provides one path forward. In this issue, Rudolph et al. turned to mediation to reanalyze boys' MTO data in an attempt to understand what drove the harmful effects boys experienced, examining behavioral and substance abuse outcomes and a set of mediators describing boys' school, neighborhood, and social environments.[7] In the context of that work, this commentary highlights key methodologic and conceptual decision points for epidemiologists turning to mediation in an effort to autopsy harmful policy experiments and improve interventions. First, we propose distinct goals of mediation analyses in such situations. Second, we discuss several common choices for mediation estimands and the questions they answer. Third, we assess the utility of mediation for unpacking treatment heterogeneity, particularly in the case of housing interventions.

THE GOAL OF MEDIATION IN POLICY EXPERIMENTS

It is worth making explicit the goals of using mediation to understand harmful policy experiments. Three may be common: First, researchers may be interested in leveraging the experiment to answer etiologic questions that are broadly applicable to a discipline (advancing theory), treating the policy experiment per se as a nuisance. Second, researchers may want to understand what went wrong (in general or for a specific subgroup) to supplement the original intervention with additional supports that could help any harmed subgroup(s)

Editors' note: Related article appears on page 336. From the [a]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; and [b]Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, San Francisco, CA.
The authors report no funding and conflicts of interest.
LOUISA H. SMITH is a doctoral student in epidemiology at the Harvard T.H. Chan School of Public Health. She is interested in understanding and quantifying bias and in developing and applying methods for complex treatments and exposures, particularly those affecting children's health and development. GABRIEL L. SCHWARTZ is a social epidemiologist and postdoctoral fellow at the UCSF Institute for Health Policy Studies. His work examines how social policy, social stratification, and the places we live shape health inequities across the life course, with a particular focus on housing and segregation.
Correspondence: Louisa H. Smith, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave., Boston, MA 02115. E-mail: Louisa_h_smith@g.harvard.edu.

ISSN: 1044-3983/21/323-347
DOI: 10.1097/EDE.0000000000001335

enjoy the same benefits of treatment as other groups. Finally, researchers may want to understand what went wrong (in general or for a specific subgroup) to develop a new, entirely redesigned policy intervention that avoids such pitfalls.

CHOOSING AN ESTIMAND

Various mediation estimands could be used to address these goals. Natural direct and indirect effects[8,9] have become popular in part because they attractively sum to the total effect.[10] This allows for a causal effect to be partitioned into a portion that acts through a specific mediator and another that does not. Intuitively, natural effects assess the causal effects of treatment mediated (indirect) or not mediated (direct) through some mediator(s) by fixing the mediator to the specific value it would have taken under treatment or no treatment. Despite natural effects' appealing qualities, Rudolph et al[7] eschew them for good reason. Natural effects require strong identification assumptions[9,11,12] and are difficult to interpret:[13] in this situation, the natural direct effect, for example, would be a contrast between the probability of a given behavioral or substance use outcome had every boy's family not received a voucher, and the probability of the outcome had all families received vouchers although each boy was exposed to the school, neighborhood, and social environments he would have lived in had his family not received a voucher. Since we do not know what environment that would be for individual boys, this does not directly correspond to a real-life intervention, making it more appropriate for etiologic or mechanistic exploration than for determining actionable policies.

One alternative is the controlled direct effect, which has been recognized as a more actionable estimand for public health questions.[8,13,14] This estimand quantifies the treatment effect when the mediator(s) is set to some known value under both treatment conditions. The implied contrast for the MTO scenario would again be between assigning families to receive a voucher or not, but under both circumstances, boys would be exposed to the same values of the mediators—for example, living in neighborhoods with equally low poverty. Although such an intervention may be helpful for advancing theory about the treatment (does the voucher have downstream effects net of some set of hypothesized targets for intervention?), it may be unrealistic or unactionable as policy vis-à-vis the mediator. In addition, controlled direct effects have limited utility if the goal is to quantify the mediated portion of the total effect, as there are generally no corresponding indirect effects,[15,16] although the contrast between the direct and total effects can be conceptualized as the portion of the harmful effect eliminated by a policy intervention.[17]

Although natural effects can be used to ask and answer theoretical questions about mechanisms, and controlled direct effects about a possibly radical intervention on the mediator, the stochastic effects Rudolph et al[7] estimate fall somewhere in between. Instead of assigning each boy to the environment he would have experienced in the absence of a housing voucher or all boys to the same environment, these estimands instead apply distributional interventions on the mediator.[18,19] This better maps to an intervention we can imagine, and even possibly test,[20] one in which the population of boys is exposed to the distribution of environmental characteristics they would have been exposed to in the absence of the voucher. Unlike natural effects, stochastic effects also allow for identification under mediator–outcome confounding affected by the exposure. This was one motivation for Rudolph and colleagues[7]: because using the voucher could affect all mediators and outcomes when voucher receipt is the intervention, voucher uptake acts as such a confounder (as would any mediating factor that caused another mediator).

UNPACKING TREATMENT HETEROGENEITY IN POLICY INTERVENTIONS

Although stochastic estimands correspond with possibly implementable policies, Rudolph et al's[7] conclusion from their analysis is a mechanistic one: most of the harmful effects of voucher assignment on boys' behavioral and substance use outcomes operated through the environmental characteristics they investigated. Adhering to this journal's recommendations on policy implications,[21] the authors refrain from commenting on how their findings could be applied in future housing interventions. Such restraint is appropriate, as even with advanced statistical methods and rich data, their analysis can only tell us part of what we would wish to know to implement policy changes.

What is it that we would like to know to influence policy? Freeing ourselves from the restriction of existing estimands, let us reconsider the goals of mediation we introduced earlier. What questions would we ideally answer, and how do they relate to the question addressed in the analysis?

In a mechanistic sense, the question of interest may be: *Why do boys have worse outcomes if their families are assigned vouchers compared with if they are not, but girls have better outcomes?* On the other hand, policy considerations lead us to ask: *What can we do for boys whose families receive vouchers to ensure that their behavioral and substance use outcomes at least do not worsen and at best improve, while maintaining improvements in girls' outcomes?* Or perhaps: *Can we identify mediators with beneficial effects on behavior and substance use for all children and design a housing intervention to affect only those? Or target the children for whom the mediators have only positive effects?*

Contrast these with the question implied by the analysis: *If we could assign boys' families vouchers but intervene to keep the distribution of their school, social, and neighborhood environments the same as if we had not assigned them vouchers, would they still have worse behavioral and substance use outcomes than if we had not assigned them vouchers and maintained the same environments?* Although motivated by the fact that gender was a qualitative effect modifier for the effect of voucher receipt on the outcomes, the heterogeneity of the effect is addressed only by restricting the question to boys.

Learning what mediates a harmful effect among a subset of the population does not, alone, determine what went wrong for that subpopulation when the intervention worked for other groups or tell us how to implement a policy to reverse those harms although maintaining benefits for everyone else. Rudolph and colleagues' analysis shows that, although voucher receipt had generally positive effects on objective measures of boys' school and social environments, the shift in those mediators generally negatively affected the outcomes and accounted for much of the intervention's negative impact on boys. However, because we do not know anything about the role of gender in this mediation process, the implications are limited. Gender could be a qualitative effect modifier of the exposure–mediator effect (girls see effects on their school and social environments in the opposite direction), a qualitative effect modifier of the mediator–outcome effect (girls' outcomes are improved instead of harmed with improvements in their environments), or it is possible that the effect for girls is not mediated through the same process as for boys. These possibilities could have different implications for understanding what drives boys' and girls' well-being when changing neighborhoods, for developing supplemental interventions within school and social environments to complement future housing interventions, or for making decisions about which families to target with vouchers.

This example of "moderated mediation" exemplifies the challenges that come with translating an observation about the distribution of health and disease into a causal and statistical estimand and then into actionable policy. Housing interventions aimed at moving individuals into new neighborhoods present special challenges. For example, another possible pathway that may be harming boys relative to girls is that two-thirds of the experimental voucher group moved back to (on average) lower-income neighborhoods.[22] These movers, in other words, experienced heightened residential instability as a result of voucher receipt but did not retain their initial reductions in neighborhood poverty. These movers may have chosen to move because their initial moves were more damaging in terms of increased social alienation compared with those who moved once and stayed put. Which causal process is driving the harmful treatment effects among boys relative to girls, and which of those processes would be most efficient to intervene upon to correct the intervention's harmful effects on boys, is a difficult question, with a mountain of assumptions required to get any leverage. Importantly, because boys have siblings, any intervention at the family level also has to consider its effects on girls. Ideally, we want to prioritize how to most efficiently alter or supplement the original intervention such that older boys also experience benefits, without losing the benefits that other children accrue; but figuring out how to do so in a single, causally identified model may be unrealistically ambitious.

This is not to say that we should not even attempt to answer these questions. Analyses like Rudolph and colleagues'[7] can guide future work that delves deeper into the strongest mediators, their different effects among boys relative to girls, and the best place for intervention. This can move us toward hypotheses about what drove treatment heterogeneity and which policies may need to be fixed (or abandoned). Emerging work on stochastic mediation estimands[23,24] allows for answering questions that may be more targeted toward policy interventions with heterogeneous treatment effects, now that we know, mechanistically, that a given set of mediators are important.

Critically, any attempt to use statistical models to develop targets for supplementary interventions or policies to implement them requires not only careful thought but the right quantitative data to perform the relevant analysis (and often a lot of it). When we do not have that, or when we need to refine the hypotheses generated by analyses like Rudolph et al's[7], qualitative work can step in to further assess potential mediators and inform improvements in intervention design.

REFERENCES

1. Rothwell J. *Sociology's revenge: Moving to Opportunity (MTO) revisited*. The Brookings Institution. May 6, 2015. Available at: https://www.brookings.edu/blog/social-mobility-memos/2015/05/06/sociologys-revenge-moving-to-opportunity-mto-revisited/. Accessed 5 January 2021.
2. Chetty R, Hendren N, Katz LF. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am Econ Rev.* 2016;106:855–902.
3. Clampet-Lundquist S, Kling JR, Edin K, Duncan GJ. Moving teenagers out of high-risk neighborhoods: how girls fare better than boys. *AJS.* 2011;116:1154–1189.
4. Kling JR, Liebman JB, Katz LF. Experimental analysis of neighborhood effects. *Econometrica.* 2007;75:83–119.
5. Sanbonmatsu L, Ludwig J, Katz LF, et al. *Moving to Opportunity for Fair Housing Demonstration Program—Final Impacts Evaluation*. US Department of Housing and Urban Development, Office of Policy Development and Research; 2011.
6. Schmidt NM, Glymour MM, Osypuk TL. Adolescence is a sensitive period for housing mobility to influence risky behaviors: an experimental design. *J Adolesc Health.* 2017;60:431–437.
7. Rudolph KE, Gimbrone C, Díaz I. Helped into harm: mediation of the effect of a housing voucher intervention on long-term mental health and substance use outcomes in adolescent boys. *Epidemiology.* 2021;XX:XX–XX.
8. Pearl J. Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann; August 2001:411–420.
9. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992;3:143–155.
10. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol.* 2013;42:1511–1519.
11. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, August 2005:357–363.
12. Tchetgen Tchetgen EJ, Vanderweele TJ. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology.* 2014;25:282–291.
13. Naimi AI, Kaufman JS, MacLehose RF. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *Int J Epidemiol.* 2014;43:1656–1661.
14. Kaufman JS. Commentary: gilding the black box. *Int J Epidemiol.* 2009;38:845–847.
15. Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiol Perspect Innov.* 2004;1:4.

16. VanderWeele TJ. Controlled direct and mediated effects: definition, identification and bounds. *Scand Stat Theory Appl.* 2011;38:551–563.
17. VanderWeele TJ. Policy-relevant proportions for direct effects. *Epidemiology.* 2013;24:175–176.
18. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. In: Proceedings of the *Twenty-second* Conference on Uncertainty in Artificial Intelligence. Arlington, VA: AUAI Press; 2006:138–146.
19. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator–outcome confounder. *Epidemiol Camb Mass.* 2014;25:300–306.
20. Moreno-Betancur M, Carlin JB. Understanding interventional effects: a more natural approach to mediation analysis? *Epidemiology.* 2018;29:614–617.
21. Editor-in-Chief. Our policy on policy. *Epidemiology.* 2001;12:371–372.
22. Kingsley GT, Pettit KLS. *Have MTO families lost access to opportunity neighborhoods over time?* Urban Institute. 2008. Available at: https://www.urban.org/sites/default/files/publication/31586/411637-Have-MTO-Families-Lost-Access-to-Opportunity-Neighborhoods-Over-Time-.PDF. Accessed 15 December 2020.
23. Loh WW, Moerkerke B, Loeys T, Vansteelandt S. Heterogeneous indirect effects for multiple mediators using interventional effect models. *arXiv.* 2020. Available at: http://arxiv.org/abs/1907.08415. Accessed 5 January 2021.
24. Moreno-Betancur M, Moran P, Becker D, Patton GC, Carlin JB. Mediation effects that emulate a target randomised trial: simulation-based evaluation of ill-defined interventions on multiple mediators. *ArXiv190706734 Stat.* July 13, 2020. Available at: http://arxiv.org/abs/1907.06734. Accessed 10 August 2020.