Evaluating Public Health Interventions: 5. Causal Inference in Public Health Research—Do Sex, Race, and Biological Factors Cause Health Outcomes?

Counterfactual frameworks and statistical methods for supporting causal inference are powerful tools to clarify scientific questions and guide analyses in public health research. Counterfactual accounts of causation contrast what would happen to a population's health under alternative exposure scenarios.

A long-standing debate in counterfactual theory relates to whether sex, race, and biological characteristics, including obesity, should be evaluated as causes, given that these variables do not directly correspond to clearly defined interventions. We argue that sex, race, and biological characteristics are important health determinants. Quantifying the overall health effects of these variables is often a natural starting point for disparities research.

Subsequent assessments of biological or social pathways mediating those effects can facilitate the development of interventions designed to reduce disparities. (*Am J Public Health.* 2017;107:81–85. doi: 10.2105/AJPH.2016.303539)

M. Maria Glymour, ScD, and Donna Spiegelman, MS, ScD

> "Epidemiology is the science of understanding the causes and distribution of population health so that we may intervene to prevent disease and promote health."
> —*Keyes and Galea*[1]

In this fifth column in the Evaluating Public Health Interventions series, we begin a set of commentaries on causal inference. Many public health researchers were taught that it is best to avoid discussion of causation in interpreting findings from observational studies; any reference to causation was thought to overreach the evidence. The hesitation to embrace causal methods may be exacerbated by recent controversies about counterfactuals and reports arguing that race, sex, and biological states such as obesity are not quantifiable causes of health outcomes (see references 2–5 and responses). In this commentary, we argue that causal inference methods are valuable tools for researchers focusing on public health and health disparities. Counterfactual thinking, and the quantitative tools derived from it, can be as fruitfully applied to studies of race, sex, and biological states as to studies of any other health risk factors.

COUNTERFACTUAL FRAMING OF CAUSATION

Comprehensive evidence from ideal randomized trials is rarely available for answering public health questions, so methods to support causal inference in observational settings are required. Modern causal inference in health research relies on a flexible and powerful framework arising from counterfactual perspectives (see the box on the next page). Counterfactualists define the effect of an exposure on an outcome—for example, the effect of college completion on subsequent development of depression—as a contrast between potential outcomes. The effect of interest may be conceptualized as the difference in the incidence of depression among individuals having completed college and the incidence among the same individuals had they not completed college. Because each individual either does or does not complete college, only one of these potential outcomes can be observed; the other remains "counterfactual."

The individual-level difference in such cases can never be directly observed, not even in a randomized trial. But with appropriate data and a set of clearly stated assumptions, the average causal parameter can be estimated for a population, for example to describe the incidence in the population if everyone were to complete college as opposed to if no one were to complete college. Specifying the counterfactual contrast requires choosing the unit of observation (in our example, individuals), the outcome (depression), the exposure (college), and how the two potential outcomes would be compared (e.g., according to difference in risk).

Counterfactual thinking elucidates why certain statistical analyses of nonrandomized studies can produce unbiased causal effect estimates. Such methods offer a rigorous scaffold upon which causal

ABOUT THE AUTHORS

M. Maria Glymour is with the Department of Epidemiology and Biostatistics, University of California, San Francisco. Donna Spiegelman is with the Departments of Epidemiology, Biostatistics, Nutrition, and Global Health, Harvard T. H. Chan School of Public Health, Boston, MA.

Correspondence should be sent to M. Maria Glymour, ScD, Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158 (e-mail: mglymour@epi.ucsf.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted October 17, 2016.
doi: 10.2105/AJPH.2016.303539

WHAT IS THE COUNTERFACTUAL FRAMEWORK FOR CAUSAL INFERENCE?

In the counterfactual framework for causal inference, causal inferences can be drawn when the distribution of the observed outcomes among those who did not receive the intervention equals in expectation the distribution that would have been observed had those who received the intervention not received it. Because no individual can both receive an intervention and not receive it at the same time, causal inferences implicitly contrast actual values with counterfactual values. Various old and new methods of study design and analysis aim to approximate as closely as possible the ideal contrast of counterfactuals with those actually observed. These methods include randomization, matching at the design phase, nearly saturated regression, nearly saturated propensity score matching or adjustment, doubly robust methods, and instrumental variable methods. The methods traditionally used to control for confounding are intended to produce counterfactual contrasts. To relate terminology from older literature to this current framework, it can be said that when exchangeability between those receiving and not receiving an intervention is obtained, the causal counterfactual can be estimated. When exchangeability (i.e., no confounding) is achieved, the contrast of actual outcomes among people who received the intervention in comparison with those who did not receive the intervention can provide the causal effect of the intervention. Analogously, we can replace the words "receive the intervention" with "are exposed" or "have the risk factor" and proceed with a counterfactual framework for causal inference.

inferences in nonrandomized studies can be drawn. It is only by invoking these causal inference methods that we can justify claims such as "if we make this change, it will improve public health." Without such a framework, researchers who cannot conduct randomized trials are relegated to making only vague pronouncements regarding associations, often impeded from recommending actions to remediate inequalities and improve public health.

When specifying a causal question, it is useful to imagine an intervention that corresponds to the groups to be contrasted. Describing an intervention helps identify ambiguities in the definition of the exposure and clarifies the public health relevance of the question. Describing a conceptual intervention can be useful even if the intervention is not feasible or ethical.

"NO MANIPULATION, NO CAUSATION"

Many advocates argue that the counterfactual framework necessarily entails that only manipulable interventions can be evaluated as causes. An early version of this argument precluded consideration of sex and race as possible causal factors[6]; more recent work argues that although they may be causes, it is not possible to quantify the magnitude of the impact of nonmanipulable factors.[4,5] Here we refer to this argument as the "no manipulation, no causation" (NMNC) perspective. Advocates of this view often write and teach as if there were consensus among causal inference researchers, but this is not the case. Leading thinkers have debated this point for more than 30 years.[3–8]

Two types of arguments are commonly offered to defend the NMNC perspective, but we find neither convincing. First, with respect to race and sex, it is often argued that these characteristics are so intrinsic to an individual that it is impossible to imagine the same individual but of a different race or sex, as required to construct the counterfactual contrast.[9] This perspective assumes that researchers know which particular characteristics are absolutely essential to an individual's being. Such definitions are unavoidably subjective and arbitrary. More troubling, this perspective implies that the most powerful factors—those that shape so much of our lives—are not causes at all.

The second line of argument given against conceptualizing race and sex as causes is that the contrast under consideration is poorly defined, because either the characteristics cannot be changed or the "effect" of these characteristics would depend on how we change them. In other words, although they may be causes, unless the intervention is specified, we cannot quantify the effects.[4] We agree that specifying an intervention helps reduce ambiguity in interpretation of the findings, but this is not a sufficiently compelling concern to justify the most extreme version of the NMNC position. Every intervention definition includes ambiguities; even if we describe an intervention, we can only hope to reduce ambiguity to some reasonable degree, a criterion that is inevitably subjective and contextual. If the intervention is simply taking a pill, must we take it with water? In the morning? It is impossible and unnecessary to eliminate all ambiguity in applying causal inference methods to establish causality.

When evaluating the causal effect of sex, it is useful to specify the contrast one has in mind. Sex, "a biological construct premised upon biological characteristics enabling sexual reproduction,"[10(pp694–695)] should not be conflated with gender, "a social construct regarding culture-bound conventions."[10(p694)] Thus, a causal evaluation of the effect of sex would correspond to a counterfactual contrast of people carrying XX versus XY chromosomes. Given this definition of sex, evaluating the total effect of sex on cumulative mortality would be relatively straightforward. For example, one could follow a birth cohort of people with XX versus XY chromosomes using rigorous causal inference methods to adjust for potential selection bias (differential selection of babies into the cohort in a way that depends jointly on sex and risk of mortality) and information bias (resulting from errors in dates of death, which may or may not depend on sex, and sex-dependent differential loss to follow-up). Once the total effect of sex is established, the mechanisms can be dissected through causal mediation analysis.[11]

Such mechanistic evidence should guide interventions designed to mitigate inequalities. Sex chromosomes may influence health by shaping gender and gendered social interactions, differentially exposing males and females to discrimination, poverty, violence, and other socially modifiable experiences (and, of course, sex chromosomes are not the only determinants of such exposures). Sex-linked biological characteristics influence health in part through a sequence of highly context-specific social experiences that are themselves amenable to causal analysis. In an attempt to resolve the NMNC conundrum, it is now sometimes argued that one could conceptualize an intervention to change chromosomal sex, such as by somehow switching one X for a Y chromosome in a fertilized egg.[12] Although this hypothetical intervention is not humanly feasible, many NMNC advocates find implausible interventions acceptable.[4]

If causal inference about sex were taken off the table, interventional research aimed at reducing disparities by targeting hypothesized mechanisms might be hindered as a consequence. Adopting a counterfactual framework for sex will only serve to facilitate important research on related constructs such as gender relations and gender identity.[13,14]

RACE AS A CAUSE

Conceptualizing race as a cause is less straightforward, because an individual's race cannot be altered by a single genetic switch. As with sex, there is ambiguity in what people mean by any particular racial categorization, and culturally recognized racial categories evolve over time and differ between places. Although disparities researchers may adopt inclusive definitions of race, for example viewing race as reflecting "common geographic origins, ancestry, family patterns, language, cultural norms and traditions, *and* the social history of particular groups,"[15(p70)] there is much heterogeneity in these definitions across health research. As a result, interpretation of racial effects in health research has been markedly inconsistent.

Could we specify a hypothetical intervention that corresponds to the effect of race to clarify the meaning? In addition to changing genetic variants that are associated—in whichever culture we are examining—with the racial groups being compared, under hypothetical interventions attempting to quantity the full impact across the legacy of US history, we may need to consider changes in the race of an individual's family members along with the physical, social, and cultural inheritance of generations of racial inequality.

Interventions that would change all social and physical features corresponding with racial groupings are generally not feasible, but describing such interventions helps clarify interpretations of effect estimates for any given racial contrast. Imagining the scope of interventions that correspond to contemporary racial inequalities in the United States can also illuminate why many interventions have been found to achieve only modest reductions in racial health disparities. Given the range of characteristics plausibly linked with what we measure as race, it is not surprising that ambitious, multifaceted strategies are needed to eliminate disparities.[16]

Some disparities researchers have proposed a useful work-around to the NMNC contradiction, suggesting that we do not talk about the effects of race on health but only the potential effects of various interventions designed to reduce racial inequalities in health.[3] Although a step forward from the usual NMNC perspective, this seems unnecessarily limiting; it is helpful to begin investigations into interventions to eliminate racial inequalities by estimating the effect of race on health in the current context. We can evaluate the causal effect of race in any particular culture at any particular time while simultaneously recognizing the fluidity of racial categories across time and place. Evaluations of the determinants of such categorizations (i.e., the social process of defining race), although not commonly the purview of public health, are also amenable to counterfactual thinking.

One of the most unfortunate possible consequences of the debate on whether the health effects of race and sex are legitimately evaluable within the counterfactual framework through use of rigorous causal inference methods would be to deter disparities researchers from adopting rigorous approaches to causal inference. Quantifying the overall effects of sex and race on health is often a natural starting point for assessing the mechanisms mediating those effects. To reduce disparities, such mechanistic analysis is critical for identifying the most fruitful points along the causal pathway to intervene and for evaluating the interventions developed as a result of this analysis.

BIOLOGICAL CHARACTERISTICS AS CAUSES

A particularly broad interpretation of the NMNC criterion excludes efforts to estimate the causal effects of any "state" or biological characteristic, for example body mass index (BMI). Under this extreme NMNC perspective, it is argued that we can only meaningfully discuss the causal effects of particular interventions on BMI (e.g., exercise regimens); it is not meaningful to discuss the causal effects of alternative BMI levels.[17] One concern raised is that BMI violates the consistency assumption, which requires that the exposure under examination be a clearly defined construct that has the same meaning for each person (or unit) included in a study (i.e., "there were a number of ways in which the treatment could have been assigned, but all those ways would have resulted in the same observed outcome"[18]). The plausibility of this assumption requires careful consideration when evaluating the effects of a biological state.

For example, BMI differences may reflect differences in multiple underlying physiological variables, including muscle mass, fat mass, and central versus peripheral adiposity. One would not be able to estimate the effect of muscle mass by comparing people with different amounts of fat mass, although both correspond to differences in BMI. In our view, this does not preclude evaluating BMI or other physiological states as causal factors. It is possible that the most common, plausible pathways to higher BMIs produce the same overall health

outcomes, and further research is under way in this area.[19]

"Fat-hand" interventions—interventions that clumsily trigger multiple mechanisms beyond the specific target—are another challenge in evaluating the effects of BMI or other physical characteristics. This metaphor invokes the image of a hand reaching for a chess piece but accidentally knocking over adjacent pieces.[20] In relation to BMI, exercise interventions may influence health through mechanisms outside of their effect on BMI, such as changes in lung capacity or neurotrophin expression.

Concerns about consistency and "fat-hand" mechanisms merit attention when specifying causal questions and interpreting observational evidence, but neither should preclude evaluation of the causal effects of physical characteristics. Estimating the health effects of biological characteristics is critical for public health research, providing a way to anticipate the likely consequences of diverse sets of interventions. Consider, as another example, hypertension. Although there are many ways to reduce blood pressure to a healthy range (e.g., exercise, medication), these diverse interventions largely have similar effects. Regarding hypertension as a characteristic with causal health consequences allows us to explore different approaches to intervening upon the condition, with a reasonable expectation that, if successful, any of them might have health benefits.

As our biological understanding deepens, we recognize that some physiological states we previously thought were clearly defined constructs are in fact heterogeneous. An illustration is the emergence of the distinction between high-density lipoprotein and low-density lipoprotein cholesterol from the earlier total cholesterol measure. There is no hope of completely eliminating ambiguity from our language; we aim only to be sufficiently clear to enable scientific progress.

CAUSAL INFERENCE FOR SOCIAL VARIABLES

Counterfactual frameworks and similar arguments about the importance of the consistency assumption[21] apply to social variables, such as education,[22] or even macro-level social processes.[23] If we had compelling observational research demonstrating that completing college causally improves later health, any number of strategies could then be devised to improve college graduation rates. Of course, we must always consider the possibility that any proposed intervention may be "fat hand" with respect to education. For example, forcing American Indian children into boarding schools intended to extinguish cultural identity likely had numerous adverse health effects unrelated to schooling per se.[24]

Furthermore, many social variables may operate more like cholesterol than like hypertension, in that different elements of the variable have different health effects. For example, income from lottery winnings or windfall payouts may not have the same health consequences as other income sources.[25]

CONCLUSION: POINTS OF AGREEMENT

Our hope for this commentary was to clarify the importance of counterfactual frameworks and modern statistical methods for supporting causal inferences in public health research on disparities, social determinants of health, and other drivers of population patterns of health and disease. We consider the NMNC perspective extreme, mistaken, and, if accepted by applied researchers, likely to impede public health research. Nevertheless, there is substantial agreement. Neither the NMNC perspective nor the perspective articulated in this commentary, which embraces causal inference for race and sex, has claim to higher moral ground. Many leading advocates of both perspectives are vigorous proponents of social justice and believe, for example, that individual and structural racism play critical roles in racial inequalities in health.

We also agree that "the potential outcomes framework is the ideal tool to frame causal discussions about how to eliminate racial disparities,"[4(p678)] with the caveat that the framework's utility extends well beyond this more limited scope and that specifying conceivable interventions, even if those interventions are not humanly feasible, is often invaluable when defining a cause and drawing causal inferences.

The ongoing dissemination of causal inference thinking and methods has led to a paradigm shift in the regard with which well-formulated and well-conducted observational research is held. Researchers who focus on race, sex, or physiological characteristics will find many useful tools in causal inference. Adoption of these tools will strengthen research aimed at mitigating or eliminating health disparities. Future columns will discuss some of the tools and their applications for causal inference in public health research. AJPH

CONTRIBUTORS

M. M. Glymour wrote the first draft of the commentary. D. Spiegelman conceived of the commentary. Both of the authors revised and edited substantially for content.

ACKNOWLEDGMENTS

This commentary was jointly sponsored by the National Institutes of Health Office of the Director and the National Institute of Environmental Health Sciences (grant DP1 ES025459).

We gratefully acknowledge generous critical feedback from Sari Reisner, Tyler VanderWeele, Kerith Conron, Jay Kaufman, and James Robins.

REFERENCES

1. Keyes KM, Galea S. *Epidemiology Matters: A New Introduction to Methodological Foundations*. New York, NY: Oxford University Press; 2014.

2. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;Epub ahead of print.

3. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014; 25(4):473–484.

4. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674–680.

5. VanderWeele TJ. On causes, causal inference, and potential outcomes. *Int J Epidemiol*. In press.

6. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396): 945–960.

7. Glymour C. Statistics and causal inference—statistics and metaphysics. *J Am Stat Assoc*. 1986;81(396):964–966.

8. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, England: Cambridge University Press; 2009.

9. Kaufman J, Cooper R. Seeking causal explanations in social epidemiology. *Am J Epidemiol*. 1999;150(2):113–120.

10. Krieger N. A glossary for social epidemiology. *J Epidemiol Community Health*. 2001;55(10):693–700.

11. VanderWeele T. *Explanation in Causal Inference: Methods for Mediation and*

Interaction. New York, NY: Oxford University Press; 2015.

12. VanderWeele TJ, Hernán MA. Causal effects and natural laws: towards a conceptualization of causal counterfactuals for nonmanipulable exposures, with application to the effects of race and sex. In: Berzuini C, Dawid P, Bernardinelli L, eds. *Causality*. New York, NY: John Wiley & Sons; 2012:101–113.

13. Conron KJ, Landers SJ, Reisner SL, Sell RL. Sex and gender in the US health surveillance system: a call to action. *Am J Public Health*. 2014;104(6): 970–976.

14. Reisner SL, Poteat T, Keatley J, et al. Global health burden and needs of transgender populations: a review. *Lancet*. 2016;388(10042):412–436.

15. Williams DR, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. *Ann N Y Acad Sci*. 2010; 1186(1):69–101.

16. Cooper LA, Ortega AN, Ammerman AS, et al. Calling for a bold new vision of health disparities intervention research. *Am J Public Health*. 2015;105(suppl 3): S374–S376.

17. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes (Lond)*. 2008; 32(suppl 3):S8–S14.

18. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20(1):3–5.

19. Lim S, Meigs JB. Ectopic fat and cardiometabolic and vascular risk. *Int J Cardiol*. 2013;169(3):166–176.

20. Scheines R. The similarity of causal inference in experimental and non-experimental studies. *Philos Sci*. 2006; 72(5):927–940.

21. Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Curr Epidemiol Rep*. 2016;3(1):63–71.

22. Cohen AK, Syme SL. Education: a missed opportunity for public health intervention. *Am J Public Health*. 2013; 103(6):997–1001.

23. Nandi A, Sweet E, Kawachi I, Heymann J, Galea S. Associations between macrolevel economic factors and weight distributions in low- and middle-income countries: a multilevel analysis of 200 000 adults in 40 countries. *Am J Public Health*. 2014;104(2): e162–e171.

24. Adams DW. *Education for Extinction: American Indians and the Boarding School Experience, 1875–1928*. Lawrence, KS: University Press of Kansas; 1995.

25. Bruckner TA, Brown RA, Margerison-Zilko C. Positive income shocks and accidental deaths among Cherokee Indians: a natural experiment. *Int J Epidemiol*. 2011;40(4):1083–1090.