

# Race and Sex Are Causes

Clark Glymour<sup>a</sup> and Madelyn R. Glymour<sup>b</sup>

In this issue of *EPIDEMIOLOGY*, VanderWeele and Robinson<sup>1</sup> state:

“Part of the challenge of interpreting race coefficients causally is that, in the formal causal-inference literature, effects are often defined in terms of counterfactual or potential outcomes, which are in turn defined as the outcomes that would result under hypothetical interventions. <sup>10–23</sup> There are, however, no reasonable hypothetical interventions on race when race itself is the exposure. Here we attempt to provide a causal interpretation of race coefficients in regressions without defining potential outcomes for race itself.” “...the race coefficient in the regression could be interpreted as the racial inequality that would remain if the family and neighborhood SES distribution of the black population were set equal to that of the white population.”

“Essentially, we give a plausible causal interpretation of the race coefficient by considering how much a racial inequality could be eliminated by intervening on a different variable, namely socioeconomic status, which is more manipulable than race.”

A minor point we leave for last: the authors’ technical argument<sup>1</sup> depends on a specific hypothetical intervention that in their model is logically unnecessary for the estimation, through an intervention on SES, of the regression coefficient for race or the effect of race on an outcome.

More importantly, we find the motivation and discussion wrong footed in several ways: (1) the notion that the random substitution of SES values is a “reasonable hypothetical” alternative to changing people’s races; (2) the notion that race and sex are not manipulable in principle; and (3) the assumption that causal claims make sense only when they correspond to potential intervention effects.

It is reasonable to consider what the inequality in an outcome would be according to a model if SES were equally distributed across races, and it makes sense to note that, on the model, the result is the same as if all black people were made white. It is not sensible to claim or imply that randomly assigning white SES values to black people is a “reasonable hypothetical intervention.” Aspects of SES may be manipulated, but that will not give the same outcome as the “impossible” manipulation of race, and an intervention that encompasses all of SES is both vaguely defined and practically impossible. (For instance, black young people could be subsidized to provide the same income distribution as for white youths, but how could it be ensured that blacks complete the same distribution of university majors in the same proportions as whites? And yet education is decidedly an element of SES.) Indeed, many factors that might readily be accepted as possible social factors in epidemiologic outcomes—weight, social circle, education, or occupation, for example—are either practically or ethically difficult to effectively intervene upon or composed of enough interacting subfactors that an intervention upon social circle, for instance, would be extremely complicated to define. It therefore seems somewhat arbitrary that race should be disregarded as a cause, whereas other variables retain that status.

Submitted 4 February 2014; accepted 28 March 2014.

From the <sup>a</sup>Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA; and <sup>b</sup>Department of Computer Science, University of California, Los Angeles, CA. **Editors’ note:** Related articles appear on pages 473, 488, and 485.

Correspondence: Clark Glymour, Department of Philosophy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: cg09@andrew.cmu.edu.

Copyright © 2014 by Lippincott Williams & Wilkins

ISSN: 1044-3983/14/2504-0488

DOI: 10.1097/EDE.0000000000000122

Although it may be known that particular interventions are feasible, the general notion of a “reasonable hypothetical intervention” is vague, and attempts to defend the restriction of causal claims to cases where hypothetical interventions are conceivable have made it even more vague. The mass of the sun is a cause of the planetary orbits, but what is the reasonable intervention? Rubin<sup>2</sup> goes so far as to entertain an intervention that makes the sun vanish. At that point, we have no clear idea of the intended scope of “reasonable intervention.” Perhaps, it is intended that a “reasonable hypothetical intervention” in a system is any exogenous change consistent with the laws of physics, although we are not sure Rubin’s<sup>2</sup> example recognizes even that limitation and we are pretty sure the meaning would leave the Big Bang as a non-cause. With that sense of “reasonable hypothetical intervention,” race and sex should count as causes. Interventions shortly after fertilization to replace a Y chromosome with an X, or an X with a Y, are on the verge of being technically feasible and as far as we know, consistent with physical law. Insofar as race is biologically defined, changes in DNA to change race are likewise, so far as anyone knows, consistent with physical law. Insofar as it is culturally defined, it is on the same footing with SES so far as “reasonable hypothetical interventions” are concerned.

This leaves us in a somewhat unfortunate place: if causation must be defined by intervention, and interventions on race and the whole of SES are vague or impractical, how is one to frame discussions of causation as they relate to this and other vital issues? Rather than stretching and twisting the sense of “intervention” to try to make it conform with scientific and common usage about causation, we suggest that “causation” is not univocal. There is a counterfactual/interventionist notion of causation—of use when one is designing a public policy to intervene and solve a problem—and an historical, or more exactly, etiological notion—often of use when one is identifying a problem to solve. The etiological notion involves a series of causal steps, for each of which there may be counterfactuals with interventions, even though the terminal events are not related by a counterfactual with or without an intervention. A time series of counterfactuals with corresponding interventions may not result in a counterfactual/intervention relation between an arbitrary point in the series and the endpoint.

Consider sex: Susan did not get the job she applied for because the prejudiced employer took her to be a woman; she presented as a woman because she was raised as a girl; she was raised as a girl because she was biologically female; and so on. The causation is palpable—Susan’s sex caused her not to get the job she applied for. The counterfactual, if Susan were male and had applied for the job, she would have gotten it, suggests a vague, miraculous transformation of Susan into some unspecified male (maybe one with the same qualifications, provided Susan did not attend any all-female schools)—but it makes no literal sense as a practical intervention. Suppose, however, a past intervention to make Susan male, say one of her X chromosomes was to be changed to some Y in utero.

To make the counterfactual come out true, the intervention must be expanded to also bring it about that in the course of life as an adult male she applies for the job. Pretty much all of the world history that would interact with her in the course of her male life would have to be intervened upon to bring it about that she, as a male, applied for the job. That would be a remarkably prescient intervention indeed and certainly not a reasonable one. The counterfactual, if Susan had been made a male in utero, Susan would have gotten the job, is almost certainly not true. Etiological causation does not direct us to practical interventions—for that, we need to focus on other causes that are feasibly and ethically manipulable. But it can provide us with a rationale for wanting to change outcomes: Susan did not get the job because of a biological fact about her that is irrelevant to her qualifications, and we think that is unjust.

The authors’ technical point provides a sufficient but not necessary condition for an intervention on SES that estimates both racial inequality in an outcome and the partial regression coefficient on race. To take the simplest model, let  $Y = b_0 + b_1R + b_2SES + e$ , with  $e$  distributed in the population with zero mean and independent of  $R$  (race) and  $SES$ .  $R$  is binary, 0 represents white race and 1 represents black race.  $SES$  is family socioeconomic status at time of birth. Then, on a simple model corresponding to the authors’ Figure 3,

$$E(Y|R=1, ses) - E(Y|set(R=0), ses) = \beta_1 + \beta_2 ses - \beta_2 ses = \beta_1 \quad (1)$$

is the expected change in outcome  $Y$  for a black person were she (miraculously) to become a white person with her original  $SES$ . We take the authors’ claim to be

$$E(Y|R=1, set[SES|R=1 \sim SES|R=0]) - E(Y|R=0) = \beta_1 \quad (2)$$

where “ $\sim$ ” means “distributed as.” How Equation (2) is obtained from the causal model and Equation (1) is a little puzzling, because Equation (1) holds for any distribution of  $SES$ . One thought is it follows from the model that

$$E(Y|R=1) - E(Y|R=0) = \beta_1 + \beta_2(E(SES|R=1) - E(SES|R=0)) \quad (3)$$

Interventions on the  $SES$  distribution that make the expected difference in outcome by race equal  $\beta_1$  must set  $E(SES|R=1) = E(SES|R=0)$ . There are, however, lots of ways to make  $E(SES|R=1) = E(SES|R=0)$ . One could set the white  $SES$  distribution equal to the actual black distribution or set both conditional  $SES$  distributions to any distribution whatsoever, as long as they are the same. The authors’ technical conclusion is correct, but seems more specific than their formal model warrants.

## ACKNOWLEDGMENTS

We thank David Danks, Medellena Maria Glymour, and Peter Spirtes for valuable discussions.

## ABOUT THE AUTHORS

CLARK GLYMOUR is the Alumni University Professor at Carnegie Mellon University. His research was supported by the James S. McDonnell Foundation. MADELYN R. GLYMOUR

is a Research Assistant to Judea Pearl at the University of California, Los Angeles, CA.

## REFERENCES

1. VanderWeele T, Robinson WR. On causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25.
2. Rubin D. Comment: which ifs have causal answers. *J Am Statistical Assoc*. 1986;81:961–962.