SOCIAL EPIDEMIOLOGY (JM OAKES, SECTION EDITOR)



The Counterfactual Implications of Fundamental Cause Theory

Ashley I. Naimi¹

Published online: 27 February 2016

© Springer International Publishing AG 2016

Abstract Causal inference lies at the center of epidemiologic research. In social epidemiology, two separate approaches to framing cause-effect relations have been considered: the counterfactual (or potential outcomes) framework and the theory of fundamental causes. The relations between these two frameworks have not yet been articulated. In this paper, I review the counterfactual and fundamental cause frameworks, and show how they capture different notions of cause-effect relations. Additionally, I show how the counterfactual and fundamental cause frameworks can be integrated to provide a more rigorous treatment of causality in social epidemiology. In particular, I show how counterfactual quantities can be used to evaluate predictions that follow from fundamental cause theory, assess the relations between and roles of various social resources in a given health disparity, and generate evidence on the potential interventions to mitigate health disparities.

Keywords Causation · Fundamental cause · Counterfactual cause · Potential outcomes · Health disparities · Social stratification · Social epidemiology

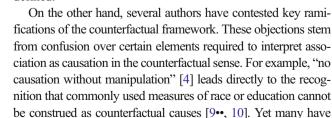
Introduction

As the science of public health, epidemiology has long been devoted to generating actionable knowledge to improve health

This article is part of the Topical Collection on Social Epidemiology

Ashley I. Naimi ashley.naimi@pitt.edu

Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, Parran 503, Pittsburgh, PA 15261, USA



outcomes [1]. Because of these pragmatic aims, causality is a permanent and prominent feature of the field. Several theoretic, analytic, and heuristic tools have been developed to structure thinking about cause-effect relations, analyze population health data, and interpret empirical results. These include the Bradford Hill criteria [2], the sufficient component cause model [3], the potential (or counterfactual) outcomes framework [4], and fundamental cause theory [5••].

These tools have different implications for what might be considered "causal," and what the nature of this causal relation might be. For example, depending on context, the Bradford Hill criteria might yield different conclusions about the nature of a suspected cause-effect relation that the sufficient component cause framework [6]. Similarly, the theory of fundamental causes might suggest that a particular association is causal, whereas the counterfactual framework might suggest otherwise.

Indeed, certain differences between the theory of fundamental causes and the counterfactual framework have led to an apparent causal impasse in social epidemiology. One the one hand, standard causal interpretations in social epidemiology have been weak. Associations from observational studies are often used to support claims that policy interventions to modify social determinants of health will improve population health and reduce health disparities [7, 8]. However, the precise connections between empirical analyses and proposed interventions are often nebulous, unarticulated, and ill-defined.



mis-interpreted this to imply that the constructs represented by variables such as race and education cannot cause health outcomes [11–14].

The purpose of this paper is to provide some perspective on causal inference in social epidemiology. I first describe the counterfactual (or potential outcomes) framework and highlight the reasons why social exposures such as race and education cannot be construed as counterfactual causes. I next review the theory of fundamental causes and note how the very attributes that preclude interpreting race or education as counterfactual causes may actually justify their characterization as fundamental causes. Finally, I outline how the counterfactual framework and fundamental cause theory can be integrated with much potential. In particular, I demonstrate how counterfactual quantities can be used to test predictions that follow from fundamental cause theory, assess the relations between and roles of various social resources in a given health disparity, and generate evidence on the potential interventions to mitigate health disparities.

The Counterfactual Framework

The potential (or counterfactual) outcomes framework is becoming increasingly popular in a wide array of empirical sciences, including epidemiology [15, 16]. The foremost strength of this framework is the clarity it offers in defining causal effects, particularly in the context of observational studies [17]. For a given outcome Y, exposure X, and individual i, $Y_i(x)$ represents the potential outcome for individual i that would be observed if (possibly contrary to fact) their exposure value was set to X=x. Using this formalization, it is easy to define a range of causal effects of interest, such as the causal risk difference:

$$P[Y_i(x) = 1] - P[Y_i(x^*) = 1]$$

This risk difference can be interpreted as what would be observed if the entire study population were exposed to x versus if the entire study population were exposed to x^* [18]. When x represents a characteristic with a clear and manipulable assignment mechanism, contemplating the risk that would be observed if the entire study population were exposed to x tends to evoke little controversy [19] (p452). Such is not the case for social exposures, including measures of socioeconomic position, gender, and race [10].

The causal status of such exposures has been heavily debated, particularly for race [4, 9••, 12–14, 20, 21]. Across the empirical sciences, race is most often quantified by questionnaire. Specifically, in responding to a question about one's "race," an individual (self or other) chooses from among several responses, which are tallied into different racial categories, such as "non-Hispanic black" (x = 1) or "non-Hispanic white" (x = 0). These commonly used racial and ethnic

classification schemes underlie much of the research on heath disparities. The debate centers on what to make of estimates representing quantities such as:

$$P[Y_i (x = 1) = 1].$$

In the counterfactual framework, when x=1 denotes non-Hispanic black, this quantity would be interpreted as the risk of Y that would be observed if everyone's race were set to "non-Hispanic black." Risk differences such as the one defined above would thus serve to answer research questions that query what would be observed if everyone in a given population were black versus if everyone were white. Such questions are not the same as questions pertaining to what would be observed if all individuals were exposed to the same levels of discrimination, stress, opportunity, or societal resources that are implicated in research on racial and social health disparities [9••, 22].

Commonly used racial/ethnic classification schemes capture long-established and systemic features of the social, historical, political, economic, and scientific systems from which sampled individuals arise [23–25]. For example, Feagin comments on the long-term effects of a single instance of systemic racial discrimination: the federal Homestead Act passed by the U.S. government in the 1860s, and which remained in effect until the 1930s. By this act, the U.S. government

provided about 246 million acres of land ... at low or no cost for about 1.5 million farm homesteads. Because of the extensive racial exclusion and violence directed at African Americans, including those recently freed from slavery, those who gained access to these wealthgenerating resources were almost entirely white. The homesteads of about 160–320 acres provided land resources on which many white families, including new European immigrant families, built up substantial wealth in the initial generation and subsequent generations [26] (p3).

Indeed, demographic projections suggest that roughly a quarter of the American population in the year 2000 were descendants of homestead recipients [27] and were thus beneficiaries of the wealth, privilege, and power generated by this government program.

As proxies of the past, modern racial classification schemes are a reflection of centuries-old institutional practices such as the Homestead Act that underlie the differential allocation of a nation's resources [26, 28, 29]. Thus, quantifying the risk of a health outcome that would be observed if everyone's racially patterned experiences (as captured by contemporary racial classification schemes) were set to what would have been had everyone been non-Hispanic white or black would, in effect, require intervening on the past and changing history. Race relations have deep and complicated historical roots [28,



30–33], and this complexity precludes interpreting contrasts of the risk of a particular health outcome among different racial groups as counterfactual causes. Interpreting race as a counterfactual cause would require asking what would be observed if all of the cultural, historical, and ancestral experiences of non-Hispanic black individuals would have been switched to what non-Hispanic white individuals experienced.

Said another way, the manner in which individuals become assigned (either by self- or other-report) to a given racial/ethnic category results from the complex interplay of long-standing social, historical, political, and economic processes. This complexity in the assignment mechanism leads to the violation of key assumptions required for causal inference and thus, the non-identifiability of the causal effect of race as a contrast of counterfactual outcomes [34, 35]. Similar problems are encountered for exposures such as education [10], gender [36], neighborhood effects [37], and other variables of common interest in social epidemiology.

However, while this complexity precludes interpreting social exposures as causes in the counterfactual framework, it does not imply that the constructs represented by variables such as race are not causal in a more general sense. For example, VanderWeele and Hernán note that it is possible to formulate causal statements that cannot be achieved by manipulation [36] (p106). Additionally, Paul Holland, the originator of the phrase "no causation without manipulation" also suggests that, in a very particular sense, race is not a cause [22] (p101). Moreover, contrasts of the risk of a health outcome between racial/ethnic groups can validly be interpreted as quantitative expressions of long-standing race relations [11, 22, 38] and are thus not without meaning.

Fundamental Cause Theory

A gamut of health outcomes have long been socially patterned [39]. Prior to the epidemiologic transition, mortality in the West was largely attributable to infectious diseases. This infectious disease mortality was concentrated in communities of low socioeconomic status minority populations with limited resources. When this infectious etiology was overtaken by death due to chronic disease, the social patterns in the distribution of mortality remained. That is, despite the dramatic etiologic change, minority populations with limited resources continued to bear the brunt of the excess mortality. The theory of fundamental causes was developed to explain this persistence [40••].

Fundamental causes are characterized by four key features [5••, 40••, 41••]: (i) they are related to numerous health outcomes, (ii) they affect health outcomes through multiple potential pathways, (iii) they involve access to resources that can be used to avoid risks or mitigate the negative impact of disease, and (iv) their effects persist over time via the

replacement of intervening mechanisms. Both race [41••] and socioeconomic status [5.., 40..] have been postulated as fundamental causes. High-SES individuals and privileged racial groups have access to a variety of flexible material and social resources (knowledge, money, power, prestige, and beneficial social connections) that enable them to maintain lower risks of death and disease, irrespective of the underlying etiology driving these outcomes [42]. As noted by Phelan, Link, and Tehranifar [40..], (pS29) "[i]t is their capacity to be used flexibly by individuals and groups that places [these] resources at the center of fundamental cause theory." Furthermore, these flexible material and social resources and their precursors have historically been allocated to individuals in positions of power and privilege (e.g., high-SES individuals of a given racial category). Thus, in some respect, the very reasons for which race and socioeconomic status cannot be considered counterfactual causes are why they are causal in the "fundamental" sense.

Several reviews have documented evidence on the features that characterize fundamental causes [40••, 41••, 42]. The unique policy implications of the fundamental cause theory have also been articulated [40••, 43]. In particular, the theory predicts that policies that do not depend on material and social resources are more likely to reduce health disparities. For example, exhorting women to consume more green leafy vegetables prior to and during gestation will likely lead to a different pattern in the disparity of neural tube defects than fortifying grains, rice, and cereals with folic acid. This feature provides a means by which researchers can empirically assess fundamental cause theory predictions and may facilitate prioritizing policy and interventions to reduce health disparities.

However, the theory does not provide guidance on how flexible resources should be modified to reduce health disparities and improve population health. Phelan and Link [40••] (p312) argue that "health inequalities resulting from a fundamental cause cannot be eradicated by addressing intervening mechanisms." Rather "[t]he the long-term impact of the fundamental cause can only be eliminated by reducing inequalities in the flexible resources." Yet the distinction between strategies to address intervening mechanisms and strategies to reduce inequalities in flexible resources can be unclear. For example, money is a flexible resource that can be altered in several ways [10]. A policy to increase minimum wage might be construed as an intervention to reduce inequalities in flexible resources [40••], but can the same be said of a conditional cash transfer program that increases income by a similar amount?

Additionally, one noted limitation is that fundamental cause theory does not imply anything specific about the linkages between flexible resources [42]. Thus, based on fundamental cause theory alone, one could not determine whether reducing a socioeconomic health disparity would best be accomplished by modifying resources linked to, for example, money, prestige, or knowledge.



The Counterfactual Implications of Fundamental Cause Theory

Fundamental cause theory leads directly to predictions for questions such as "what would the magnitude of a health disparity be if a particular risk factor were modified in a given population." Such questions are counterfactual in that they relate to the risk of a health outcome among different racial or social groups that would be observed if a third variable was modified. Naimi et al. [44] termed these quantities counterfactual disparity measures (CDMs) and reviewed several techniques that can be used to (i) evaluate predictions from the theory of fundamental causes; (ii) assess linkages between flexible resources, and how they might contribute to a given disparity; and (iii) generate evidence on the potential of a variety of interventions to mitigate health disparities.

To frame this illustration, I will highlight the counterfactual implications of fundamental cause theory in a hypothetical scenario involving the racial disparity in breastfeeding practices. Breastfeeding has been shown to have several infant health benefits [45], and there is a strong racial disparity in breastfeeding practices in the U.S. [46]. Breastfeeding counseling is a component of prenatal care [47], and thus, one may ask whether and to what extent prenatal care contributes to the racial disparity in breastfeeding.

More precisely, one may ask what the racial disparity in breastfeeding practices would be if all women received prenatal care. Such a question can be quantified using counterfactual disparity measures [10, 44]. Letting Y be an indicator of whether a woman breastfed or not, X be an indicator of non-Hispanic black status or not, and M be an indicator of whether a woman participated in a prenatal care program, one can define the CDM on the difference scale as:

CDM
$$(m) = P [Y (m) | X = 1] - P [Y (m) | X = 0]$$

where P[Y(m)] is the probability of breastfeeding if a woman's prenatal care status was set to some value m. This quantity captures the extent to which the racial disparity will be affected by altering participation in prenatal care and can thus be used to generate evidence on the potential that prenatal care has to mitigate (or aggravate) health disparities [44].

An estimate of this quantity can also be used to evaluate predictions from the theory of fundamental causes. Consider that most prenatal care programs provide information and counseling on the benefits of breastfeeding [47]. Thus, the effectiveness of prenatal care on breastfeeding initiation/duration depends on the availability of a woman's educational, monetary, and social resources to act on this counseling. Because non-Hispanic black women tend to have fewer educational, monetary, and social resources [48], fundamental cause theory would predict that prenatal care will increase the racial disparity in breastfeeding practices [40••]. More formally, it follows from fundamental cause theory that:

$$RD < CDM (m = 1),$$

where RD = P ($Y \mid X = 1$) – P ($Y \mid X = 0$) is the actual risk difference in breastfeeding in the population comparing non-Hispanic black to non-Hispanic white women. Additionally, according to fundamental cause theory, the increased risk difference that would be observed if all women received prenatal care would likely be due to a greater increase in breastfeeding among women who were not classified as non-Hispanic black, rather than a lowering of the probability of breastfeeding among non-Hispanic black women.

Finally, because the effectiveness of prenatal care depends on a woman's existing set of flexible resources (as measured by, e.g., educational level), it follows that the disparity in breastfeeding if all women received prenatal care would be greater among those with lower educational levels relative to higher educational levels. Letting *Z* be an indicator of educational status (e.g., less than high school education versus high school or more), one could further test fundamental cause theory predictions by estimating the following counterfactual disparity measures:

CDM
$$(m \mid Z = 0) = P[Y(m) \mid X = 1, Z = 0] - P[Y(m) \mid X = 0, Z = 0]$$

CDM $(m \mid Z = 1) = P[Y(m) \mid X = 1, Z = 1] - P[Y(m) \mid X = 0, Z = 1]$

These quantities can be interpreted as:

- CDM $(m \mid Z = 0)$: The racial disparity in breastfeeding that would be observed if prenatal care was set to a given level (m) in the entire population among those with less than a high school education (Z = 0).
- CDM $(m \mid Z = 1)$: The racial disparity in breastfeeding that would be observed if prenatal care were set to a given level (m) in the entire population among those with high school or more (Z=1).

Because acting on the counseling provided in a prenatal care setting would be easier for those with higher educational



levels, fundamental cause theory would predict that CDM ($m=1 \mid Z=0$) > CDM ($m=1 \mid Z=1$), or that the racial disparity in breastfeeding if all women received prenatal care would be greater among those with less than a high school education relative to those with high school or more.

Finally, counterfactual disparity measures can be used to assess the relative importance and potential linkages between flexible resources underlying a health disparity. That is, it might be of interest to assess whether resources related to income or resources related to education played more of a role in shaping the breastfeeding disparity that would be observed under different prenatal care scenarios. This may be accomplished by estimating

CDM
$$(m \mid Z = 0, V = 0),$$
 CDM $(m \mid Z = 0, V = 1),$

$$CDM(m \mid Z = 1, V = 0),$$
 $CDM(m \mid Z = 0, V = 1),$

where V=0 and V=1 denote values of income deemed low and high, respectively. One would conclude that resources related to income are more relevant in shaping the disparity that would be observed under a given prenatal care scenario if, for example, [CDM $(m \mid Z = 1, V = 0) - CDM (m \mid Z = 0, V = 0)] > [CDM <math>(m \mid Z = 0, V = 1) - CDM (m \mid Z = 0, V = 0)]$.

Methods to estimate these quantities have been illustrated in previous research [44].

Conclusion

Because of its practical orientation, causal inference lies at the center of epidemiologic research. The counterfactual framework and the theory of fundamental causes have both contributed to shaping how social epidemiologists think about cause-effect relations. While these distinct approaches frame causality in different ways, their combination may yield powerful insights on how to best mitigate health disparities.

Compliance with Ethical Standards

Conflict of Interest Ashley I. Naimi declares no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.



References

Papers of particular interest, published recently, have been highlighted as:

- Of major importance
- Rosen G. A history of public health. Baltimore: Johns Hopkins University Press; 1993.
- Hill A. The environment and disease: association or causation? Proc R Soc Med. 1965;58(5):295–300.
- 3. Rothman KJ. Causes. Am J Epidemiol. 1976;104(6):587-92.
- 4. Holland PW. Statistics and causal inference. J Am Stat Assoc. 1986;81(396):945–60.
- 5.•• Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;35(Extra Issue):80–94. The original paper introducing the notion that socioeconomic status acts as a fundamental cause of health outcomes, and outlining the implications of fundamental cause theory for health disparities.
- Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. Emerg Themes in Epidemiol. 2005;2(1):11.
- Auger N, Roncarolo F, Harper S. Increasing educational inequality in preterm birth in Quebec, Canada, 1981–2006. J Epidemiol Community Health. 2011;65(12):1091–6.
- Cutler D, Lleras-Muney A. Education and health: evaluating theories and evidence. In: House J, Schoeni R, Kaplan G, et al., editors.
 Making Americans healthier: social and economic policy as health policy. New York: Russell Sage; 2008.
- 9.•• Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. Am J Epidemiol. 1999;150(2):113–20. This article was the first to articulate the implications of the counterfactual framework for defining cause-effect relations in social epidemiology. The article emphasizes the difficulties encountered as a result of the non-random allocation of social exposures, but also comments on the problems that arise because of their "non-manipulable" nature.
- Naimi AI, Kaufman JS. Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. Curr Epidemiol Rep. 2015;2(1):52–60.
- Krieger N. Does racism harm health? Did child abuse exist before 1962? On explicit questions, critical science, and current controversies: an ecosocial perspective. Am J Public Health. 2003;93(2): 194–9.
- Bollen KA and Pearl J. Eight myths about causality and structural equation models. In: Morgan SL, eds. Handbook of causal analysis for social research. Springer; 2013. p. 301–328.
- Glymour C, Glymour MR. Commentary: race and sex are causes. Epidemiol. 2014;25(4):488–90.
- Marcellesi A. Is race a cause? In: Philosophy of science assoc. 23rd biennial Mtg. San Diego: PSA; 2012.
- Rubin DB. Causal inference using potential outcomes. J Am Stat Assoc. 2005;100(469):322–31.
- Sekhon J. The Neyman-Rubin model of causal inference and estimation via matching methods. In: Box-Steffensmeier JM, Brady HE, and Collier D, eds. The oxford handbook of political methodology. Oxford University Press; 2008.
- Robins JM, Greenland S. Causal inference without counterfactuals: comment. J Am Stat Assoc. 2000;95(450):431–5.
- Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60(7):578–86.
- VanderWeele TJ. Explanation in causal inference: methods for mediation and interaction. Oxford: Oxford University Press; 2015.

- Vanderweele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiol. 2014;25(4):473–84.
- Krieger N. On the causal interpretation of race. Epidemiol. 2014;25(6):937.
- Holland PW. Causation and race. In: Zuberi T, Bonilla-Silva E, editors. White logic, white method: racism and methodology. Plymouth: Rowman & Littlefield Publishers, Inc; 2008. p. 93–110.
- Krieger N. Refiguring "race": epidemiology, racialized biology, and biological expressions of race relations. Int J Health Services. 2000;30(1):211–6.
- Graves JL. The Emperor's new clothes: Biological theories of race at the millennium. New Brunswick: Rutgers University Press; 2001
- Krieger N. Stormy weather: race, gene expression, and the science of health disparities. Am J Public Health. 2005;95(12):2155–60.
- Feagin JR. Systemic racism: a theory of oppression. New York: Routledge; 2006.
- Williams T. The Homestead Act: Our earliest national asset policy.
 Center for social development, Washington University. Working Paper 00–9. St Louis; 2000
- Jacobson MF. Whiteness of a different color: European immigrants and the alchemy of race. Cambridge: Mass. Harvard University Press; 1998.
- Nobles M. Shades of citizenship: race and the census in modern politics. Stanford: Stanford University Press; 2000.
- Zinn H. A people's history of the United States: 1492–2001. New York: HarperCollins; 2003.
- Branch T. Parting the waters: America in the King years 1954–63.
 Simon & Schuster; 1988.
- 32. Branch T. Pillar of fire: America in the King years, 1963–65. Simon & Schuster; 1998.
- Branch T. At Canaan's edge: America in the King years, 1965–68.
 Simon & Schuster: 2006.
- VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. J Causal Inference. 2013;1(1):1–20.
- Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiol. 2011;22(3):368–77.
- 36. VanderWeele TJ and Hemán MA. Causal effects and natural laws: towards a conceptualization of causal counterfactuals for non-manipulable exposures, with applications to the effects of race and sex. In: Berzuini C, Dawid AP, and Bernardinelli L, eds. Causality: statistical perspectives and applications. John-Wiley & Sons; 2012: 101–112.

- VanderWeele TJ. Ignorability and stability assumptions in neighborhood effects research. Stat Med. 2008;27(11):1934

 –43.
- Zuberi T. Deracializing social statistics: problems in the quantification of race. Ann Am Acad Pol Soc Sci. 2000;568:172–85.
- Antonovsky A. Social class, life expectancy and overall mortality. Milbank Mem Fund Q. 1967;45(2):31–73.
- 40.•• Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav. 2010;51 Suppl 1:28–40. A review of the theory of fundamental causes and the empirical evidence supporting it, with a particular emphasis on the policy implications of the framework.
- 41.•• Phelan JC, Link BG. Is race a fundamental cause of inequalities in health? Annu Rev Sociol. 2015;41:311-30. This paper situates race in the fundamental cause framework. In particular, it outlines the relations between race, socioeconomic status, and health outcomes to show how race and racism generate inequities in power, prestige, freedom, neighborhood context, and health care, both with and independent of socioeconomic status, to shape health disparities.
- Glymour MM, Acendano M, I K. Socioeconomic status and health.
 In: Berkman LF, Kawachi I, Glymour MM, editors. Social epidemiology. New York: Oxford University Press; 2014. p. 17–62.
- Carpiano RM, Link BG, Phelan JC. Social inequality and health: future directions for the fundamental cause explanation. In: Lareau A, Conley D, editors. Social class: how does it work? New York: Russell Sage; 2008. p. 232–63.
- Naimi AI, Schnitzer ME, Moodie EEM, et al. Mediation analysis for health disparities research. Am J Epidemiol. 2015; In Press.
- Gertosio C, Meazza C, Pagani S, et al. Breast feeding: gamut of benefits Minerva Pediatr. 2015: Published Ahead of Print, May 29, 2015
- Allen JA, Li R, Scanlon KS, et al. Progress in increasing breastfeeding and reducing racial/ethnic differences—United States, 2000–2008 births. MMWR Morb Mortal Wkly Rep. 2013;62(5):77–80.
- Lind JN, Ahluwalia IB, Perrine CG, et al. Prenatal breastfeeding counseling—Pregnancy Risk Assessment Monitoring System, United States, 2010. MMWR Surveill Summ. 2014;63(2):14–9.
- Williams DR, Mohammed SA, Leavell J, et al. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. Ann N Y Acad Sci. 2010;1186:69–101.

