Curr Epidemiol Rep (2015) 2:52–60
DOI 10.1007/s40471-014-0030-4

SOCIAL EPIDEMIOLOGY (JM OAKES, SECTION EDITOR)

Counterfactual Theory in Social Epidemiology: Reconciling Analysis and Action for the Social Determinants of Health

Ashley I. Naimi · Jay S. Kaufman

Published online: 27 January 2015

Abstract There is a strong and growing interest in applying formal methods for causal inference with observational data in social epidemiology. A number of challenges in defining, identifying, and estimating counterfactual-based causal effects have been especially problematic in social epidemiology, particularly for commonly used exposures such as race, education, occupation, or socioeconomic position. The purpose of this article is to revisit these challenges in light of the conceptual and analytic advancements in causal inference over the last two decades. We focus on a central assumption for causal inference known as the stable unit treatment value assumption, which can be divided into two component assumptions: counterfactual consistency and the absence of interference. We give simple hypothetical examples to illustrate how and why these assumptions are often violated in research on the social determinants of health (e.g., education, race/ethnicity, socioeconomic position) and provide strategies that can be used to sidestep these assumptions. In particular, we note that a recently proposed mediation analysis strategy can be used to explore questions about health disparities in a more formal causal inference framework. We emphasize that a central obstacle to estimating causal effects variables such as race, education (e.g., high school versus no high school), or occupation is the need to identify an intervention (possibly hypothetical) that will lead to changes in the exposure of interest.

This article is part of the Topical Collection on *Social Epidemiology*

A. I. Naimi (✉)
Department of Obstetrics and Gynecology, McGill University, 687 Pine Ave West, Room F 432, Montreal, QC H3A 1A1, Canada
e-mail: ashley.naimi@mcgill.ca

J. S. Kaufman
Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

Keywords Epidemiology · Causality · Social determinants of health · Social class · Counterfactual consistency · Interference

Epidemiology is the study of the distribution and determinants of disease in the population. Because it is the science of public health, the epidemiologist's end goal is to change something in the world so as to improve population health [1, 2]. Social epidemiology is the study of how social factors, broadly writ, are related to health and disease. Many social epidemiologists seek to improve the health and well-being of all individuals, with a particular emphasis on the deprived and disenfranchised, usually defined with reference to some characteristic that classifies individuals as belonging to a certain group, such as gender, race/ethnicity, or social rank. Social rank or position is often defined as a function of education, income, and occupation, each of which can be measured at a fixed point in time during the study time scale (e.g., baseline), or repeatedly over a study's follow-up.

Observational research findings in social epidemiology have been traditionally used to reason about the effects of interventions aimed at reducing the impact of deleterious social exposures, or increasing the impact of beneficial ones. Recently, this practice has led to some disquieting observations [3••]. Rigorous intervention trials seeking to estimate the effects of policies to improve social determinants of health in epidemiology [4–6] as well as decades of work seeking to reduce health disparities [7–9] have shown results that diverge from expectations set by observational analyses. This has led to much debate on how to analyze data in social epidemiology [3••, 10–14]. From this debate, several routes have been identified as a way forward. Among them include the analysis of observational data using a formal approach outlined in the field of causal inference [11–15].

Since the late 1990s, several authors examined the implications of causal inference concepts in social epidemiology [16–18]. A number of challenges in defining, identifying, and estimating counterfactual-based causal effects were found to be especially problematic in social epidemiology. The purpose of this article is to revisit these issues in light of the conceptual and analytic advancements in causal inference over the last two decades. We focus on a central assumption in causal inference known as the stable unit treatment value assumption, which can be divided into two component assumptions: counterfactual consistency and the absence of interference. We give simple hypothetical examples to illustrate how and why these assumptions are often violated in research on the social determinants of health (e.g., education, race/ethnicity, socioeconomic position) and provide strategies that can be used to sidestep these assumptions. In particular, we note that a recently proposed mediation analysis strategy can be used to explore questions about health disparities in a more formal causal inference framework [19, 20]. We emphasize that a central obstacle to estimating causal effects of commonly used variables such as race, education (e.g., high school versus no high school), or occupation is the need to identify an intervention (possibly hypothetical) that will lead to changes in the exposure of interest [3••, 21].

Inferring Causal Effects from Observational Data

In this review, we consider as an example the educational disparity in preterm birth (defined as less than 37 weeks of completed gestation). A number of studies have examined the relation between maternal education and preterm birth [22–29], all of which found that low education is associated with an increased risk of preterm birth. Some interpreted their results to suggest public health interventions to increase maternal education would yield reductions in the risk of preterm birth [22, 24, 28]. However, while increasing levels of maternal education will undoubtedly have numerous social benefits, several problems arise when making inferences about the effects of public health interventions aimed at increasing maternal education.

To frame the specifics of our discussion, we introduce the hypothetical dataset of 15 subjects presented in Table 1. For didactic purposes, assume these are surveillance data acquired from birth certificates in a well-defined geographic region using gold standard measurement tools. Birth certificates often contain measures of gestational age and maternal education (e.g., number of years attained) at the time of birth. These two *measured and observable* variables are represented in columns 2 and 3 of Table 1, where: X is an indicator of whether the woman had a high-school education or more (X=1) or less than high school (X=0) at the time of birth; and Y is an indicator of whether the woman gave birth preterm (Y=1). As we explain in more details below: column 3 represents an *unmeasured but observable* variable denoting the mechanism by which each woman attained their observed level of education, and we let columns 4 and 5 represent the *unobservable* potential outcomes under exposure $Y(x=1)$ and no exposure $Y(x=0)$.

Table 1 Hypothetical dataset of 15 subjects with an observed exposure level of maternal education (X), an observed preterm birth outcome (Y), an observable but unmeasured assignment mechanism variable (J_x), and the potential outcomes that would have been observed under exposure [$Y(x=1)$] and no exposure [$Y(x=0)$]

ID	X	Y	J_x	$Y(x=1)$	$Y(x=0)$
1	1	1	0	0	1
2	1	0	1	0	1
3	0	1	1	1	1
4	0	0	0	0	0
5	0	0	0	1	0
6	0	0	1	1	1
7	1	1	0	0	0
8	1	0	0	1	0
9	0	0	1	0	1
10	1	0	1	0	1
11	1	1	0	0	1
12	0	0	1	1	1
13	0	1	0	0	1
14	0	0	0	0	0
15	1	1	1	1	0

SUTVA: Counterfactual Consistency

Potential outcomes, such as those listed in Table 1, have become central to defining causal effects in epidemiology. Because our exposure is whether a woman had at least a high-school education (X=1), the potential outcome in column 4, $Y(x=1)$ might be interpreted as what would have been observed if a woman had attained this level of education. Furthermore, for the same woman, $Y(x=0)$ might represent what would have happened if she had not reached this educational threshold. However, we will show that these definitions require certain assumptions that are not met in our hypothetical example.

The potential outcome framework is also known as the Neyman-Rubin causal model [30], and it was Rubin who first recognized that use of this notation implicitly makes a "stable unit treatment value assumption" (SUTVA) or the assumption that [31]: (*i*) there is only one "version" of the exposure; and (*ii*) no subject's exposure can affect another subject's potential outcome. Regarding assumption (*i*), "version" is meant to

connote, for example, different ways in which a subject might get exposed to a particular level of X, which may result in different causal effects. Multiple versions of the exposure could lead to many possible versions of the potential outcome, which would void the use of this simple notation, and violate what later became known as the counterfactual consistency assumption.

Counterfactual consistency is an unverifiable assumption requiring a subject's potential outcome under the observed exposure value is indeed their observed outcome. This assumption is more likely to hold when the exposure corresponds to a well-defined intervention [32•, 33•, 34]. Counterfactual consistency allows researchers to link observed data collected for a given study to the potential outcomes. Specifically, if a mother's observed exposure value is "high school or more," counterfactual consistency requires that this mother's potential outcome $Y(x)$ under $x=1$ = "high school or more" is equivalent to the observed outcome for this woman in our data, denoted Y. For example, if we examine subject with $ID=2$ in Table 1, we note that this person's exposure was $x=1$ and that the potential outcome under exposure $Y(x=1)=1$ is equal to the observed outcome $Y=1$. This might suggest that counterfactual consistency in our hypothetical example of the causal effect of maternal education on preterm birth was met.

However, if we examine subject with $ID=1$ in Table 1, we note something different. This person's exposure was $x=1$, but their potential outcome $Y(x=1)=0$ is not equal to the observed outcome $Y=1$. Thus, counterfactual consistency is violated in our example. The reason is explained by referring to column 4 of Table 1, which provides information on the two possible mechanisms by which each woman was able to attain a high-school education or more. If there are multiple ways of acquiring a high-school education (and thus of getting the value $x=1$), each of which might have different impacts on the outcome, then—as is seen with subject $ID=1$ in Table 1—it is not logical to assume that the observed value Y corresponds to the potential outcome under the observed exposure.

VanderWeele [33•] was the first to use this reasoning to propose a refinement to the counterfactual consistency assumption. He argued that, in making the assumption, researchers additionally assume (often implicitly) that the various ways in which a particular subject may have attained a particular level of the exposure are irrelevant. He called this assumption treatment variation irrelevance. For example, in our hypothetical data, subject $ID=2$ may have acquired a high-school education or more because of, e.g., a governmental cash transfer conditional on graduating from high school. This mechanism, indexed in column 4 of Table 1, is denoted $J_x=1$. On the other hand, subject $ID=1$ graduated from high school because of some other mechanism, denoted $J_x=0$. Treatment variation irrelevance requires that, for woman i, $Y_i(x,J_x=1)=Y_i(x,J_x=0)=Y_i(x,\bullet)$, where the "•" denotes collapsing over all values of J_x. In words, this assumption requires that the *way* in which a particular mother acquired a certain level of education is irrelevant to the outcome. If treatment variation irrelevance is met, then consistency follows naturally as $Y=Y(X,\bullet)$, where Y is the observed outcome and X is the observed exposure [32•].

This over-simplified example is meant to provide intuition on the counterfactual consistency assumption and emphasize the importance of well-defined interventions that can alter the value of the exposure of interest in causal inference [35••]. At times, one can get around mild violations of treatment variation irrelevance using stochastic counterfactuals [36], which allows for the assumption that $Y_i(x, J_x = 1) =^d Y_i(x, J_x = 0) =^d Y_i(x, \bullet)$, where "$=^d$" stands for equal in distribution [33•]. This argument effectively assumes that the potential outcomes for a given exposure level attained by two different mechanisms may be different, but that they come from the same distribution. Thus, on average, one could expect the potential outcomes to be the same for different exposure mechanisms. However, for variables like maternal education that do not even approximate well-defined interventions, there may be countless ways in which a mother can attain a given level of education, each of which may have a dramatically different effect on the outcome of interest [37].

Our stated example is arguably a case in point. There is some evidence to suggest that early parental investment in children has a more dramatic impact on later health and well-being than incentives to increase skills at a later age [38–41]. Thus, if $J_x=0$ in our hypothetical example includes a subset of women who were able to acquire a high-school education due to some intervention that occurred early in childhood (e.g., high-quality early education program [42]), there would likely be systematic differences in the distribution of potential outcomes between women with $J_x=0$ and women who acquired a high-school education due to the conditional cash transfer, Pell grant, or scholarship (mechanism $J_x=1$), even though they may all have the same value of X.

Note that our focus on education is merely didactic. Other exposures of common interest in social epidemiology are subject to samilar counterfactual consistency problems. These include routinely used racial and ethnic classification measures [19, 43], measures of sex or gender [44], neighborhood-level variables [18], and socioeconomic status [45].

SUTVA: No Interference

A second component of the stable-unit treatment value assumption is that one person's exposure does not affect another person's potential outcome [46••]. Because of the nature of typical exposures of interest, interference poses several challenges for causal inference in social epidemiology. For instance, a subject's health outcome after receiving a voucher

to move to a more affluent neighborhood may depend on whether members of the subject's social network (e.g., neighbors) were also given vouchers [47]. Similarly, the effects of an educational intervention in one subject may "spillover" into other subjects if, for example, the exposed subject influences the health behavior of other subjects because of their exposure status [48]. Indeed, educational interventions have long been identified as potentially subject to interference [49, 50].

Unlike violations of counterfactual consistency, however, interference is not merely a nuisance that requires resolution [46••]. Rather, the presence of interference among units gives rise to different causal effects that may be of interest to social epidemiologists. Hudgens and Halloran [51] defined direct, indirect, total, and overall effects in the presence of interference. Illustrating these effects requires defining new potential outcomes, as presented in Table 2.

The hypothetical data in Table 2 can be thought of as coming from a trial to assess the effect of a comprehensive intervention that engaged women in several healthy behaviors during pregnancy. In particular, we let ID_i denote whether a woman came from one of five possible birthing classes in which the intervention took place. As in Table 1, Y denotes whether a woman experienced a preterm birth. We let X denote whether a woman was exposed to the intervention (X=1) or whether she was subject to the standard birthing class protocol (X=0). Thus, in contrast to Table 1, we assume X in Table 2 represents whether a subject was exposed to a well-defined educational intervention, and consistency is upheld.

The complexity caused by interference is due to how the potential outcomes are defined. When interference is present, one subject's potential outcome becomes a function of their own exposure, and the exposure from other subjects in their group. In columns 5 to 8 of Table 2, there are 2^2=4 potential outcomes, each defined as a function of a possible exposure allocation strategy:

Table 2 Hypothetical dataset of 10 subjects from five different groups with a group variable ID_i, subject identifier ID_j, an observed intervention variable (X) and preterm birth outcome (Y), and the unobserved potential outcomes $Y(\boldsymbol{x}_1)$ to $Y(\boldsymbol{x}_4)$ under four possible exposure allocation strategies

ID_i	ID_j	X	Y	$Y(\boldsymbol{x}_1)$	$Y(\boldsymbol{x}_2)$	$Y(\boldsymbol{x}_3)$	$Y(\boldsymbol{x}_4)$
1	1	0	1	0	0	1	1
1	2	1	1	1	1	1	1
2	1	0	0	1	0	0	0
2	2	1	0	1	1	0	1
3	1	0	1	0	0	1	1
3	2	1	0	0	1	0	0
4	1	1	0	0	0	1	0
4	2	0	1	0	1	0	0
5	1	0	0	0	0	0	1
5	2	1	1	1	0	1	0

$$\boldsymbol{x}^{(1)} = \{0,0\} \quad \boldsymbol{x}^{(2)} = \{1,0\}$$
$$\boldsymbol{x}^{(3)} = \{0,1\} \quad \boldsymbol{x}^{(4)} = \{1,1\}$$

For example, $\boldsymbol{x}^{(2)}$ denotes an exposure allocation strategy for a given group in which the first subject is exposed and the second is not. This corresponds to the realized exposure allocation strategy for group 4 in Table 2. To define the causal effect of such exposures, we require a measure that describes the outcome for subject j in group i that would have been observed had their exposure been set to x and had the exposure vector for all remaining subjects in group i been set to $\boldsymbol{x}_{i(j)}$. Here, the vector $\boldsymbol{x}_{i(j)}$ represents the exposure values for subjects in group i not including subject j. Because, in our example, there is only one other individual in each group, we refrain from using bold notation and use $x_{i(j)}$ instead. Following Hudgens and Halloran [51], we define

$$Y_{ij}\left(x_{i(j)}, x_{ij}\right)$$

to be a preterm birth indicator for subject j in group i that would be observed had their exposure value been set to x_{ij} and had the exposure for the remaining subject in group i been set to $x_{i(j)}$. For example, referring to Table 2, one potential outcome for subject 2 in group 3 can be written as: $Y_{32}(0,1)$, which corresponds to the outcome that would be observed if this subject had been exposed and the other subject in group 3 been unexposed. Because this corresponds to the observed exposure status for the group, under counterfactual consistency, the potential outcome $Y_{32}(\mathbf{x}_3) = Y_{32}(0,1)$ is equal to the observed outcome Y=0.

This notation enables us to define four different outcome measures for preterm birth in our hypothetical study. The *individual average potential outcome* is:

$$\overline{Y}_{ij}\left(x_{i(j)}, x\right) \equiv \sum_k Y\left(x_{i(j)} = x_k, x_{ij} = x\right) Pr\left(X_{i(j)} = x_k \middle| X_{ij} = x\right),$$

where $Pr(X_{i(j)}=x_k|X_{ij}=x)$ is the probability that subjects in group i other than subject j received a particular exposure x_k. For example, the individual average potential outcome for subject 2 in group 3 if they had been exposed is:

$$\overline{Y}_{32}\left(x_{3(2)}, 1\right) = Y\left(x_{3(2)} = 0, 1\right) Pr\left(X_{3(2)} = 0 \middle| X_{32} = 1\right) + Y\left(x_{3(2)} = 1, 1\right) Pr\left(X_{3(2)} = 1 \middle| X_{32} = 1\right).$$

Similarly, the *group average potential outcome* $\overline{Y}_i\left(x_{i(j)}, x\right) = \sum_{j=1}^{n_i} \overline{Y}_{ij}\left(x_{i(j)}, x\right)/n_i$, where n_i is the total number of subjects in group i. For example, the group 3 average potential outcome under exposure is: $\overline{Y}_3\left(x_{3(j)}, 1\right) = \left[\overline{Y}_{32}\left(x_{3(2)}, 1\right) + \overline{Y}_{31}\left(x_{3(1)}, 1\right)\right]/2$. Finally, we can average over all groups to get the *population average potential outcome*, as: $\overline{Y}\left(x_{i(j)}, x\right) = \sum_{i=1}^{N} \overline{Y}_i\left(x_{i(j)}, x\right)/N$, where N is

the total number of groups. Thus, the population average potential outcome under exposure for Table 2 is:

$$\overline{Y}(x_{i(j)},1) = \left[\overline{Y}_1(x_{1(j)},1)+\overline{Y}_2(x_{2(j)},1)+\overline{Y}_3(x_{3(j)},1)+\overline{Y}_4(x_{4(j)},1)+\overline{Y}_5(x_{5(j)},1)\right]/5.$$

We refer the reader to Box 1 for a summary and interpretation of these measures of occurrence.

Box 1: Measures of Occurrence Under Interference.

$\overline{Y}_{ij}(x_{i(j)},x)$	Individual average potential outcome. The potential outcome for individual j in group i, averaged over the set of exposure scenarios for other group members.
$\overline{Y}_i(x_{i(j)},x)$	Group average potential outcome. The mean of all individual average potential outcomes for group i.
$\overline{Y}(x_{i(j)},x)$	Population average potential outcome. The mean of all group average potential outcomes.

The counterfactual measures of occurrence outlined in Box 1 can be used to define four different causal effects when interference is present. Hudgens and Halloran [51] introduced causal effects in the presence of interference that consist of overall and total effects, and a decomposition of the total effect into direct and indirect components. Tchetgen Tchetgen and VanderWeele [46••] provide a technical review of these effects and derive finite-sample confidence interval estimators. VanderWeele and Tchetgen Tchetgen [52] provide alternative definitions of direct and indirect effects in the presence of interference that rely on a decomposition of the overall rather than the total effect. We summarize the effects presented by Hudgens and Halloran [51] here and provide some intuition on the exposure effects they capture. The *individual average direct effect* is defined as the effect of switching a subject's exposure status, averaged over all possible exposure states for other subjects in the group:

$$\overline{DE}_{ij}(\boldsymbol{x}_{i(j)},x_{ij}) \equiv \overline{Y}_{ij}(x_{i(j)},1)-\overline{Y}_{ij}(x_{i(j)},0)$$

In our example, this equation captures the average effect of the educational intervention on the subjects who received it. For instance, subject j's educational intervention may impact subject j's outcome because it induces a beneficial change in subject j's behavior. This effect excludes any average effects that may have occurred through other individuals in the group. For instance, subject j's educational intervention may impact subject j'''s outcome because the change in subject j's behavior also induced a change in subject j'''s behavior (where j and j' index different individuals). To capture this latter type of effect, one has the *individual average indirect effect*, defined as:

$$\overline{IE}_{ij}\left(\boldsymbol{x}_{i(j)},\boldsymbol{x}'_{i(j)}\right) \equiv \overline{Y}_{ij}(x_{i(j)},0)-\overline{Y}_{ij}\left(\boldsymbol{x}'_{i(j)},0\right).$$

This contrast would yield the causal effect of switching the other subject's exposure from, say, $x_{i(j)}=1$ to $x'_{i(j)}=0$. This contrast has also been termed "spillover" effects [47] because the intervention effects are indirectly transmitted to other individuals. Such indirect or spillover effects may be of particular interest to social epidemiologists because of the roles that peer influence and social interaction play in human societies [53–55]. We may also be interested in the *individual average total effect*, which combines both the direct and indirect effect as:

$$\overline{TE}_{ij}(\boldsymbol{x}_{i(j)},\boldsymbol{x}'_{i(j)}) \equiv \overline{Y}_{ij}(x_{i(j)},1)-\overline{Y}_{ij}(x'_{i(j)},0).$$

Finally, we may be interested in the *individual average overall effect*, defined as:

$$\overline{OE}_{ij}(\boldsymbol{x}_i,\boldsymbol{x}'_i) \equiv \overline{Y}_{ij}(\boldsymbol{x}_i)-\overline{Y}_{ij}(\boldsymbol{x}'_i),$$

which compares, for example, the effect of exposure assignment strategies $\boldsymbol{x}^{(2)}$ to $\boldsymbol{x}^{(4)}$, as defined above.

While the presence of interference leads to several interesting causal contrasts, it can also add a substantial degree of complexity to the definition of individual-level potential outcomes [46••]. This complexity depends in part on the size of the cluster. For a binary exposure, there are 2^k possible potential outcomes for a cluster of size k. For cluster sizes of greater than 2, definitions of individual average direct and indirect effects are more nuanced [56, 52], and statistical inference for these effects is more involved, but still possible [46••]. Finally, as noted by Sobel [47], ignoring interference might lead researchers to conclude that an exposure under study is beneficial, even when it is actually harmful for all individuals in a given group.

A Mediation Approach

Counterfactual consistency violations in social epidemiology are often due to how social exposures are defined. Because typical exposures in social epidemiology are difficult to construe as intervention-based exposures, the counterfactuals of interest are not well-defined [35••]. It is unclear how one might manipulate adult socioeconomic status (SES), especially when defined as a composite measure of education, income, and occupation. In particular, one can often identify several different ways of manipulating each component of SES. For example, income can be increased in one lump sum or via regular installments during a specific period. For those social exposures that do correspond to well-defined interventions, the assumption of no interference can be difficult to satisfy because of the nature of the connections between human beings in society. These challenges curtail our ability to define and estimate counterfactually based causal effects in social epidemiology. Yet, despite these challenges, both causal inference and counterfactual thinking remain central features of social epidemiology [16, 57•, 58].

One potential strategy to deal with these issues was introduced recently in a conceptual paper by VanderWeele and Robinson [19] and implemented in a study seeking to assess strategies to mitigate educational disparities in preterm birth [20]. The general objective of this work is to facilitate the process of addressing questions about health disparities in a formal counterfactual framework. The basic logic of the approach follows from a few commonly accepted principles in social epidemiology:

1. Social exposures such as race, income, education, or occupation, or commonly used measures of socioeconomic position act as "fundamental causes" of a range of health outcomes [59, 60] and thus affect a host of potential factors that subsequently affect the health outcome under study. These factors are often identified as potential mediators of the relation between the social exposure and health outcome of interest.
2. A subset of these potential mediators may be variables that correspond more easily to well-defined interventions and that better satisfy the no interference assumption.
3. In spite of the violation of counterfactual consistency and no interference, *associations* based on social exposures are critical to understanding how health outcomes are distributed across populations and can contain decisive information on the severity or trajectory of a health disparity in a given population.

It follows from these observations that a formal causal mediation analysis [61] can be used to assess the magnitude of the disparity that would remain under an intervention (possibly hypothetical) on a modifiable mediator of interest, thus enabling a reconciliation of analysis and action from within the counterfactual framework in social epidemiology.

One previously published example of this approach involves the relation between education and preterm birth. Low levels of maternal education have consistently been associated with an elevated risk of preterm birth in a range of settings [25–28, 62]. Some have speculated that increasing the duration between births in a given woman may reduce the social disparities in preterm birth [63, 64], since several markers of social status—including race and education—are also associated with short birth intervals. Naimi et al. used a causal mediation framework to formally address this question [20] and estimate a parameter that corresponds to the magnitude of the educational disparity in preterm birth that would be observed under hypothetical interventions to alter birth intervals in the population. They defined a contrast:

$$\psi = \mathbb{E}\left[Y(m)\middle|X=1\right] - \mathbb{E}\left[Y(m)\middle|X=0\right]$$

where $X \in \{0,1\}$ denotes whether a given woman had a high-school education or more, M denotes a birth interval variable,

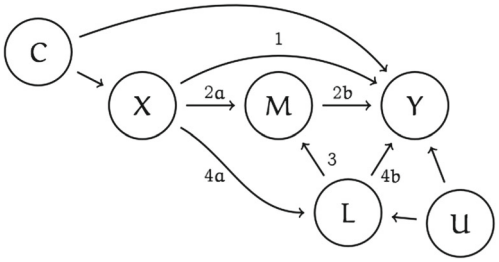

Fig. 1 Diagram representing the relations between exposure (X), mediator (M), and outcome (Y), as well as the exposure-outcome confounders (C) and mediator-outcome confounders affected by the exposure (L). This scenario commonly encountered in the causal mediation literature can be used to address questions in social epidemiology in a formal counterfactual framework using causal mediation analysis tools, when the exposure is a social determinant of health (e.g., race, education, income, occupation)

and $Y(m)$ is the preterm birth status that would be observed under an intervention that sets birth intervals of all women in the population to a value m. Thus, referring to Fig. 1, ψ corresponds to the risk difference for the X–Y relation that would be observed under an intervention on M, thus blocking the association between X and Y that occurs through paths *2a* and *2b*, as well as paths *4a*, 3, and *2b*. The approach can also be modified to estimate the risk difference for a disparity that would be observed under a more realistic intervention on the mediator in which only a subset of the population is affected by the hypothetical intervention of potential interest [20].

Discussion: The First Causal Inference Problem

Understanding cause–effect relations is arguably one of the most important objectives of epidemiologic research. Yet, using observational data to do so is subject to well known difficulties. Two challenges that are most commonly identified in social epidemiology are unmeasured confounding and reverse causation [e.g., 58]. While they are challenging issues, the first challenge when dealing with observational data to estimate counterfactually defined causal effects is the non-manipulable exposure problem [19, 45, 65]. In the counterfactual framework, confounding is defined as the presence of statistical dependence between the exposure and the potential outcomes [66••, §7.4]. Thus, concepts of confounding are logically dependent on the clearly defined potential outcomes, which depend on the ability to identify an intervention that can lead to changes in the exposure. Moreover, reverse or reciprocal causation, sometimes misleadingly referred to as "simultaneity" [58, 67], refers to a process in which the exposure and outcome are related in a time-dependent feedback loop. Such circumstances commonly arise in epidemiologic research with longitudinal data

[68]. These circumstances can sometimes be dealt with analytically [69], but require at least being able to clearly define the potential outcomes as a function of the exposure of interest. Hence, the emphasis on the need to identify (possibly hypothetical) interventions that can lead to changes in the exposure [3••, 35••] and the difficulty of estimating the causal effects of non-manipulable exposures such as race/ethnicity or sex/gender [65].

Characterizing social exposures (such as, e.g., race/ethnicity) as non-manipulable has not been without its critics. For example, Krieger has argued that considering race as a non-manipulable exposure "in effect relegate[s] 'race' to an intrinsic trait" [70, p196] and relies on the "unsubstantiated claim that 'race' is an a priori innate biological (i.e., genetic) property of individuals" [71, p937]. Such "essentialist" notions of race/ethnicity are indeed problematic in medical research [72], but they are not implied by counterfactual concepts such as non-manipulability. The causal inference literature defines a non-manipulable parameter as one that cannot be estimated in a randomized trial with (possibly sequential) exposure assignments [73]. As it would be impossible to conduct a randomized trial in which individuals are assigned to commonly used racial/ethnic classifications, it follows that race is not a manipulable exposure.

The solution to this problem would be to identify variables of interest in social epidemiology that fit better into the paradigm of intervention-based effect estimation [3••]. Indeed, Krieger has pointed out [70] such variables can include several manipulable characteristics related (either directly or indirectly) to commonly used racial classifications, including discriminatory practices [74], social policies [75], or even behavioural practices [63]. Though no panacea, use of manipulable exposure variables would do much to resolve the confusion over the interpretation of results in social epidemiology [1, 12, 13, 14].

Decades of work in social epidemiology has now established that disadvantaged social groups—whether defined by racial/ethnic classifications, education, income, occupation, or other characteristics related to the social, political, or economic realms—are strongly associated with a host of adverse health outcomes. Causal inference theory has much to offer social epidemiologists in their pursuit of reducing health disparities and the overall burden of disease.

Compliance with Ethics Guidelines

Conflict of Interest Both AI Naimi and JS Kaufman declare no conflicts of interest.

Human and Animal Rights and Informed Consent All studies by the authors involving animal and/or human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013;178:1185–91.
2. Breslow L. Musings on sixty years in public health. Annu Rev Public Health. 1998;19:1–15.
3. •• Harper S, Strumpf EC. Social epidemiology: questionable answers and answerable questions. Epidemiology. 2012;23:795–8. *An excellent commentary on the challenges in translating empirical research findings in social epidemiology into policy.*
4. Osypuk T, Tchetgen Tchetgen E, Acevedo-Garcia D, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: results from a randomized trial. Arch Gen Psychiatry. 2012;69:1284–94.
5. Schmeiser MD. Expanding wallets and waistlines: the impact of family income on the BMI of women and men eligible for the Earned Income Tax Credit. Health Econ. 2009;18:1277–94.
6. Jacob BA, Ludwig J, Miller DL. The effects of housing and neighborhood conditions on child mortality. J Health Econ. 2013;32:195–206.
7. Marmot M. Fair society, healthy lives: the Marmot Review. Strategic review of health inequalities in England post-2010. URL:http://www.instituteofhealthequity.org/projects/fair-society-healthy-lives-the-marmot-review.
8. Mackenbach JP. Can we reduce health inequalities? An analysis of the English strategy (1997–2010). J Epidemiol Community Health. 2011;65:568–75.
9. Bambra C, Smith KE, Garthwaite K, Joyce KE, Hunter DJ. A labour of Sisyphus? Public policy and health inequalities research from the Black and Acheson Reports to the Marmot Review. J Epidemiol Community Health. 2011;65:399–406.
10. Kawachi I. Editorial: isn't all epidemiology social? Am J Epidemiol. 2013;178:841–2.
11. Galea S, Link BG. Six paths for the future of social epidemiology. Am J Epidemiol. 2013;178:843–9.
12. Oakes JM. Invited commentary: paths and pathologies of social epidemiology. Am J Epidemiol. 2013;178:850–1.
13. Muntaner C. Invited commentary: on the future of social epidemiology—a case for scientific realism. Am J Epidemiol. 2013;178:852–7.
14. Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: off-roading with social epidemiology—exploration, causation, translation. Am J Epidemiol. 2013;178:858–63.
15. Galea S, Link BG. Galea and link respond to "pathologies of social epidemiology", "social epidemiology and scientific realism", and "off-roading with social epidemiology". Am J Epidemiol. 2013;178:864.
16. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. Am J Epidemiol. 1999;150:113–20.
17. Kaufman JS, Kaufman S, Poole C. Causal inference from randomized trials in social epidemiology. Soc Sci Med. 2003;57:2397–409.
18. Oakes J. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med. 2004;58:1929–52.
19. Vanderweele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology. 2014;25:473–84.
20. Naimi AI, Moodie EE, Auger N, Kaufman JS. Stochastic mediation contrasts in epidemiologic research: interpregnancy interval and the

educational disparity in preterm birth. Am J Epidemiol. 2014;180: 436–45.
21. Kaufman JS, Harper S. Health equity: utopian and scientific. Prev Med. 2013;57:739–40.
22. Auger N, Abrahamowicz M, Park AL, Wynant W. Extreme maternal education and preterm birth: time-to-event analysis of age and nativity-dependent risks. Ann Epidemiol. 2013;23:1–6.
23. Kaufman J, MacLehose R, Torrone E, Savitz D. A flexible Bayesian hierarchical model of preterm birth risk among US Hispanic subgroups in relation to maternal nativity and education. BMC Med Res Methodol. 2011;11:51.
24. Auger N, Roncarolo F, Harper S. Increasing educational inequality in preterm birth in Quebec, Canada, 1981–2006. J Epidemiol Community Health. 2011;65:1091–6.
25. Petersen CB, Mortensen LH, Morgen CS, et al. Socio-economic inequality in preterm birth: a comparative study of the Nordic countries from 1981 to 2000. Paediatr Perinat Epidemiol. 2009;23:66–75.
26. Morgen CS, Bjork C, Andersen PK, Mortensen LH, Nybo Andersen AM. Socioeconomic position and the risk of preterm birth—a study within the Danish National Birth Cohort. Int J Epidemiol. 2008;37:1109–20.
27. Luo ZC, Wilkins R, Kramer MS, for the Fetal, and of the Canadian Perinatal Surveillance System IHSG. Effect of neighbourhood income and maternal education on birth outcomes: a population-based study. Can Med Assoc J. 2006;174:1415–20.
28. Reagan PB, Salsberry PJ. Race and ethnic differences in determinants of preterm birth in the USA: broadening the social context. Soc Sci Med. 2005;60:2217–28.
29. Parker JD, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. Ann Epidemiol. 1994;4:271–8.
30. Sekhon J. The Neyman–Rubin Model of causal inference and estimation via matching methods. In: Box-Steffensmeier JM, Brady HE, Collier D, editors. The Oxford handbook of political methodology. Oxford University Press; 2008. URL http://www.oxfordhandbooks.com/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-11.
31. Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat. 1978;6:34–58.
32.• Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009;20:3–5. *An accessible introduction to counterfactual consistency, conditions under which it is violated, and strategies on how to deal with certain circumstances under which it may be violated.*
33.• VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009;20:880–3. *An accessible re-framing of the counterfactual consistency assumption in terms of treatment variation irrelevance.*
34. Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? Epidemiology. 2010;21:872–5.
35.•• Hernán MA. Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite? Am J Epidemiol. 2005;162:618–20. *An excellent commentary on the how well-defined interventions relate to well defined potential outcomes, and their role in causal inference.*
36. Robins J, Greenland S. The probability of causation under a stochastic model for individual risk. Biometrics. 1989;45:1125–38.
37. Cutler D, Lleras-Muney A. Education and health: evaluating theories and evidence. In: House J, Schoeni R, Kaplan G, Pollack H, editors. Making Americans healthier: social and economic policy as health policy. New York: Russell Sage Foundation; 2008.
38. Heckman JJ, Stixrud J, Urzua S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J Labor Econ. 2006;24:411–82.
39. Cunha F, Heckman JJ. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. J Hum Resour. 2008;43:738–82.
40. Carneiro P, Crawford C, Goodman A. Impact of early cognitive and non-cognitive skills on later outcomes. http://discovery.ucl.ac.uk/16164/1/16164.pdf. Accessed Nov 2013, Centre for the Economics of Education. 2007.
41. Miyamoto K, Chevalier A. Education and health. In: OECD, Improving health and social cohesion through education. doi:10.1787/9789264086319-6-en. Accessed Nov 2013: OECD Publishing. 2010.
42. Campbell F, Conti G, Heckman JJ, et al. Early childhood investments substantially boost adult health. Science. 2014;343:1478–85.
43. Kaufman JS, Cooper RS. Commentary: considerations for use of racial/ethnic classification in etiologic research. Am J Epidemiol. 2001;154:291–8.
44. Rubin DB. Comment: which ifs have causal answers. J Am Stat Assoc. 1986;81:961–2.
45. Glymour C, Glymour MR. Commentary: race and sex are causes. Epidemiology. 2014;25:488–90.
46.•• Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. Stat Methods Med Res. 2012;21:55–75. *The most complete and comprehensive review of interference in causal inference to date.*
47. Sobel ME. What do randomized studies of housing mobility demonstrate? J Am Stat Assoc. 2006;101:1398–407.
48. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: a case study of causal inference for multilevel observational data. J Am Stat Assoc. 2006;101:901–10.
49. Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. Stat Sci. 1990;5:472–80.
50. Rosenbaum PR. Interference between units in randomized experiments. J Am Stat Assoc. 2007;102:191–200.
51. Hudgens MG, Halloran ME. Toward causal inference with interference. J Am Stat Assoc. 2008;103:832–42.
52. Vanderweele TJ, Tchetgen Tchetgen EJ. Effect partitioning under interference in two-stage randomized vaccine trials. Stat Probab Lett. 2011;81:861–9.
53. Blume L, Blume L, Durlauf S. Identifying social interactions: a review. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco: Jossey-Bass; 2006. p. 287–315. *chap. 12.*
54. Berkman LF, Kawachi I, Glymour MM, editors. Social epidemiology. 2nd ed. New York: Oxford University Press; 2014.
55. Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. J Am Stat Assoc. 2013;108:469–82.
56. VanderWeele TJ, Tchetgen Tchetgen EJ. Bounding the infectiousness effect in vaccine trials. Epidemiology. 2011;22:686–93.
57.• Krieger N. Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med. 1994;39:887–903. *Though not dealing with formal framework for causal inference using potential outcomes, this paper stands out in the literature on causal inference in social epidemiology. It provides several compelling arguments on how social and political circumstance shape the nature of causal questions in epidemiologic research.*
58. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: mechanisms and causal evidence. Ann N Y Acad Sci. 2010;1186: 56–68.
59. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;35:80–94.
60. 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):S28–40.
61. Vansteelandt S. Estimation of direct and indirect effects. In: Berzuini C, Dawid P, Bernardinelli L, editors. Causality: statistical

perspectives and applications. West Sussex: Wiley; 2012. p. 126–50. *chap. 11*.

62. Auger N, Gamache P, Adam-Smith J, Harper S. Relative and absolute disparities in preterm birth related to neighborhood education. Ann Epidemiol. 2011;21:481–8.
63. Hogue CJ, Menon R, Dunlop AL, Kramer MR. Racial disparities in preterm birth rates and short inter-pregnancy interval: an overview. Acta Obstet Gynecol Scand. 2011;90:1317–24.
64. Khoshnood B, Lee KS, Wall S, Hsieh HL, Mittendorf R. Short interpregnancy intervals and the risk of adverse birth outcomes among five racial/ethnic groups in the United States. Am J Epidemiol. 1998;148:798–805.
65. VanderWeele TJ, Hernán 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, Bernardinelli L, editors. Causality: statistical perspectives and applications. Chichester: Wiley; 2012. p. 101–12.
66.•• Hernán MA, Robins J. Causal inference. Forthcoming. Chapman/Hall, http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/. Accessed 12 Mar 2014. *A comprehensive and accessible treatment of causal inference in epidemiologic research.*
67. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. J Epidemiol Community Health. 2008;62:858–61.
68. Rothman KJ, Greenland S, Lash T. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer; 2008.
69. Robins JM, Greenland S, Hu FC. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. J Am Stat Assoc. 1999;94:687–700.
70. 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;93:194–9.
71. Krieger N. On the causal interpretation of race. Epidemiology. 2014;25.
72. Morning AJ. The nature of race how scientists think and teach about human difference. Berkeley: University of California Press; 2011.
73. Robins J, Richardson T. Alternative graphical causal models and the identification of direct effects. In: Keyes KM, Ornstein K, Shrout PE, editors. Causality and psychopathology: finding the determinants of disorders and their cures, chap. Alternative graphical causal models and the identification of direct effects. Oxford University Press. 2011;103–58.
74. Williams DR, Neighbors HW, Jackson JS. Racial/ethnic discrimination and health: findings from community studies. Am J Public Health. 2008;98:S29–37.
75. Almond D, Chay KY, Greenstone M. Civil rights, the war on poverty, and black-white convergence in infant mortality in the rural South and Mississippi. Tech. Rep., Massachusetts Institute of Technology, Department of Economics. 2006.