Curr Epidemiol Rep (2016) 3:63–71
DOI 10.1007/s40471-016-0069-5

SOCIAL EPIDEMIOLOGY (JM OAKES, SECTION EDITOR)

The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose Is Not a Rose

David H. Rehkopf[1] · M. Maria Glymour[2] · Theresa L. Osypuk[3]

Published online: 16 February 2016
© Springer International Publishing AG 2016

Abstract The assumption that exposures as measured in observational settings have clear and specific definitions underpins epidemiologic research and allows us to use observational data to predict outcomes in interventions. This leap between exposures as measured and exposures as intervened upon is typically supported by the consistency assumption. The consistency assumption has received extensive attention in risk factor epidemiology but relatively little emphasis in social epidemiology. However, violations of the consistency assumption may be especially important to consider when understanding how social and economic exposures influence health. Efforts to clarify the definitions of our exposures, thus bolstering the consistency assumption, will help guide interventions to improve population health and reduce health disparities. This article focuses on the consistency assumption as considered within social epidemiology. We explain how this assumption is articulated in the causal inference literature and give examples of how it might be violated for three common exposures in social epidemiology research: income, education, and neighborhood characteristics. We conclude that there is good reason to worry about consistency assumption violations in much of social epidemiology research. Theoretically motivated explorations of mechanisms along with empirical comparisons of research findings under alternative operationalizations of exposure can help identify consistency violations. We recommend that future social epidemiology studies be more explicit to name and discuss the consistency assumption when describing the exposure of interest, including reconciling disparate results in the literature.

Keywords Social class · Epidemiology · Causal inference · Potential outcomes · Bias

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

✉ David H. Rehkopf
drehkopf@stanford.edu

M. Maria Glymour
mglymour@epi.ucsf.edu

Theresa L. Osypuk
tosypuk@umn.edu

[1] Department of Medicine, Stanford University School of Medicine, 1070 Arastradero Road, Palo Alto, CA 94304, USA

[2] Department of Epidemiology and Biostatistics, University of California, San Francisco, School of Medicine, 550 16th Street, San Francisco, CA 94158, USA

[3] Division of Epidemiology and Community Health, University of Minnesota School of Public Health, West Bank Office Building, Suite, 300, 1300 S. Second Street, Minneapolis, MN 55454, USA

Rose is a rose is a rose is a rose. – Gertrude Stein [1]

Introduction

A single term may encompass a multitude of meanings. Ambiguity in the meaning of a construct can impede scientific research and the replication and translation of findings. In epidemiology, this challenge is acute when translating evidence from observational to intervention settings. To help make the leap from observational evidence to intervention design, epidemiologists typically invoke three assumptions: exchangeability, positivity, and consistency. Social epidemiology, addressing how social and economic determinants influence health and disease throughout the life course, has long struggled with the exchangeability (no confounding) assumption. The positivity assumption—that in all covariate strata

some individuals are treated while others are untreated—has also received important research attention from social epidemiologists. But, social epidemiologists have paid less attention to the consistency assumption, which entails that the exposure is defined with enough specificity that different variants of the exposure do not have different effects on the outcome. Both theoretical evidence and the limited available empirical evidence suggest that violations of the consistency assumption are plausible for social exposures. Attending to such violations may help us design more effective interventions.

Causal Criteria of Consistency

Moving from an observed association between two factors to understanding whether one factor actually *caused* the other is a common goal for epidemiology research. We adopt a counterfactual or potential outcomes approach to defining a cause as: if the *cause* did not occur, the chance of the *outcome* occurring would be different than if the cause did occur. We can more concisely describe this using potential outcomes notation where $\Pr(Y_{X=1})$ represents the probability distribution of Y if the variable X were set to the value 1. This is called a "potential" outcome because X may or may not take the value 1. We define causal effects as contrasts of such potential outcomes. For example, the risk difference measure of the effect of X on Y would be defined as the difference in the probability that Y was 1 if we set everyone in the population to $X=1$ versus if we set everyone in the population to $X=0$ (or any other specific value besides 1):

$$\Pr(Y_{X=1} = 1)\text{-}\Pr(Y_{X=0} = 1) \tag{1}$$

Causal inference relies on the process of replacing the potential outcomes in the causal parameter with observed outcomes that can be estimated from data. For example, we may estimate the risk difference by contrasting the probability that Y is 1 among those people in the population with $X=1$ versus the probability that Y is 1 among those people for whom $X=0$:

$$\Pr(Y = 1|X = 1)\text{-}\Pr(Y = 1|X = 0) \tag{2}$$

But, under what circumstances can we substitute observed outcomes, such as $\Pr(Y=1|X=0)$ for potential outcomes such as $\Pr(Y_{X=0}=1)$? It is impossible to observe both $Y_{X=0}$ and $Y_{X=1}$ for even one person, much less for everyone in the population, as demanded in Eq. 2. In the analysis of quantitative data, the core criteria for causal inference are exchangeability, positivity, and consistency. These criteria form a link between observable features of the data and potential outcomes that define causal effects. Most efforts for deducing causal estimates in epidemiology have been focused on whether or not there are violations to the criterion of exchangeability (i.e., no confounding), that those individuals receiving the treatment should be considered as exchangeable (with respect to potential outcomes) with those not receiving the treatment and vice versa. That is, they should be identical on average for characteristics that may influence the outcome except for the treatment itself. This is the advantage of randomized trials, since exchangeability is inherent to the study design by virtue of random assignment. Specifically, because of randomization, we believe we can modify Eq. 1 to estimate the distribution of the potential outcomes $\Pr(Y_{X=1}=1)$ for the whole sample based only on those people who were randomized to receive $X=1$ and similarly estimate the distribution of the potential outcomes $\Pr(Y_{X=0}=1)$ based only on those people who were randomized to receive $X=0$:

$$\Pr(Y_{X=1} = 1|X = 1)\text{-}\Pr(Y_{X=0} = 1|X = 0) \tag{3}$$

In observational studies, exchangeability is achieved via other design or analytic techniques, including most commonly through covariate adjustment in multiple regression, but also by restriction, matching, or weight-based approaches, all of which are intended to mimic randomization [2, 3]. But, note that in Eq. 3, we still have a potential outcome. We need another assumption to be able to directly estimate those values as in Eq. 4, and this assumption seems so obvious it is sometimes overlooked. We need to assume that $\Pr(Y_{X=1}=1|X=1)=\Pr(Y=1|X=1)$: among those people with $X=1$, their actual value of Y matches the value Y would take if we set $X=1$. The consistency assumption requires that there are no two "flavors" or versions of treatment such that $X=1$ under both versions but the potential outcome for Y would be different under the alternative versions. This modest looking assumption—which in recent causal inference literature has been labeled the consistency assumption—is often overlooked.

It is important to note that the consistency assumption we discuss here is only very indirectly related to the Bradford Hill Criterion of "consistency" [4], which invokes the view that if most studies find a similar result, the association is more likely to be causal than if different studies report different results [5]. We are also not focused here on a consistent statistical estimator, i.e., an estimator that moves closer to the truth in larger sample sizes [6].

Rather, we are focused on describing the consistency criterion that is fundamental to the potential outcomes approach to causal inference. The consistency assumption implies that an individual's potential outcome under his or her observed exposure history is the outcome that will actually be observed for that person. Based on an earlier conception [7•] restated by Pearl: "for all A and B, if A is true, then if B would have prevailed (counterfactually) had A been true, it must be true already." [8] Perhaps part of what has led to less attention to this criterion is that on its own it can seem obtuse and a bit

circular. As noted by Pearl, consistency can, depending on the exact causal framework used, be either presented as an assumption to support inference or as an axiom to define counterfactuals [8]. This perhaps offers another reason why consistency may be the most ignored of the assumptions for causal inference—it is invisible because it is so fundamental to causal inference itself. What consistency critically *implies* is that *the exposure specified in the analysis must have enough precision that any variation within the exposure specification would not result in a different outcome*.

Consider the effect of socioeconomic status (SES) on health. While many measures of SES exist, composite measures that combine a person's income, education, and occupational prestige are common. For descriptive social epidemiology, composite measures are sometimes advantageous for capturing a more general underlying construct that could be, for instance, compared across place and time, in particular when the variation in the components of that composite index may differ in ways that are not of substantive interest. However, for causal studies, in order to avoid violation of the consistency assumption, it is important that regardless of which component of SES is changed, the same effect would occur. Without analyzing it directly, we can consider the plausible validity of this assumption by turning to literature that has estimated separate associations between each SES component of income, education, and occupation, with a specific health outcome. Based on a large literature in this area, we conclude that for the construct of socioeconomic status as operationalized using a composite, the assumption of consistency is likely violated [9, 10]. Interventions on income, education, or occupation may each have different impacts on a particular health outcome. An SES index could be considered to be a "compound treatment," a common cause of violation of the consistency assumption [11]. While not invoking the consistency assumption, this has led others to the same conclusion that for etiologic work, it is best to avoid composite measures of SES [12].

Many interpretations of the consistency assumption require that we grapple with the manipulability criterion, a controversial topic in social epidemiology. This debate for social epidemiology was laid out in an influential commentary by Kaufman and Cooper "Seeking Causal Explanations in Social Epidemiology," [13] where they explored the difficulties of establishing a clear counterfactual when examining social factors' influences on health. They recommended that researchers "seek causal explanations only for definable interventions." Consistency is often motivated by linking to specific interventions rather than simply to specificity of exposure (e.g., by focusing on whether obesity is changed by diet or by exercise, rather than by focusing on whether obesity is due to excess central or peripheral adipose tissue) [14••]. Currently, researchers in social epidemiology have not reached consensus on the manipulability criterion. However, even if people believe that non-manipulable factors can be causes, the consistency assumption may be violated if an intervention to change the exposure is not clearly and specifically described. For example, if intervening on SES by increasing education has different health consequences than intervening on SES by increasing income, SES is not a well-defined intervention and violates the consistency criterion. Telling other researchers that increasing SES improves health is thus insufficient to guide the development of specific interventions.

Ambiguity in the definition of interventions to change exposure is central to violations of consistency [15•]. Cole and Frangakis define consistency in terms of the potential outcome definition of $Y_{j(x,k)}$, where for individual j the exposure X is set to a specific value x by intervention k. Therefore, the consistency assumption holds if the observed $Y_{j(x)} = Y_{j(x,k)}$ for all values of k that set X to x. As noted by VanderWeele [16•], the range of possible k (the means by which exposure occurs) will vary depending on the specific x (the specific exposure tested) [16•]. The task then within social epidemiology is for an investigator to consider the range of possible interventions (k) that could elicit the same value of exposure (x). Interventions need to be described with enough specificity that any additional variation not specified is irrelevant. However, if the intervention description is needlessly specific, it will preclude future replications.

A second related but distinct concern is effect measure modification. Effect measure modification occurs when the effect of an exposure of interest on an outcome is different in some subgroups of the population as compared to others—for example, if the effect of social support on heart disease differs between men and women. In many cases, it is ambiguous whether a variation should be considered a feature of the *population* or a feature of the *intervention*. For example, we might find that beginning school at ages 3–4 years has larger cognitive benefits than enrolling in school at ages 6–7 years. We could say there are two versions of the treatment "begin school at age 3" versus "begin school at age 6," or we could say there is one treatment "begin school," which could be applied to 3 or to 6 years old. Focusing on two interventions would frame this as a consistency problem; focusing on the age of children who receive a treatment frames this as an effect modification problem. The key distinction is whether the attribute violating the consistency assumption is more clearly thought of as a separate factor or a characteristic of the exposure itself. Very closely related is the challenge of transportability [11], or how causes identified in one population can or cannot be applied to other populations [17]. For practical purposes, consistency violations and effect measure modification have similar implications: to design an intervention you must decide precisely who will receive the intervention and what the content of the intervention will be. To be useful to intervention design, observational studies must therefore specify both the population and the exposure.

A third issue related to but distinct from consistency considers causal pathways, or mediation. Exposures that influence an outcome via multiple mediating pathways with different magnitudes of effect do not necessarily violate the consistency assumption. Multiple mediating pathways, however, may imply the potential for violations of consistency because it is easy to imagine several closely related interventions that trigger some but not all of the mediating pathways. For example, education is thought to influence health via improvements in knowledge and cognitive skills, credentials that are valued on the labor market, status improvements, and changes to the individual's social network. It is easy to imagine variations on educational experiences that have larger or smaller effects on just one of these mechanisms, e.g., cognitive skills or prestige. Indeed, articulating the theorized mechanisms linking exposure and outcome may help us define clear and specific exposures that fulfill the consistency criterion.

Potential Consistency Assumption Violations in Social Epidemiology

Although some social epidemiology research is directed towards adherence to the consistency assumption, it is usually framed in terms of the specificity of the exposure and generalizability, without formal linkages to the consistency assumption as a criterion for causal inferences. Thus, while not evaluated directly, in this section of the paper, we discuss examples of how the consistency assumption may be violated for research on three commonly studied factors in social epidemiology: income, education, and neighborhoods. We describe below examples of research in social epidemiology and related fields that through their specific counterfactual contrast examine a particular k for these three common exposures and provide some evidence that consistency violations are a concern in social epidemiology. For all of the exposures we describe, the possible k variations (the means by which exposure is enacted) based on timing and duration also apply.

Income

Multiple reviews and chapters on the relationship between income and health have noted the varied associations across studies; these reviews generally focus on the importance of duration and timing of exposures, the outcomes examined, and violations of exchangeability [18]. All three of these issues are important for defining specific causal effects of income on health, but some of the variability in effect estimates for the relationships between income and health may also be due to lack of attention to the consistency criteria. In examining income effects on health, several types of income changes have been evaluated. These include sources such as regular earnings [19], paychecks [20], tax refunds [21], gaming profit disbursements [22], cash transfers [23], inheritances [24], and lottery winnings [25]. For some of these differences in k, the distinction may be subtle. For example, regular earnings are typically disbursed as a paycheck, so earnings and paychecks could be considered effectively the same treatment with respect to the consistency assumption. Analyses, however, have differentiated within these exposures by using methods to examine earnings over periods of time of months or years, or focus on the period of time when the funds from a paycheck are actually received, resulting in a potentially different effect. It is not currently clear whether, for a particular health outcome, these various mechanisms of receiving income result in a different effect (Fig. 1).

While some of the literature is adherent to the consistency criterion by carefully framing results in reference to the exact type of income transfer examined, it is generally a mistake to extrapolate from a particular exposure to the effects of income more generally. For example, in describing the implications of the conditional cash transfer from the *Oportunidades* program, the authors wrote: "Our results suggest that the cash transfer component of *Oportunidades* is associated with better outcomes in child health, growth, and development" [23]. In contrast, when describing the goal of research examining the effect of lottery winnings on health, the authors wrote: "This paper has asked whether money makes individuals healthier" [26], a description of a causal estimate that has a wider range of k.

There are a few examples where multiple types of income exposures have been examined in the same study using similar methods, sometimes with identical outcomes and populations. This literature suggests that income from different sources has different health effects. One of the best examples of attention to the consistency criterion is work that stipulates a specific k of short-term duration and compared the effects of other multiple k sources of money (i.e., from social security payments, regular wage payments for the military, tax rebates, and Alaska fund payments) on short-term mortality [27]. The similarity of results across these k income sources suggests that for short-term effects of income, these sources effectively meet the consistency assumption. In the examination of short-term effects of tax credits on risk factors for mortality, however, impacts were primarily beneficial [21], suggesting that k could differ for the alternative exposure of consistent, large benefits from the Earned Income Tax Credit (EITC). Qualitative work can also play an important role in considering violations of the consistency assumption. Work on EITC recipients has shown that individuals view the lump sum tax refund payment in a different way than regular paychecks and that what they spend money on varies depending on the type of income received [28].

Fig. 1 56 million £ lottery winners Nigel Page and Justine Laycock. Does a dollar of lottery winnings affect their health in the same way as a dollar of income from work?

Education

The overwhelming majority of health research operationalizes education as either years completed (often in two or three categories, such as less than 12 versus 12 or more) or degree completion (often in two or three categories, such as a high school diploma or bachelor's degree). Although many researchers acknowledge the potential importance of school quality [29], aspects of quality such as variations in school term duration (which now averages around 180 days per year but has historically varied by nearly a factor of two, especially across segregated schools) are not captured in years completed or degrees obtained [30]. Differences in the timing of education with respect to developmental stage are also disregarded in conventional measures of education based on years or degrees completed. The literature suggests that the multiple attributes of education that are not typically specified may be differentially associated with health outcomes, thus use of standard measures of level of education may violate the consistency assumption. In contrast to years of education, Manly has conceptualized literacy—the capacity to use printed and written information to function in society—as a marker of school quality [31]. Teaching literacy skills is a key goal of education, and literacy is plausibly a powerful mediator for schooling effects. Manly finds that accounting for literacy substantially accounts for racial disparities in age-related dementia outcomes. She has also shown large regionally based disparities in literacy in older adults [32]. Literacy could be improved via multiple different treatments that are within the domain of education, such as higher quality schooling, longer duration of schooling, or mechanisms unrelated to formal schooling. However, if literacy mechanisms are essential to the effects of schooling, it implies that educational interventions focusing on improved literacy will have different health consequences than interventions that, for example, focus only on the number of years of education. Nguyen reports that literacy accounts for roughly 19 % of the effect of education on mortality, although this may be an underestimate due to the limited measure of literacy available [33]. Another approach to intervening within the overall construct of education is on class size. The Tennessee STAR (Student/Teacher Achievement Ratio) trial randomized children in 79 schools to small class sizes, with or without certified teacher's aides to assist children in the classroom [34]. Evidence on the health effects of these variations would help us understand whether consistency violations are a concern in typical definitions of education. To date, however, evidence on the health effects of Tennessee STAR is limited and mixed, although assignment to high-quality classrooms appeared to benefit earnings [35].

Research using compulsory schooling laws (CSLs) as natural experiments for the health effects of education has indirectly grappled with the consistency challenge, because CSL changes typically either introduce an extra year of schooling at an earlier age (e.g., reducing the school entry age from 7 to 6 years) or an extra year of schooling in adolescence (e.g., increasing the school leaving or work permit age from 14 to 15). Several instrumental variable studies based on CSLs find larger effect estimates than conventional analyses [36–39], and one interpretation is that delivering an additional year of schooling at these ages has larger consequences than the effect of an additional year of schooling at later developmental periods. To our knowledge, there is no direct empirical evaluation of this explanation for the discrepancy. However, several studies suggest that early interventions focused on improved

cognitive environments (e.g., among children under 5) have enduring health benefits [38, 40, 41]. These studies typically include both cognitive and non-cognitive (e.g., nutritional) enrichment for children, so it is difficult to disentangle strictly educational exposures from other benefits.

Figure 2 illustrates the challenge of identifying interventions that correspond with the relevant mechanisms via which education may influence health. Of the several possible interventions shown, few would be reflected in typical measurements of education (e.g., years of education completed) and each way of intervening might have a different consequence on the mediators and thus on health. Conversely, the consequences of attending extra years of schooling might depend quite strongly on how much classroom instruction occurred during each year of school, the quality of the teachers, the norms, talents, and socioeconomic resources of classmates, the prestige of the school, and so forth. In short, the link between what we measure in most observational studies of education, and what matters for health, is not necessarily close. As a result, we are unsure how to best intervene.

Neighborhoods

Studies examining how neighborhood context influences health are quite heterogeneous in terms of the specific neighborhood construct, as well as the operationalizations of specific constructs. The most common exposure among neighborhood effects studies is the SES of a neighborhood, which is typically used to measure the construct of underlying neighborhood quality. A common way to operationalize neighborhood SES is via an index of different dimensions of SES, e.g., median area income, proportion of neighborhood households below poverty line, proportion of residents without a high school education, and proportion of unemployed males of working age [42]. This exposure encounters the same threat of consistency violation described above as for individual SES indices. For example, if decreasing the proportion of people in poverty may have different effects than increasing the average high school graduation rate.

Another potential violation of consistency concerns the method by which the construct of neighborhood quality is changed. Two approaches are typically conceptualized as approaches to modifying neighborhood environments: (1) place-focused interventions, often called neighborhood revitalization policies, that improve the conditions in disadvantaged neighborhoods and (2) people-focused interventions that help households obtain housing in higher quality neighborhoods. This second approach improves neighborhood quality for individuals by virtue of a specific household that moves. Housing policy experts recommend that both place and people-based approaches are necessary to reduce neighborhood inequality [43], but specific interventions within each of these approach types may have different impacts on the health of neighborhoods. One of the strongest studies for assessing the changes of neighborhood and housing context for health is the Moving to Opportunity (MTO) study, where volunteer low income households were randomized to receive a government housing voucher to subsidize rental costs in private units, compared to in-place control groups in public housing. The MTO study demonstrated that receiving a rental voucher resulted in substantial improvements of neighborhood context and strong improvements in mental health for low-income predominantly minority female household heads and their daughters in the treatment group compared to controls. Treatment effects were stronger for those without health vulnerabilities at baseline [44, 45]. The randomized design often (but not necessarily) means that consistency is met because exposure is specific and assigned by the program: study

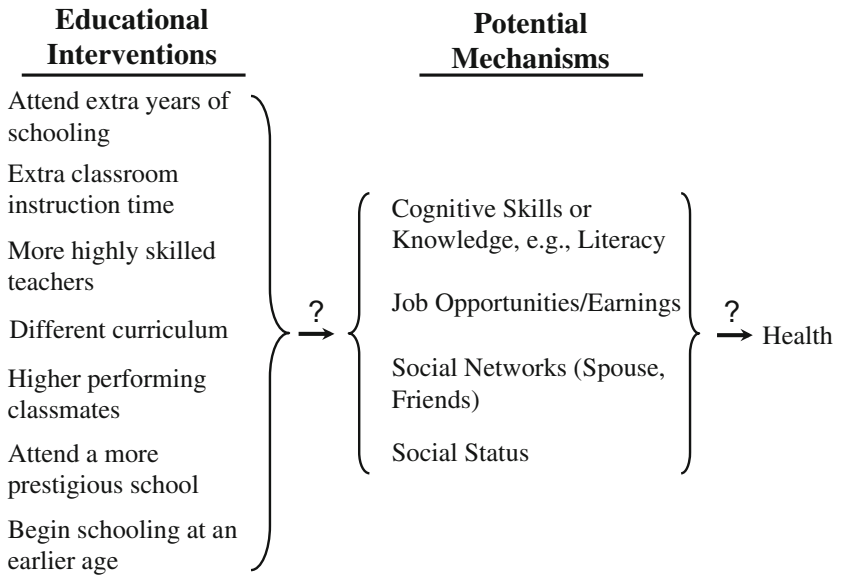

Fig. 2 Links between possible interventions on education and potential mechanisms affecting health. Not all interventions influence the same mechanisms, and it is unknown which mechanisms are most important for particular health outcomes

participants were assigned to a specific experimental group, with a well-defined treatment protocol in each arm. However, more typical of observational studies in the field, changes in neighborhood poverty exposure may be achieved via any of several mechanisms, including household moves unassisted by a mobility policy, neighborhood revitalization, or via neighborhood gentrification. One could easily imagine that the health consequences of moving from a high-poverty to a low-poverty neighborhood might differ from the health consequences of improvements to the current neighborhood. If these modes differ in their effects on health, this suggests a potential violation of consistency. Specifying whether neighborhood change is achieved by moving, as opposed to by place-based change, would improve inferences and provide better guidance to develop interventions.

While neighborhood revitalization is one key mechanism to improve neighborhood quality, there are many possible ways to operationalize it. Moreover, in practice, revitalization programs often implement multiple simultaneous changes with concentrated place-based investment, which influence multiple dimensions of a neighborhood (e.g., transportation, housing, economic development). This bundled treatment makes it difficult to meet the consistency assumption. Again, narrowing the exposure could help, if such improvements were implemented (or able to be evaluated) in isolation. For example, neighborhood redesign to adopt smart growth principles of street connectivity to improve walkability, or increasing greenspace and vegetation. However, much of this literature fails to manipulate an exposure or to isolate a specific exposure [46].

Notably, neither random assignment, nor manipulability, guarantees consistency because of ambiguities in the treatment protocols and bundled treatments. Many social and economic policies that have been rigorously designed and evaluated (i.e., via experiments) deliver bundled treatments [47]. Bundled treatments may be problematic for the consistency assumption if particular dimensions of the treatment affect health differently.

Is Consistency Always Important?

Ralph Waldo Emerson wrote “a foolish consistency is the hobgoblin of little minds” [48], and we agree that in some settings, the consistency assumption is less important. Particularly, early in the arc of a research question, it may be important to cast a wide net, examine unclearly defined constructs, and try to integrate evidence across studies with measures that do not clearly correspond to a specific intervention. Vandenbroucke et al. note that one challenge in recent causal inference literature is that one does not necessarily know in advance if an intervention is well-specified (i.e., consistent), because we do not know exactly how the intervention will work [49]. Just as research designs are often more vulnerable to confounding or violations of exchangeability when forging new scientific territory, initial research designs may also appropriately back-burner consistency with the goal of evaluating whether any component of a large set of related but somewhat ill-defined constructs influences the outcome. As a research area matures and researchers move towards efforts to intervene, however, the consistency assumption must be directly addressed to guide intervention development.

Conclusion

We have three recommendations for how social epidemiologists should address the consistency assumption in their work. First, studies should be explicit about the assumptions of their version of treatment or exposure (k) and draw on both theoretical and empirical evidence to identify potentially relevant variations. Second, until results come to light that suggest that certain dimensions of the version of treatment (k) do not violate the consistency assumption, studies should endeavor to examine exposures with a reduced range of such versions. Third, as VanderWeele suggests [16•], whenever possible, researchers should explicitly test multiple different definitions of an exposure. Although there is not a definitive test for the consistency assumption, there are ways to evaluate how plausible it is for a specific exposure definition by using data from both within and across studies.

The epidemiologic literature that speaks to the substantive magnitude of differences in effect estimates is also useful for thinking about how important it may be to pay closer attention to consistency. For example, quantitative comparison of studies has shown that there is a close match between trials and observational studies when there is a precisely defined exposure, such as a medical treatment [50]. When results of randomized trials diverge from observational evidence, it may be because studies did not use precisely the same exposure, for example, in studies of beta-carotene [51–53]. Although the exchangeability assumption is usually assumed to account for differences between RCTs and observational evidence, consistency assumption violations may also play a role.

Violations of the consistency assumption in social epidemiology may provide another explanation for why observational study results are often heterogeneous, in addition to differences deriving from the study populations or confounding structures. The call for consequential social epidemiology entails more attention to the consistency assumption [54, 55]. Fulfilling the consistency assumption allows for a closer connection between observational studies and inferences about actions based on those studies. In practice, this aligns with the in all policies approach [56, 57], a cross-sectoral view of how policies shaping social determinants of health outside of

the health sector (e.g., in the education, housing and/or workforce sectors) can be used to improve health.

Compliance with Ethical Standards

Conflict of Interest David H. Rehkopf, M. Maria Glymour, and Theresa L. Osypuk declare that they have 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.

Funding David H. Rehkopf was supported by the National Institute on Aging of the National Institutes of Health under award number K01AG047280. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

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

1. Stein G, Sacred Emily. Geography and plays. Boston: The Four Seas Company; 1922, p. 178.
2. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5): 550–60.
3. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci: Rev J Inst Math Stat. 2010;25(1):1.
4. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295–300.
5. Rothman K, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95(S1):S144–50.
6. Kiefer J, Wolfowitz J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann Math Stat. 1956:887-906.
7.• Gibbard A, Harper WL. Counterfactuals and two kinds of expected utility: Springer Netherlands ; 1981, pp 153-190. [**A fundamental development of counterfactual theory which describes the requirement of consistency for inference**.]
8. Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? Epidemiology. 2010;21(6):872–5.
9. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, et al. Socioeconomic status in health research: one size does not fit all. JAMA. 2005;294(22):2879–88. doi:10.1001/jama.294.22.2879.
10. Adler NE, Rehkopf DH. US disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health. 2008;29:235–52.
11. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology (Cambridge, Mass). 2011;22(3).
12. Duncan GJ, Magnuson KA. Off with Hollingshead: Socioeconomic resources, parenting, and child development. Socioeconomic status, parenting, and child development. 2003:83-106.
13. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. Am J Epidemiol. 1999;150(2):113–20.
14.•• Hernán M, Taubman S. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. International Journal of Obesity. 2008;32:S8–S14. **This article gives an overview of the importance of the consistency assumption for causal inference in epidemiology illustrated using the example of studies of the effects of obesity on mortality**.
15.• Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009;20(1):3–5. **A summary of the importance of the consistency assumption.**
16.• VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009;20(6):880–3. **This article clarifies the basis of the consistency assumption and provides details about the range of treatments.**
17. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. J Causal Inference. 2013;1(1):107–34.
18. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: mechanisms and causal evidence. Ann N Y Acad Sci. 2010;1186(1):56–68.
19. Benzeval M, Judge K. Income and health: the time dimension. Soc Sci Med. 2001;52(9):1371–90.
20. Catalano R, McConnell W. Psychiatric emergencies: the check effect revisited. J Health Soc Behav. 1999;79–86.
21. Rehkopf DH, Strully KW, Dow WH. The short-term impacts of Earned Income Tax Credit disbursement on health. Int J Epidemiol. 2014:dyu172.
22. Wolfe B, Jakubowski J, Haveman R, Courey M. The income and health effects of tribal casino gaming on American Indians. Demography. 2012;49(2):499–524.
23. Fernald LC, Gertler PJ, Neufeld LM. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. Lancet. 2008;371(9615):828–37.
24. Smith JP. Healthy bodies and thick wallets: the dual relation between health and economic status. J Econ Perspect: J Am Econ Assoc. 1999;13(2):144.
25. Gardner J, Oswald AJ. Money and mental wellbeing: a longitudinal study of medium-sized lottery wins. J Health Econ. 2007;26(1):49–60.
26. Apouey B, Clark AE. Winning big but feeling no better? The effect of lottery prizes on physical and mental health. Health Econ. 2015;24(5):516–38.
27. Evans WN, Moore TJ. The short-term mortality consequences of income receipt. J Public Econ. 2011;95(11):1410–24.
28. Romich JL, Weisner T. How families view and use the EITC: advance payment versus lump sum delivery. Nat Tax J. 2000:1245-65.
29. Cohen AK, Syme SL. Education: a missed opportunity for public health intervention. Am J Public Health. 2013;103(6):997–1001.
30. Glymour MM, Manly JJ. Lifecourse social conditions and racial and ethnic patterns of cognitive aging. Neuropsychol Rev. 2008;18(3):223–54. doi:10.1007/s11065-008-9064-z.
31. Manly JJ, Schupf N, Tang MX, Stern Y. Cognitive decline and literacy among ethnically diverse elders. J Geriatr Psychiatry Neurol. 2005;18(4):213.
32. Liu SY, Glymour MM, Zahodne LB, Weiss C, Manly JJ. Role of place in explaining racial heterogeneity in cognitive outcomes among older adults. J Int Neuropsychol Soc. 2015;21(09):677–87. doi:10.1017/S1355617715000806.
33. Nguyen T, Tchetgen Tchetgen E, Kawachi I, Gilman SE, Walter S, Glymour MM. Comparing alternative effect decompostion methods: the role of literacy in mediating educational effects on mortality. Epidemiology. 2015.
34. Nye B, Hedges LV, Konstantopoulos S. The long-term effects of small classes: a five-year follow-up of the Tennessee class size experiment. Educ Eval Policy Anal. 1999;21(2):127–42.
35. Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D. How does your kindergarten classroom affect your

earnings? Evidence from Project STAR. Nat Bur Econ Res 2011;126(4):1593–660.
36. Lleras-Muney A. The relationship between education and adult mortality in the United States. Rev Econ Stud. 2005;72(1):189–221.
37. Silles MA. The causal effect of education on health: evidence from the United Kingdom. Econ Educ Rev. 2009;28(1):122–8.
38. Glymour MM, Kawachi I, Jencks CS, Berkman LF. Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. J Epidemiol Community Health. 2008;62(6):532–7. doi:10.1136/jech.2006.059469.
39. Nguyen T, Tchetgen Tchetgen E, Kawachi I, Gilman SE, Walter S, Liu S, Manly J, Glymour M. Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. Ann Epidemiol. 2016;26(1):71–76.e3. doi:10.1016/j.annepidem.2015.10.006.
40. Campbell F, Conti G, Heckman JJ, Moon SH, Pinto R, Pungello E, et al. Early childhood investments substantially boost adult health. Science. 2014;343(6178):1478–85.
41. Conti G, Heckman JJ, Pinto R. The effects of two influential early childhood interventions on health and healthy behaviors. Nat Bur Econ Res. 2015.
42. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science. 1997;277(5328):918–24.
43. Katz B. Neighbourhoods of choice and connection: The evolution of American neighbourhood policy and what it means for the United Kingdom: Joseph Rowntree Foundation; 2004.
44. Osypuk TL, Tchetgen Tchetgen EJ, Acevedo-Garcia D, Earls FJ, Lincoln A, Schmidt NM, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: results from a randomized trial. Arch Gen Psychiatry. 2012;69(12):1284–94.
45. Kling JR, Liebman JB, Katz LF. Experimental analysis of neighborhood effects. Econometrica. 2007;75(1):83–119.
46. Lindberg RA, Shenassa ED, Acevedo-Garcia D, Popkin SJ, Villaveces A, Morley RL. Housing interventions at the neighborhood level and health: a review of the evidence. J Public Health Manag Pract. 2010;16(5):S44–52.
47. Osypuk TL, Joshi P, Geronimo K, Acevedo-Garcia D. Do social and economic policies influence health? A review. Curr Epidemiol Rep. 2014;1(3):149–64.
48. Emerson RW, John Davis Batchelder Collection (Library of Congress). Essays. Boston; J. Munroe and Company; 1841. 303 p. p.
49. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology—the need for a pluralistic approach. Int J Epidemiol. 2016;dvy341.
50. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887–92.
51. Druesne-Pecollo N, Latino-Martel P, Norat T, Barrandon E, Bertrais S, Galan P, et al. Beta-carotene supplementation and cancer risk: a systematic review and metaanalysis of randomized controlled trials. Int J Cancer. 2010;127(1):172–84.
52. Van Poppel G. Epidemiological evidence for beta-carotene in prevention of cancer and cardiovascular disease. Eur J Clin Nutr. 1996;50:S57–61.
53. Peto R, Doll R, Buckley JD, Sporn M. Can dietary beta-carotene materially reduce human cancer rates? 1981;290(5803):201–208
54. Nandi A, Harper S. How consequential is social epidemiology? A review of recent evidence. Curr Epidemiol Rep. 2014;2(1):61–70.
55. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013;178(8):1185–1191. doi:10.1093/aje/kwt172.
56. Collins J, Koplan JP. Health impact assessment: a step toward health in all policies. JAMA. 2009;302(3):315–7.
57. Marmot MG, Allen J. From science to policy. In: Berkman LF, Kawachi I, Glymour M, editors. Social epidemiology. 2nd ed. New York: Oxford University Press; 2014.