Meaningful Causal Decompositions in Health Equity Research

Definition, Identification, and Estimation Through a Weighting Framework

John W. Jackson[a,b,c,d,e]

Abstract: Causal decomposition analyses can help build the evidence base for interventions that address health disparities (inequities). They ask how disparities in outcomes may change under hypothetical intervention. Through study design and assumptions, they can rule out alternate explanations such as confounding, selection bias, and measurement error, thereby identifying potential targets for intervention. Unfortunately, the literature on causal decomposition analysis and related methods have largely ignored equity concerns that actual interventionists would respect, limiting their relevance and practical value. This article addresses these concerns by explicitly considering what covariates the outcome disparity and hypothetical intervention adjust for (so-called allowable covariates) and the equity value judgments these choices convey, drawing from the bioethics, biostatistics, epidemiology, and health services research literatures. From this discussion, we generalize decomposition estimands and formulae to incorporate allowable covariate sets (and thereby reflect equity choices) while still allowing for adjustment of non-allowable covariates needed to satisfy causal assumptions. For these general formulae, we provide weighting-based estimators based on adaptations of ratio-of-mediator-probability and inverse-odds-ratio weighting. We discuss when these estimators reduce to already used estimators under certain equity value judgments, and a novel adaptation under other judgments.

Keywords: Allowability; Causal inference; Disparity; Equity; Ethics; Interventional effects; Mediation analysis; Oaxaca–Blinder decomposition

Editors' note: Related article appears on page 291.

Submitted September 21, 2019; accepted December 7, 2020

From the [a]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [b]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [c]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [d]Johns Hopkins Center for Health Equity, Baltimore, MD; and [e]Johns Hopkins Center for Health Disparities Solutions, Baltimore, MD.

This research was supported by a grant from the National Heart Lung and Blood Institute (K01HL145320).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: John W. Jackson, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 614 N. Broadway Room E-6543, Baltimore, MD 21205. E-mail: john.jackson@jhu.edu.

ISSN: 1044-3983/21/3202-0282

DOI: 10.1097/EDE.0000000000001319

(*Epidemiology* 2021;32: 282–290)

Health disparities represent differences across socially privileged versus socially marginalized groups that society considers inequitable, avoidable, and unjust.[1] Interventions that address disparities[2] usually affect risk factors that are overrepresented among marginalized groups. Often their evidence base draws from studies that compare measures of disparities before and after adjustment for a risk factor (the difference method[3]). But the changes seen after such adjustments may be due to confounding, selection bias, or information bias. The results may not imply that the risk factor studied is the one to be intervened upon. Causal decomposition methods[4–8] compare to a counterfactual disparity under a hypothetical intervention on a target risk factor. They overcome the limitations of simple adjustment through sound study design and unverifiable assumptions to rule out alternative explanations (bias). They ask a simple question: how disparities in outcomes would change if disparities in a targeted factor (that affects the outcome) were removed?

Unfortunately, estimators used for causal decomposition have ignored how disparities and hypothetical interventions are defined, limiting their relevance in health equity research. What should a disparity measure for the outcome condition on or standardize over? Surveillance reports that track health disparities usually adjust for age and sex, but some decomposition estimators adjust for all covariates needed to identify a causal effect. Ideally, disparity measures should reflect judgments about what constitutes an inequitable difference in the outcome.

Estimators have also ignored how hypothetical interventions are defined. As such, the hypothetical intervention, meant to remove disparities in a targeted factor, may not reflect equity concerns that actual interventionists would respect. For example, an intervention to remove disparities in healthcare should depend on clinical status, but not socioeconomic status which is irrelevant for medical care. Hypothetical interventions should reflect judgments about what constitutes an inequitable difference in the target.

In this article, we outline a framework for defining disparities and interventions in causal decomposition analysis and provide estimands, nonparametric formulae, and weighting estimators to implement it. Specifically, we draw from the bioethics, biostatistics, epidemiology, and health services research literatures to consider when covariates are "allowable" for adjusting disparity measures and hypothetical interventions. The estimands, formulae, and estimators we propose partition covariates into "allowable" sets that define the disparity and intervention, and a "non-allowable" set that is needed to identify the causal effect. Under a motivating example, we use this framework to examine estimators of natural and interventional effects (mediation analysis) and discrimination (Oaxaca–Blinder decomposition) and the equity judgments they imply and consider a meaningful alternative.

MOTIVATING EXAMPLE AND NOTATION

We ground our ideas through a motivating example from clinical medicine: how to reduce disparities in hypertension control by intervening on decisions to intensify antihypertensive treatment?[9] Suppose a healthcare system administrator wants to address disparities in hypertension control across race/ethnicity and tasks us with forming a cohort to study them. Patients are enrolled at their first visit if, on the basis of their systolic blood pressure (L_1), they can be classified as hypertensive (Y_1; 1 = yes, $L_1 \geq 140$ mmHg, 0 = no, $L_1 < 140$ mmHg). For simplicity we ignore diastolic blood pressure. At 6 months follow-up systolic blood pressure (L_2) and uncontrolled hypertension (Y_2; 1 = yes, $L_2 \geq 140$ mmHg, 0 = no, $L_2 < 140$ mmHg) are measured. Disparities in hypertension control may arise through clinical uncertainty.[10] When providers know less about their patients' medical condition (e.g., due to poor provider communication) their decision-making may rely on stereotypes. We are interested in how eliminating disparities in treatment decisions would affect disparities in hypertension control.

We thus record the patient's race R_0 where $R_0 = r_0$ represents membership in the marginalized group (e.g., blacks) and $R_0 = r_0'$ the privileged group (e.g., whites), demographics age X_0^{age} and sex X_0^{sex}, whether antihypertensive treatment was intensified at the initial visit (M_1; 1 = yes, 0 = no) as well as socioeconomic factors such as educational attainment (X_0^{edu}) and private health insurance (X_0^{ins}) that may implicitly affect treatment and hypertension control. We also record diabetes diagnosis (X_0^{dia}) which, as a marker of cardiovascular risk, predicts blood pressure and may influence treatment decisions.

The subscripts in our notation grossly indicate a temporal ordering, where "0" indicates variables that are realized before baseline, "1" indicate those realized at baseline, and "2" indicates those realized at follow-up. We assume that some associations between race/ethnicity R_0, X_0, and L_1 are driven by historical processes H_0 (e.g., slavery, Jim Crow, federal and local housing policies) as depicted in our causal graph (Figure), provided for intuition. We allow that unmeasured factors U_0 may correlate repeated measures of systolic blood pressure L_1 and L_2 and possibly other variables but do not independently predict treatment intensification M_1. In our notation, random variables are written in uppercase and their realizations in lowercase. For any variable V, the probability $P(V = v \mid W = w)$ is abbreviated as $P(v \mid w)$.

To develop general formulae, we will discuss two nonoverlapping sets of "allowable" covariates, those used to define the disparity in hypertension control (the outcome Y_2), denoted as $\boldsymbol{A}_1^{\boldsymbol{y}}$, and those used to further define the intervention on treatment intensification (the targeted factor M_1), denoted as $\boldsymbol{A}_1^{\boldsymbol{m}}$. These sets are restricted to variables not affected by treatment intensification M_1. While the time subscript on these sets indicates measurement at baseline, they can include variables measured before baseline as well.

We let $G_{m_1|\boldsymbol{a}_1^y, \boldsymbol{a}_1^m}$ denote a hypothetical stochastic intervention[11,12] to set the conditional distribution of treatment intensification M_1 among blacks to the distribution among whites with identical values for $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$, denoted as $P(m_1 \mid R_0 = r_0^{'}, \boldsymbol{a}_1^{\boldsymbol{y}}, \boldsymbol{a}_1^{\boldsymbol{m}})$. Suppose we choose $\boldsymbol{A}_1^{\boldsymbol{y}}$ as X^{age} and X^{sex}, and $\boldsymbol{A}_1^{\boldsymbol{m}}$ as X^{dia} and L_1. Among blacks, under $G_{m_1|\boldsymbol{a}_1^y \boldsymbol{a}_1^m}$, their treatment is intensified according to a random draw from the distribution among whites who share the same values for $\mathrm{X}_0^{\mathrm{age}}$, X_0^{sex}, X_0^{dia}, and L_1.

MEANINGFUL DISPARITY AND INTERVENTION DEFINITION

The disparity in hypertension control arises in large part because contextual and other cardiovascular risk factors, and also access to and quality of medical care, are differentially distributed across race.[13] In the United States, these inequities result from centuries of injustice, including slavery, Jim Crow, government policy, and other forms of structural, cultural, and personal racism.[14] In our example, causal decomposition analysis asks how, among those with uncontrolled hypertension at baseline, the disparity in hypertension control at follow-up Y_2 is affected by intervening to eliminate the disparity in treatment decisions M_1.

Defining a disparity is a complex process involving decisions about what is fair and just in the distribution of health and its determinants.[15–19] We focus on which covariates are considered allowable for adjustment in defining outcome disparities and interventions, and how these choices relate to equity value judgments. The notion of allowability has been discussed in the context of medical goods and used to define healthcare disparities in counterfactual terms.[20,21] Our discussion broadens the concept to health disparities and decomposition analysis but avoids counterfactual definitions of disparity.

Allowability

In the bioethics literature, many define a health disparity as an avoidable, systematic difference between socially advantaged versus marginalized groups, wherein the marginalized

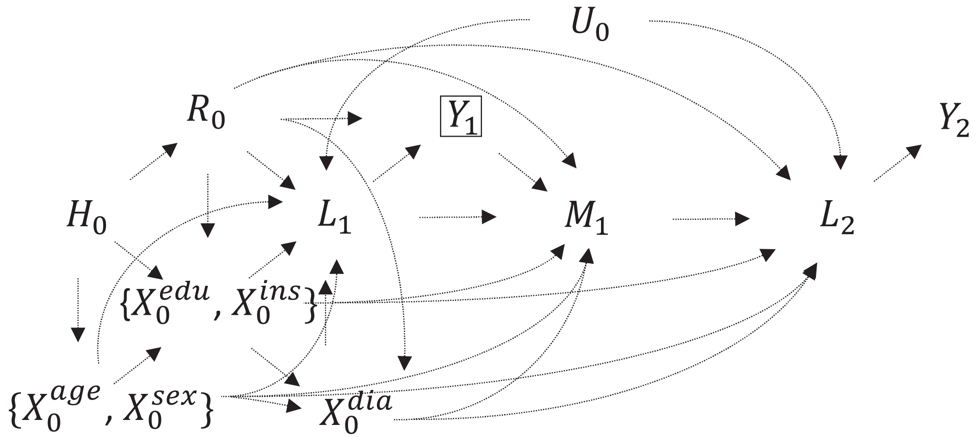

FIGURE. Causal diagram depicting relationships between history (H_0), race (R_0), demographics age and sex (X^{age} and X^{sex}), socioeconomic covariates educational attainment and private health insurance (X^{edu} and X^{ins}), diabetes (X^{dia}), baseline and follow-up blood pressure (L_1 and L_2), hypertensive status at baseline and subsequent control (Y_1 and Y_2), and treatment intensification M_1. Additional arrows could be allowed from U to the covariates X. The subscripts denote, grossly, the time of realization: "0" pre-baseline, "1" baseline, "2" follow-up. The box around Y_1 indicates that the population of interest is those with uncontrolled hypertension at baseline. To simplify the graph, brackets group covariates with similar relationships. Within the subset $\{X_0^{age}, X_0^{sex}\}$ no causal relationship is specified; within the subset $\{X_0^{edu}, X_0^{ins}\}$ X_0^{edu} causes X_0^{ins}.

group is further disadvantaged on health.[22,23] In the health services research literature, disparities in healthcare are often defined using the Institute of Medicine definition, as differences in healthcare services that are not due to differences in underlying health needs or preferences.[20,24] A careful reading of these definitions recognizes that, in defining disparities, both avoid detailing a causal model for how they arise. While there are some objections to this,[21] there are practical and scientific reasons why this may be desirable.[22,23]

These definitions encourage us to consider what sources of difference might be considered fair or allowable and to take these off the table when measuring disparity. By corollary, non-allowable sources are those that are unjust and thus contribute to disparity. For health outcomes, allowable sources typically include demographic factors such as age and sex.[25] For medical goods, allowable sources typically include clinical status, history, and presentation.[20,24] All other factors are often considered non-allowable. Although these represent default choices in many studies, other positions become clear when issues of modifiability, amenability to intervention, social contract, and purpose are considered.

Modifiability and Amenability to Intervention

In defining health disparity measures, it is common to see innate, apparently non-modifiable factors such as chronological age, sex, and even somatic genotype treated as allowable and adjusted for. While society has profound power over the historical distribution of innate factors (through war, genocide, racism, and policy), for a fixed living population, it has no ability to modify them. On these grounds, some might argue that innate differences leading to differences in health are not necessarily unfair.[19]

Some object to using strict modifiability to decide whether to treat a covariate as allowable. The conceptualization and role of innate characteristics in daily, civic, and economic life are entirely under society's control.[26] Moreover, while society cannot change characteristics defined at birth, it can address their effects.[16,22] Social programs can be made age and sex appropriate, increasing their effectiveness. Targeted cancer therapies have been developed for genetic profiles. If one considers a disparity as any difference placing a marginalized group at further disadvantage, and society can address the effects of innate characteristics, one might then consider their differences as contributors to disparity when society fails to adequately respond to them. Thus, innate differences could be considered non-allowable and not adjusted for. The decision to not adjust for innate factors is best applied when the marginalized group is disadvantaged on the innate factor. Failing to adjust otherwise could mask unjust differences.

Consider these arguments in our motivating example. First, we have preexisting conditions. Blacks disproportionately encounter barriers to care, such as lack of health insurance, leading to higher prevalence of chronic conditions at baseline encounters. By that point, a patient's clinical history is beyond the control of the clinician and healthcare system. However, this history can be managed for better prognosis, for example, by consulting a specialist. Clinical guidelines recommend tailored treatment protocols for patients with diabetes, kidney disease, and heart failure. Thus, healthcare can respond to disparate preexisting conditions. Failure to do so equitably would contribute to unjust differences in prognosis. Second, we have age. Blacks are typically younger than whites, and increasing age predicts poor hypertension control. Not adjusting for age would mask unjust differences in hypertension control. Overall, when measuring disparities in prognosis, if there are disparities in preexisting conditions, one might reasonably decide to treat them as non-allowable (and not adjust). If blacks are younger than whites, one might reasonably decide to treat age as allowable (and adjust).

Social Contract

When defining hypothetical interventions that address disparities in goods, allowability choices should consider the social contract. The distribution of any good is ideally governed by norms and conditions that society has agreed upon as fair and just. That is, we believe that decisions about goods are fair if and only if they are based upon ideal criteria that reflect our shared norms and values. That is, in defining interventions on goods, we treat our idealized criteria as allowable.

These arguments have clear implications for interventions that address medical goods. Medical ethics dictates that decisions be clinically appropriate. For addressing disparities in diagnosis, allowable sources could include information needed to accurately differentiate between syndromes, such as presentation and test results. For addressing disparities in treatment, allowable sources could include factors that indicate and modify treatment effectiveness, such as comorbid conditions.[24] For addressing disparities in social conditions, perhaps through a community health worker, allowable sources could include social needs. In our motivating example, the hypothetical intervention to intensify treatment would need to consider age, sex, baseline blood pressure, and diabetes but ignore socioeconomic status. Otherwise, the intervention would not only be unethical, but also inequitable. It would preserve racial differences in treatment that operate through racial differences in socioeconomic status.[5] Observing the social contract reflects that our goal is not equal treatment, but equitable treatment.[24]

When defining disparities in outcomes that represent goods, allowability choices should also consider the social contract. This again would treat idealized decision-inputs as allowable. An exception can be made when one seeks to reduce racial differences in goods by eliminating disparities in a criterion governing their distribution. Suppose one were studying a lower rate of listing for transplantation among blacks compared with whites. Listing decisions often consider anticipated social support. An overall racial difference in listing partly due to differences in social support might be concerning, even considered disparate. Here, social support could be treated as non-allowable (and not adjusted for) so it could be studied as a hypothetical intervention.

Purpose

In defining outcome disparity measures, allowability must also consider their use. In surveillance and quality assessment, disparity measures can track how well a society, institution, or actor meets objectives. In these settings, when factors the actor does not control are treated as non-allowable (and not adjusted for), this can lead to deleterious effects. In our example, if an external body were benchmarking clinical practices based on disparities in hypertension control rates without any risk adjustment, those that serve marginalized populations with comorbidities may score worse. Those clinics might then be incentivized to avoid complex patients.[27] When causal decomposition analysis is applied to performance assessment measures, it is advisable to study the one used in practice, which might call for preexisting conditions to be treated as allowable, for the sake of risk adjustment.[28]

Measurement

We now turn to how outcome disparities are defined in our causal decomposition analysis. For the outcome, we define disparity as the mean outcome difference across levels of social groups, where the distribution of allowable covariates is standardized. Here, we use the pooled distribution as the standard. In defining the hypothetical intervention, which is intended to remove disparities in a targeted factor, we adjust for allowable covariates through conditioning. We have chosen these simple statistical measures of disparity because the observed disparities can be estimated directly from the data with minimal assumptions and are consistent with widely adopted definitions of disparity.[22–24] These definitions assume common support for the allowable covariates across race. Otherwise, adjustment would fail to remove the influence of allowable covariates. Defining disparities using the distribution among blacks or whites, rather than the pooled as the standard may weaken this common support assumption.

MEANINGFUL DISPARITY DECOMPOSITION

Here, we present general causal decomposition estimands defined by allowable covariates for the disparity in hypertension control (the outcome Y_2) and the hypothetical intervention to intensify treatment (the targeted intervention M_1). Under assumptions, we provide identifying formulae and weighting-based estimators. Importantly, these formulae and estimators can incorporate confounders that are considered non-allowable, without using them to define outcome disparities or the intervention. All expressions condition on the population of interest, persons with uncontrolled hypertension at baseline.

Definition

The observed disparity in uncontrolled hypertension is:

$$\sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2|r_0, \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) - \sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2|r_0', \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) \qquad (1)$$

The change in disparity under the intervention $G_{m_1|\boldsymbol{a}_1^y, \boldsymbol{a}_1^m}$ to remove the disparity in treatment intensification is:

$$\sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2|r_0, \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) - \sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2\left(G_{m_1|\boldsymbol{a}_1^y, \boldsymbol{a}_1^m}\right)|r_0, \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) \qquad (2)$$

The remaining disparity after the intervention $G_{m_1|\boldsymbol{a}_1^y, \boldsymbol{a}_1^m}$ is:

$$\sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2\left(G_{m_1|\boldsymbol{a}_1^y, \boldsymbol{a}_1^m}\right)|r_0, \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) - \sum\nolimits_{\boldsymbol{a}_1^y} E\left[Y_2|r_0', \boldsymbol{a}_1^y\right]P(\boldsymbol{a}_1^y) \qquad (3)$$

The covariates $\boldsymbol{A}_1^{\boldsymbol{y}}$ used to define the disparity are considered outcome-allowable. These covariates $\boldsymbol{A}_1^{\boldsymbol{y}}$, along with $\boldsymbol{A}_1^{\boldsymbol{m}}$, are used to define the intervention so both are considered target-allowable. These equations (1)–(3) are agnostic about whether components of $\boldsymbol{A}_1^{\boldsymbol{y}}$ affect components of $\boldsymbol{A}_1^{\boldsymbol{m}}$ and vice-versa, and are also agnostic about whether $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$ are affected by race R_0, but do require that $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$ are not affected by the targeted variable, treatment intensification M_1. This will be satisfied in a design where the eligibility criteria defining the population of interest, the timing of the intervention on M_1, and the start of follow-up coincide, and allowable covariates are measured at or before this moment.

Identification

The counterfactual disparity reduction (2) and residual (3) are not observed. These expressions can be identified with observational data under assumptions (see eAppendix; http://links.lww.com/EDE/B759, for formal statements). Among blacks, where $R_0 = r_0$, we assume conditional exchangeability,[29] or no unmeasured confounding of the relationship between treatment intensification M_1 and hypertension control Y_2: given the allowables $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$ and an additional set of non-allowable confounders $\boldsymbol{N}_1$, the potential outcomes under treatment intensification are independent of observed treatment intensification. We further assume positivity[29] among blacks, where there is a positive conditional probability of each observed value for treatment intensification M_1 given the allowable covariate sets $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$ used to define the estimands and the non-allowable covariates $\boldsymbol{N}_1$ used to help identify them. Subtly, it is sufficient if these conditions hold only for the values of M_1 observed among whites who share (with blacks) identical values of $\boldsymbol{A}_1^{\boldsymbol{y}}$ and $\boldsymbol{A}_1^{\boldsymbol{m}}$. As a consequence, they can be satisfied when, for some subgroups, all have their treatment intensified. We also assume common support across race R_0 for the targeted factor and allowable covariates (jointly). Finally, we assume consistency[29] among blacks, that their outcomes would be the same regardless if their values were merely observed or set by hypothetical intervention. These assumptions are strong and the ability to satisfy them will vary across substantive settings.[30]

When these assumptions hold, we can identify the proportion with uncontrolled hypertension among blacks under $G_{m_1|\boldsymbol{a}_1^{\boldsymbol{y}},\boldsymbol{a}_1^{\boldsymbol{m}}}$ as:

$$\begin{aligned}
&\sum\nolimits_{\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2\left(G_{m_1|\boldsymbol{a}_1^{\boldsymbol{y}},\boldsymbol{a}_1^{\boldsymbol{m}}}\right)|r_0,\boldsymbol{a}_1^{\boldsymbol{y}}\right]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \\
&= \sum\nolimits_{m_1,\boldsymbol{n}_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2|r_0,m_1,\boldsymbol{n}_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right]P\left(m_1|r_0',\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right) \qquad (4) \\
&\qquad P\left(\boldsymbol{n}_1|r_0,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right)P(\boldsymbol{a}_1^{\boldsymbol{m}}\,|\,r_0,\boldsymbol{a}_1^{\boldsymbol{y}})P(\boldsymbol{a}_1^{\boldsymbol{y}})
\end{aligned}$$

This allows us to identify the disparity reduction (2) given that the observed proportion of uncontrolled hypertension among blacks is:

$$\begin{aligned}
&\sum\nolimits_{\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2|r_0,\boldsymbol{a}_1^{\boldsymbol{y}}\right]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \\
&= \sum\nolimits_{m_1,\boldsymbol{n}_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2|r_0,m_1,\boldsymbol{n}_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right]P\left(m_1|r_0,\boldsymbol{n}_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right) \qquad (5) \\
&\qquad P(\boldsymbol{n}_1\,|\,r_0,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}})P(\boldsymbol{a}_1^{\boldsymbol{m}}\,|\,r_0,\boldsymbol{a}_1^{\boldsymbol{y}})P(\boldsymbol{a}_1^{\boldsymbol{y}})
\end{aligned}$$

We can also identify the disparity residual (3) given that the observed proportion of uncontrolled hypertension among whites is:

$$\begin{aligned}
&\sum\nolimits_{\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2|r_0',\boldsymbol{a}_1^{\boldsymbol{y}}\right]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \\
&= \sum\nolimits_{m_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}} E\left[Y_2|r_0',m_1,\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right]P\left(m_1|r_0',\boldsymbol{a}_1^{\boldsymbol{m}},\boldsymbol{a}_1^{\boldsymbol{y}}\right) \qquad (6) \\
&\qquad P(\boldsymbol{a}_1^{\boldsymbol{m}}|\,r_0',\boldsymbol{a}_1^{\boldsymbol{y}})P(\boldsymbol{a}_1^{\boldsymbol{y}})
\end{aligned}$$

Note that in equations (4)–(6) $P(\boldsymbol{a}_1^{\boldsymbol{y}})$ is replaced by $P(\boldsymbol{a}_1^{\boldsymbol{y}}|\,r_0)$ if the distribution of outcome-allowable covariates that standardizes the disparity measure is drawn from blacks, and by $P(\boldsymbol{a}_1^{\boldsymbol{y}}|\,r_0')$ if drawn from whites.

Estimation

Equations (4)–(6) could be estimated using Monte-Carlo integration similar to the parametric g-formula,[31] using correctly specified parametric models, as in other contexts.[32–37] A key distinction from these proposals is what the models condition on. The outcome models condition on: (a) allowable and non-allowable covariates in the observed and counterfactual scenarios for blacks, (b) only allowable covariates in the observed scenario among whites. The target factor models condition on: (a) allowable and non-allowable covariates in the observed scenario for blacks, (b) only allowable covariates in the observed scenario for whites and the counterfactual scenario for blacks. Other algorithms are possible (see eAppendix; http://links.lww.com/EDE/B759) but this one does not require non-allowables to be measured or even defined among whites.

Because simulation-based approaches are computationally intensive and require correctly specifying several models, we present two simple weighting-based estimators that can be implemented with standard statistical software routines (briefly described in the eAppendix; http://links.lww.com/EDE/B759). In the health equity context, these encompass existing estimators in the economics,[38–40] sociology,[41,42] biostatistics, and epidemiology literatures,[43–48] including those used to estimate natural direct and indirect effects,[49,50] path-specific effects,[51] interventional effects,[11,12,46] and discrimination[52,53] under certain allowability choices, and novel adaptations under others.

Ratio of Mediator Probability Weighting Estimation

The first weighting procedure is based on a ratio of probabilities for treatment intensification M_1. The disparity reduction (2) and residual (3) under the intervention $G_{m_1|\boldsymbol{a}_1^{\boldsymbol{y}},\boldsymbol{a}_1^{\boldsymbol{m}}}$ are estimated by comparing weighted means for blacks and whites. First, we estimate the observed proportion with uncontrolled hypertension Y_2 among blacks, standardized for the outcome-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$:

$$\sum_{\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}} E\left[Y_2|r_0, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right]P\left(\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right) = E[E\left[Y_2 \times w_{r_0}|r_0, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right] | r_0] \tag{7a}$$

where the weight $w_{r_0} = \dfrac{P(r_0)}{P(r_0 \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}$ (7b)

Next, we estimate the observed proportion with uncontrolled hypertension Y_2 among whites, standardized for the outcome-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$:

$$\sum_{\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}} E\left[Y_2|r_0', \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right]P(\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}) = E[E\left[Y_2 \times w_{r_0'}|r_0', \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right] | r_0'] \tag{8a}$$

where the weight: $w_{r_0'} = \dfrac{P(r_0')}{P(r_0' \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}$ (8b)

Last, we estimate the counterfactual proportion with uncontrolled hypertension Y_2 among blacks under the intervention $G_{m_1|\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}}$, which depends on the outcome- and target-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$ and $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{m}}$. Like the observed outcomes, the estimated counterfactuals are standardized for the outcome-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$:

$$\begin{aligned}&\sum_{\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}} E\left[Y_2(G_{m_1|\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}})|r_0, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right]P(\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})\\ &= E[E\left[Y_2 \times w_{r_0}^{rmpw}|r_0, m_1, \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right] | r_0]\end{aligned} \tag{9a}$$

where the weight: $w_{r_0}^{rmpw} = \dfrac{P(m_1 \mid r_0', \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(m_1 \mid r_0, \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})} \times \dfrac{P(r_0)}{P(r_0 \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}$ (9b)

The counterfactual disparity reduction is obtained by subtracting (9a) from (7a), and the residual by subtracting (8a) from (9a).

Inverse Odds Ratio Weighting Estimation

The second weighting procedure is based on a ratio of inverted odds for race R. The disparity reduction (2) and residual (3) under the intervention $G_{m_1|\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}}$ are estimated by comparing weighted means for blacks and whites. The observed proportions with uncontrolled hypertension Y_2 among black and whites, standardizing for the outcome-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$, are respectively given in (7a) and (8a). The counterfactual proportion with uncontrolled hypertension Y_2 among blacks under the intervention $G_{m_1|\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}}$, standardized for the outcome-allowable covariates $\boldsymbol{A}_\mathbf{1}^{\boldsymbol{y}}$, is given by:

$$\begin{aligned}&\sum_{\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}} E\left[Y_2(G_{m_1|\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}})|r_0, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right]P(\boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})\\ &= E[E\left[Y_2 \times w_{r_0}^{iorw}|r_0, m_1, \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}}\right] | r_0]\end{aligned} \tag{10a}$$

where the weight:

$$w_{r_0}^{iorw} = \frac{\dfrac{P(r_0' \mid m_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(r_0 \mid m_1, \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}}{\dfrac{P(r_0' \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(r_0 \mid \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}} \times \frac{P(m_1 \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(m_1 \mid \boldsymbol{n}_1, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{m}}, \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})} \times \frac{P(r_0)}{P(r_0 \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})} \tag{10b}$$

When all covariates are treated as allowable, the middle term in (10b) cancels and does not need to be estimated. As with ratio of mediator probability weighting estimation, the disparity reduction is estimated by subtracting (10a) from (7a), and the disparity residual by subtracting (8a) from (10a). Finally, with both ratio of mediator probability and inverse odds ratio weighting estimation, each of the weights are multiplied by a factor of $\dfrac{p(r_0 \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(r_0)}$ if the distribution of outcome-allowable covariates that standardizes the disparity measure is drawn from blacks, and by a factor of $\dfrac{p(r_0' \mid \boldsymbol{a}_\mathbf{1}^{\boldsymbol{y}})}{P(r_0')}$ if drawn from whites.

A CLOSER LOOK AT EXISTING ESTIMATORS

Estimators of natural and path-specific effects, their analogs, and the Oaxaca–Blinder decomposition are often applied to study disparities. We now use our general expressions to examine these estimators and the equity value judgments they convey (see eAppendix; http://links.lww.com/EDE/B759, for a formal discussion). Many of these estimators were developed in the context of mediation analysis to study the effects of exposures and others to measure discrimination.[5,54,55] When these estimators are used to study disparities, the decomposition estimands they identify are not always meaningful. The problem arises because of how they adjust for covariates. They implicitly define outcome disparities and interventions in ways that often ignore principles of modifiability, amenability to intervention, social contract, and purpose. Our generalized estimators explicitly consider allowable and non-allowable partitions, so users can be more intentional.

Natural Indirect Effect Analogue

Estimators of the natural direct and indirect effects include the regression approaches of Valeri and VanderWeele[56] and Breen et al,[57] ratio of mediator probability weighting of Hong,[41,42] natural effect models of Lange et al[43]; inverse odds ratio weighting of Tchetgen Tchetgen[45]; propensity score weighting of Huber[58]; imputation approaches of Albert[59] and Vansteelandt et al[60]; simulation approaches of Imai et al,[33] and Wang and Arah[34]; and targeted maximum likelihood estimation of Zheng and van der Laan.[48] In our example, these estimate the observed disparity (1) by conditioning the entire analysis on all covariates needed to de-confound the relationship between treatment decisions and hypertension control (or by standardizing across race). This renders all covariates as outcome-allowable. This is problematic because preexisting diabetes, low educational attainment, and lack of private insurance are risk factors for poor hypertension control and are more common among blacks. Moreover, their effects on hypertension control are amenable to intervention. Thus, treating these preexisting conditions as outcome-allowable artefactually diminishes the amount of unjust difference to be studied. Furthermore, these approaches estimate the disparity reduction (2) by allowing the hypothetical intervention on treatment intensification to depend on all covariates. This renders all covariates as target-allowable. This is problematic because, by ignoring the social contract, racial differences in treatment intensification that operate through racial differences in educational attainment would persist.

Path-specific Effect Analogues

Estimators of path-specific effects and their interventional analogues have also been applied to study disparities. In our example, all would estimate the observed disparity (1) by conditioning the entire analysis on age and sex, treating them as outcome-allowable (belonging to $\boldsymbol{A}_1^y$). But they differ in how they would treat covariates affected by race, which in our example are educational attainment, private insurance, diabetes, and baseline blood pressure. The weighting estimator proposed by VanderWeele et al[46] would treat such covariates as non-allowable (all belonging to $\boldsymbol{N}_1$, leaving $\boldsymbol{A}_1^m$ empty). Therefore, in estimating the disparity reduction (2), this intervention to set treatment decisions would only depend on age and sex. This is problematic because our society has generally agreed that treatment decisions should depend on clinical needs. Respecting the social contract would assign clinical needs as target-allowable. In contrast, the weighting approach proposed by Zheng and Van Der Laan[48] (and, upon recoding race, that of Miles et al[44]), would estimate the disparity reduction (2) by treating all race-affected covariates as allowable (all belonging to $\boldsymbol{A}_1^m$, leaving $\boldsymbol{N}_1$ empty). This intervention to set treatment decisions would depend not only on clinical needs but also educational attainment and private insurance. This is problematic because, as we argued with the natural indirect effect analogue, medical treatment decisions should not hinge on education or private insurance. Finally, a regression-based estimator of Jackson and VanderWeele[5] and a simulation-based estimator proposed by Vansteelandt and Daniel[36] make the same allowability choice as VanderWeele et al.[46] Additionally, the latter identifies a path-specific effect that does not map to the disparity reduction (2) in the presence target-outcome confounders affected by race, as in our motivating example, but rather the differential impact of two interventions (see eAppendix; http://links.lww.com/EDE/B759).

Oaxaca–Blinder Decompositions

The Oaxaca–Blinder Decomposition[52,53] has also been applied to study disparities. The approach estimates the disparity reduction (2) by using linear models to regress the outcome on treatment intensification and all covariates and carrying out what is known as a detailed decomposition with respect to treatment intensification.[5] This studies a marginal racial difference in hypertension control, with a hypothetical intervention to remove marginal differences in treatment intensification. Effectively, it treats all covariates as neither outcome-allowable nor target-allowable, leaving $\boldsymbol{A}_1^y$ and $\boldsymbol{A}_1^m$ empty. All covariates are treated as non-allowable and included through $\boldsymbol{N}_1$ only to control for confounding. Alternatively, reweighting estimators[38,39] of the Oaxaca–Blinder decomposition have been widely used. Here, the racial difference in hypertension control conditions on no covariates, leaving $\boldsymbol{A}_1^y$ empty, but the hypothetical intervention on treatment intensification conditions on all covariates, by including them in $\boldsymbol{A}_1^m$. Thus, no covariates are treated as outcome-allowable but all are treated as target-allowable. These allowability assignments share the problems listed for the estimators of natural and path-specific effect analogs.

A Meaningful Estimator

In our example, the many existing estimators we examined do not map to meaningful estimands because of the implicit allowability choices they make, which are summarized in the Table. We could, in applying the general formulae and weighting expressions, choose to deem age and sex as outcome- and target-allowable (by including them in $\boldsymbol{A}_1^y$), respecting principles of modifiability and amenability to intervention, and deem baseline blood pressure and diabetesas exclusively target-allowable (by including them in $\boldsymbol{A}_1^m$), respecting the social contract. Doing so would treat educational attainment and private insurance as non-allowable (by including them in $\boldsymbol{N}_1$) , using them to adjust for confounding but not to define the outcome disparity or hypothetical intervention.

DISCUSSION

We have shown how causal decomposition analysis can incorporate equity concerns by partitioning covariates into allowable and non-allowable subsets (the latter used for identification). When a covariate is deemed outcome-allowable, its contribution is removed so that we are left with a difference we consider unjust. The hypothetical intervention to remove disparities in a targeted factor is administered within levels of target-allowable covariates, so that the intervention is equitable. We have discussed how allowability choices can consider issues of modifiability, amenability to intervention, social contract, and purpose, reflecting value judgments about equity. We provided generalized nonparametric formulae and weighting-based estimators that are defined in terms of allowable and non-allowable subsets. Last, we discussed when these estimators reduce to existing ones under certain value judgments, unifying and clarifying various approaches from biostatistics, epidemiology, economics, and sociology in the health equity context.

Our proposal has implications for study design in causal decomposition analysis. Researchers should consider variables needed to sensibly measure disparity, and whether these are measured by the start of follow-up with common support among blacks and whites. When our estimators are used, in particular ratio of mediator probability weighting estimation and equations (4)–(6) estimated by Monte-Carlo integration,

TABLE. Allowability Designations of Estimators

	A_1^y	A_1^m	N_1
1. Oaxaca–Blinder decomposition via linear models[52,53]: all covariates are considered non-allowable	$\emptyset$	$\emptyset$	$X_0^{age}, X_0^{sex}, X_0^{edu}, X_0^{ins}, X_0^{dia}, L_1$
2. Oaxaca–Blinder decomposition estimator via reweighting functions: no covariates considered outcome-allowable; all covariates are considered target-allowable[38,39]	$\emptyset$	$X_0^{age}, X_0^{sex}, X_0^{edu}, X_0^{ins}, X_0^{dia}, L_1$	$\emptyset$
3. Natural direct/indirect effect analogue estimators[33,41–43,48,56–58]: all covariates are considered outcome- and target-allowable	$X_0^{age}, X_0^{sex}, X_0^{edu}, X_0^{ins}, X_0^{dia}, L_1$	$\emptyset$	$\emptyset$
4. Path-specific effect analog (I and II) estimator[5,36,46]: demographic covariates considered outcome- and target-allowable; remaining covariates considered non-allowable	X_0^{age}, X_0^{sex}	$\emptyset$	$X_0^{edu}, X_0^{ins}, X_0^{dia}, L_1$
5. Path-specific effect analog (III) estimator[44,47]: demographic covariates outcome- and target-allowable; remaining covariates considered target-allowable	X_0^{age}, X_0^{sex}	$X_0^{edu}, X_0^{ins}, X_0^{dia}, L_1$	$\emptyset$
6. A meaningful estimator: demographic covariates considered outcome- and target-allowable; clinical covariates considered target-allowable; socioeconomic covariates considered non-allowable	X_0^{age}, X_0^{sex}	X_0^{dia}, L_1	X_0^{edu}, X_0^{ins}

List of covariates include X^{age}, age; X^{sex}, sex; X^{edu}, educational attainment; X^{ins}, private health insurance; X^{dia}, diabetes, L_1, baseline blood pressure.
A_1^y indicates outcome- (and target-) allowable covariates; A_1^m, exclusively target-allowable covariates; N_1, non-allowable covariates; $\emptyset$, the empty set.

the non-allowables only need to be defined and measured among blacks. This is important given the way in which some non-allowable constructs, such as racial discrimination, may occur almost exclusively with racial or ethnic minorities.[8]

Regarding disparity definition, by only discussing additive measures of disparity across more versus less marginalized groups, and by ignoring group size, we implicitly entertained several value judgments.[16,17,22] Our results easily extend to other scales such as the risk ratio or odds ratio. Our approach involved a single axis of disadvantage, but could be extended to study intersectional disparities.[7,8,61] Regarding disparity measurement, selected populations induce correlations between race and outcomes through collider-stratification,[62] as in our example. Because selection occurs pre-target and conditional exchangeability nonetheless holds, we still have causal identification.[4] The impact and interpretation of this important issue for disparity measurement is left for future work.

Regarding estimation, our approach focused on a single target, and continuous and (even non-rare) binary mean outcomes, allowing for race-target, race-covariate, target-covariate, and covariate-covariate interactions. Considering earlier work on ratio of mediator probability and inverse odds ratio weighting estimation,[38,39,43,45] and also simulation-based estimation,[32] our approach should extend to multiple targets, distributional outcomes, repeated outcomes, and survival analysis but this is left for future work. Regarding the intervention, we focused on categorical targets. Extensions to continuous targets could involve estimating conditional densities, expanding on earlier work,[32,38,39] but this may prove difficult with several covariates. The weights must be estimated using correctly specified models (see eAppendix; http://links.lww.com/EDE/B759). Our focus was on conceptual issues in definition and their relevance for estimation. Future work will consider practical guidance in implementation.

CONCLUSIONS

We have outlined a framework for incorporating equity concerns into causal decomposition analysis. Our contributions should be of wide interest, particularly when there are baseline differences in non-modifiable factors or when hypothetical interventions concern socially distributed goods.

REFERENCES

1. Braveman PA, Kumanyika S, Fielding J, et al. Health disparities and health equity: the issue is justice. *Am J Public Health.* 2011;101(suppl 1):S149–S155.
2. Cooper LA, Hill MN, Powe NR. Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. *J Gen Intern Med.* 2002;17:477–486.
3. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health.* 2016;37:17–32.
4. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology.* 2014;25:473–484.
5. Jackson JW, VanderWeele TJ. Decomposition analysis to identify intervention targets for reducing disparities. *Epidemiology.* 2018;29:825–835.
6. Jackson JW. On the interpretation of path-specific effects in health disparities research. *Epidemiology.* 2018;29:517–520.
7. Jackson JW. Explaining intersectionality through description, counterfactual thinking, and mediation analysis. *Soc Psychiatry Psychiatr Epidemiol.* 2017;52:785–793.

8. Jackson JW, VanderWeele TJ. Intersectional decomposition analysis with differential exposure, effects, and construct. *Soc Sci Med.* 2019;226:254–259.
9. Boonyasai RT, Rakotz MK, Lubomski LH, et al. Measure accurately, act rapidly, and partner with patients: an intuitive and practical three-part framework to guide efforts to improve hypertension control. *J Clin Hypertens (Greenwich).* 2017;19:684–694.
10. Balsa AI, Seiler N, McGuire TG, Bloche MG. Clinical uncertainty and healthcare disparities. *Am J Law Med.* 2003;29:203–219.
11. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. Conference on Uncertainty in Artificial Intelligence. July 2006. Cambridge, MA: AUAI Press, 2006;138–146.
12. Geneletti S. Identifying direct and indirect effects in a non-counterfactual framework. *J R Stat Soc Series B*. 2007;69:199–216.
13. James SA. Epidemiologic research on health disparities: some thoughts on history and current developments. *Epidemiol Rev.* 2009;31:1–6.
14. Williams DR, Lawrence JA, Davis BA. Racism and health: evidence and needed research. *Annu Rev Public Health.* 2019;40:105–125.
15. Keppel K, Pamuk E, Lynch J, et al. Methodological issues in measuring health disparities. *Vital Health Stat 2*. 2005;141:1–16.
16. Asada Y. Which Health Distributions are Inequitable? In: *Health Inequality: Mortality and Measurement*. University of Toronto Press; 2007;27–52.
17. Harper S, King NB, Meersman SC, Reichman ME, Breen N, Lynch J. Implicit value judgments in the measurement of health inequalities. *Milbank Q.* 2010;88:4–29.
18. Messer LC. Invited commentary: measuring social disparities in health–what was the question again? *Am J Epidemiol.* 2008;167:900–904; author reply 908.
19. Fabienne P, Evans T. The social basis of disparities in health. In: Evans T, Whitehead M, Diderichsen F, Bhuiya A, Wirth M, eds. *Challenging Inequities in Health: From Ethics to Action.* Oxford University Press; 2009;24–33.
20. Cook BL, McGuire TG, Zaslavsky AM. Measuring racial/ethnic disparities in health care: methods and practical issues. *Health Serv Res.* 2012;47(3 pt 2):1232–1254.
21. Duan N, Meng XL, Lin JY, Chen CN, Alegria M. Disparities in defining disparities: statistical conceptual frameworks. *Stat Med.* 2008;27:3941–3956.
22. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health.* 2006;27:167–194.
23. Duran DG, Pérez-Stable EJ. Novel approaches to advance minority health and health disparities research. *Am J Public Health.* 2019;109:S8–S10.
24. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academies Press; 2003.
25. Adler N. Appendix D. Overview of health disparities. In: Institute of Medicine, ed. *Examining the Health Disparities Research Plan of the National Institutes of Health: Unfinished Business*. The National Academies Press; 2006:121–174.
26. Krieger N. On the causal interpretation of race. *Epidemiology.* 2014;25:937.
27. Brook RH, Iezzoni LI, Jencks SF, et al. Symposium: case-mix measurement and assessing quality of hospital care. *Health Care Financ Rev.* 1987;Spec No:39–48.
28. Braithwaite RS. Risk adjustment for quality measures is neither binary nor mandatory. *JAMA.* 2018;319:2077–2078.
29. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006;60:578–586.
30. Jackson JW, Arah OA. Invited commentary: making causal inference more social and (Social) epidemiology more causal. *Am J Epidemiol.* 2020;189:179–182.
31. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol.* 2009;38:1599–1611.
32. Machado J, Mata J. Counterfactual decomposition of changes in wage distributions using quantile regression. *J Appl Econ*. 2005;20:445–465.
33. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods.* 2010;15:309–334.
34. Wang A, Arah OA. G-computation demonstration in causal mediation analysis. *Eur J Epidemiol.* 2015;30:1119–1127.
35. Lin SH, VanderWeele TJ. Interventional approach for path-specific effects. *J Causal Inference*. 2017;5:1–10.
36. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology.* 2017;28:258–265.
37. Sudharsanan N, Biljsma M. A Generalized Counterfactual Approach to Decomposing Differences Between Populations. Max Planck Institute for Demographic Research (MDPIDR) Working Papers and Preprints WP-2019-004. 2019:1–49.
38. Barsky R, Bound J, Charles KK, Lupton JP. Accounting for the black-white wealth gap: a nonparametric approach. *J Am Stat Assoc*. 2002;97:663–673.
39. Dinardo J, Fortin N, Lemieux T. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. *Econometrica*. 1996;64:1001–1044.
40. Huber M. Causal pitfalls in the decomposition of wage gaps. *J Bus Econ Stat*. 2015;33:179–191.
41. Hong G. Ratio of mediator probability weighting for estimating natural direct and indirect effects. Joint Statistical Meeting, August 2010. Vancouver, British Columbia: American Statistical Association; 2010:2401–2415.
42. Hong G, Deutsch J, Hill HD. Ratio-of-mediator-probability weighting for causal mediation in the presence of treatment-by-mediator interaction. *J Educ Behav Stat*. 2015;40:307–340.
43. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol.* 2012;176:190–195.
44. Miles CH, Shpitser I, Kanki P, Meloni S, Tchetgen Tchetgen EJ. On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika*. 2019;107:159–172.
45. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Stat Med.* 2013;32:4567–4580.
46. Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology.* 2014;25:300–306.
47. Zheng W, van der Laan M. Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *J Causal Inference*. 2017;5:20160006.
48. Zheng W, van der Laan MJ. Targeted maximum likelihood estimation of natural direct effects. *Int J Biostat*. 2012;8:1–40.
49. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992;3:143–155.
50. Pearl J. Direct and indirect effects. In: Breese K, Koller D, eds. *Uncertainty in Artificial Intelligence*. Morgan Kaufmann; 2001;411–420.
51. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics.* 2015;71:1–14.
52. Blinder A. Wage discrimination: reduced form and structural estimates. *J Hum Resour*. 1973;8:436.
53. Oaxaca R. Male-female wage differentials in urban labor markets. *Int Econ Rev*. 1973;14:693–709.
54. Nguyen TQ, Schmid I, Stuart EA. Clarifying causal mediation analysis for the applied researcher: defining effects based on what we want to learn [published online ahead of print July 16, 2020]. *Psychol Methods*. doi:10.1037/met0000299.
55. Fortin N, Lemieux T, Firpo S. Decomposition methods in economics. In: Ashenfelter O, Card D, eds. *Handbook of Labor Economics*. Elsevier; 2011;4:1–102.
56. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods.* 2013;18:137–150.
57. Breen R, Karlson KB, Holm A. Total, direct, and indirect effects in logit and probit models. *Sociol Methods Res*. 2013;42:164–191.
58. Huber M. Identifying causal mechanisms (primarily) based on inverse probability weighting. *J Appl Econ*. 2014;29:920–943.
59. Albert JM. Distribution-free mediation analysis for nonlinear models with confounding. *Epidemiology.* 2012;23:879–888.
60. Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol Methods*. 2012;1:130–158.
61. Jackson JW, Williams DR, VanderWeele TJ. Disparities at the intersection of marginalized groups. *Soc Psychiatry Psychiatr Epidemiol.* 2016;51:1349–1359.
62. Smith LH. Selection mechanisms and their consequences: understanding and addressing selection bias. *Curr Epidemiol Rep*. 2020: doi: 10.1007/s40471-020-00241-6, 1–11.