Published in final edited form as:
*Soc Sci Med*. 2019 April ; 226: 254–259. doi:10.1016/j.socscimed.2019.01.033.

# Intersectional Decomposition Analysis with Differential Exposure, Effects, and Construct

**John W. Jackson, ScD**(1),(2), **Tyler J. VanderWeele, PhD**(3),(4)

1)Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

2)Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

3)Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

4)Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

## Abstract

In recent years a wide array of proposals for bringing intersectional perspectives into quantitative studies of health disparities have appeared, from studies of interaction, predictive discrimination, to mediation. Bauer and Scheim, in a companion set of articles, extend these proposals by developing new attribution-blind measures of perceived discrimination and using VanderWeele's 3-way decomposition to quantify its contribution to disparities through differential exposure and differential effects (sometimes called differential vulnerability or susceptibility). In this commentary, after providing an overview of causal inference interpretations with social characteristics, we provide a broad overview of old and new decomposition methods in the social sciences literature and contrast their strengths and weaknesses for studying intersectional inequalities. We then examine how different forms of differential effects can be expressed within these decompositions and discuss their utility for the purpose of informing interventions for reducing disparities. Last, we discuss the tension in social sciences research when prominent explanatory variables represent constructs that are only defined or exist for certain marginalized populations and may not neatly fit within the decomposition methods framework. Through these discussions, we aim to provide greater conceptual clarity for applied researchers who are interested in using decomposition methods and other approaches to advance intersectional equity.

## Introduction

Following calls to incorporate intersectionality into quantitative analysis (Bauer, 2014; Bowleg, 2012), a wide array of methodological proposals have appeared (Bowleg and Bauer, 2016; Bright et al., 2016; Evans et al., 2017; Gustafsson et al., 2016; Jackson, 2017; Jackson et al., 2016; Wemrell et al., 2017; Yette and Ahern, 2018) In this Issue, Bauer and Scheim present companion papers that call for an application of causal mediation analysis, VanderWeele's 3-way decomposition (Vanderweele, 2013) to study the mediating and interactive role of perceived discrimination in producing differences in psychological distress across groups defined by multiple social characteristics, race and transgender

**Correspondence** Dr. John W. Jackson, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Room E-6543, Baltimore, MD, 21205; phone 443-287-5059; john.jackson@jhu.edu.

identity. While a formal framework for using causal mediation analysis and other decompositions to unpack intersectional disparities has already been outlined (Jackson, 2017), Bauer and Scheim add to this literature by developing new measures of perceived discrimination (Scheim and Bauer, 2019) and using the 3-way decomposition to study its contribution to intersectional disparities (Bauer and Scheim, 2019) through differential exposure and differential effects (sometimes in other literature referred to as differential vulnerability or susceptibility (Diderichsen et al., 2018)). In this commentary, we aim to: 1) review more nuanced causal interpretations with social characteristics; 2) review the shared features, strengths, and weaknesses of various forms of causal decomposition analysis for health disparities research; 3) examine how best to operationalize and distinguish differential effects from differential exposure under the goal of reducing disparities; 4) highlight critical identifiability issues when explanatory variables (mediators) such as perceived discrimination represent different constructs across social groups and suggest some ways forward.

## Causal Inference with Social Characteristics, Identities, and Positions

Historically, the application of causal inference methods for studying the effects of socially defined (and possibly non-manipulable) characteristics has been met with hesitancy by some (Holland, 1986; VanderWeele and Hernán, 2012) and embraced by others (Glymour and Glymour, 2014; Glymour and Spiegelman, 2017; Pearl, 2018). The potential outcomes framework underlying causal mediation analysis itself relies on a “consistency” condition wherein one’s observed outcome under some exposure must be the same as if it had been externally assigned. Those who are hesitant to use this framework work primarily from an interventionist model and are concerned that the many ways to change a social characteristic, if they exist at all, may each lead to a different outcome, so that the counterfactuals posed are too vague for use. Those who are more comfortable see causality as a chain of responses to determinant variables, but the precise quantitative interpretation is then often somewhat less clear (Vandenbroucke et al., 2016; VanderWeele, 2016).

This debate has yielded some important clarifications. When causal inference methods are applied to observational data to study disparities the goal is usually not to estimate the effect of a social characteristic so that we might intervene on it. If the outcome is some form of decision-making, as in medical treatment, then a causal effect of the *perceived* characteristic(s) can sometimes be interpreted as a measure of discrimination, and that measurement has scientific value for informing policy (Kaufman, 2008). Such effects can be identified by well-defined field experiments where the decision-maker’s perception is under control of the investigator, as are the person’s social characteristics and other variables (e.g., symptoms) so that there is no confounding (Greiner and Rubin, 2011). Though not without challenges, one could design experiments that directly measure discrimination across intersections of social characteristics (Goff and Kahn, 2014), perhaps with discrimination defined jointly across characteristics, and quantify the contribution of each (Jackson et al., 2016).

This experimental mapping does not, however, clearly translate when outcomes are health-related, which are determined by a complex myriad of factors, many of which are unknown

(Kaufman, 2008). Contrasts across social characteristics are meant to identify disparities (inequities) that are taken to be unfair, unjust, and avoidable, regardless of whether they arise through personally-mediated discrimination or structural forms of marginalization (Braveman, 2006). These are not necessarily taken to be causal effects of the social characteristic(s), but still do have counterfactual meaning and policy value in setting priorities for population health (Duan et al., 2008; Jackson, 2017). Descriptions of health outcomes across intersected social characteristics follow in this spirit (Jackson, 2017).

## A Broader View of Decomposition Analysis for Intersectional Disparities Research

In identifying disparate treatment or outcomes, the next step is to understand how they arise so as to identify leverage points for intervention and policy (Cooper et al., 2002). Naturally, the various causal interpretations for social characteristics has produced different causal estimands, analytic strategies, and interpretations. The economics literature, concerned with decision-related outcomes, gave rise to the Oaxaca-Blinder Decomposition (Blinder, 1973; Oaxaca, 1973). Epidemiology, concerned with health-related outcomes, gave rise to causal mediation analysis, using natural direct and indirect effects (Pearl, 2012; Robins and Greenland, 1992), and randomized interventional analogues (Didelez et al., 2006; Geneletti, 2007; VanderWeele et al., 2014; VanderWeele and Robinson, 2014) henceforth referred to as interventional effects. Each of these methods represent an advance over traditional mediation analysis methods (Baron and Kenny, 1986) as they explicitly allow for interaction between the social characteristic(s) and the mediator.

The Oaxaca-Blinder Decomposition was originally proposed to measure personally-mediated discrimination—effects of perceived characteristics—by including as explanatory variables all those factors that decision-makers use to determine outcomes. In its original use, the Oaxaca-Blinder Decomposition decomposes a marginal disparity with respect to all covariates, where the “unexplained portion” can be interpreted as a measure of discrimination. However, when it is applied to study health outcomes, the causal interpretation has been less clear because the causal process is more complex, and implementations typically ignore potential confounding of explanatory variables, among other issues (Jackson and VanderWeele, 2018).

Causal mediation analysis was developed to parse the effects of exposures into direct and indirect effects with explicit control of confounders of exposure and explanatory variables (mediators). Although mediation analysis might be interpreted as a decomposition of a social characteristic’s effect, the indirect effect estimate can also be interpreted as a disparity reduction, and the direct effect as a disparity residual, that would result from removing disparities in the explanatory mediating variable (VanderWeele and Robinson, 2014). This alternative framing leads to much weaker identifying conditions, namely no unmeasured confounding of the explanatory mediating variable, along with positivity and consistency. In this framework there is no need for confounding assumptions for the social characteristic because no causal effects are specified for it. Rather the causal effects are specified for the explanatory mediating variable and how the underlying disparity would change under interventions on that mediating variable.

With explanations of intersected characteristics, some of which may be time-varying and thus defined or measured later in life, such as the onset of a mental disorder or socioeconomic status in adulthood, this has very important implications for choosing confounders for adjustment. If a causal interpretation is given to the social characteristics, then their confounders, in addition to those of the explanatory mediating variable, must be adjusted for. When even measured confounders of one characteristic are affected by another characteristic, the methods proposed by Bauer & Scheim, including the suggested interventional effects currently developed in the literature, will not suffice. When there is treatment-confounder feedback of this type, conditioning on the confounders of the temporally latter characteristic (e.g., mental illness) can lead to a selection-bias, resulting in spurious associations between the earlier characteristic (e.g. race) and the outcome. Such a scenario is plausible, as there are racial differences in risk factors for mental illness. Also, when there are unmeasured confounders of the latter characteristic, some of the components needed to measure a statistical interaction will suffer from selection-bias. Figure 1 provides a graphical portrayal of these issues.

Even when causal interpretations are avoided for the social characteristics themselves and are only made with respect to the explanatory mediating variable, there is still a strong conceptual limitation with causal mediation analysis based on natural direct and indirect effects. It does not decompose a marginal disparity but rather a conditional one, and this may not be of substantive interest (Jackson and VanderWeele, 2018). If there are differences in confounding variables (e.g. childhood circumstances) across the social characteristic examined, then a conditional disparity only examines disparities in groups defined by the social characteristic among those with the same values of the confounders. Again, this may not be what is of interest.

In previous work, we showed that one can define interventional effects that control for confounders of explanatory variables, even those induced by a social characteristic, while explaining a marginal disparity of substantive interest (Jackson and VanderWeele, 2018). Whether the disparity itself, or the type of disparity removal, are conditional on any, some, or all covariates is a decision under the control of the investigator, not dictated by the method (Jackson, 2018). These interventional effects can be defined in ways that make them identical to Oaxaca-Blinder decompositions in some cases, identical to causal mediation analyses based on natural direct and indirect effects in others, or sometimes completely new decompositions (Jackson and VanderWeele, 2018). Accessible tutorials for estimating such interventional effects are still needed. However, some existing proposals for practical implementation can be adapted (VanderWeele et al., 2014; VanderWeele and Tchetgen Tchetgen, 2016; Zheng and Van Der Laan, 2017) provided they are substantively relevant.

These interventional effects offer a rich scaffolding to study intersectional disparities (Jackson, 2017). For example, suppose that, on the additive scale, we observe higher psychological distress for sexual/gender minority blacks as compared to sexual/gender majority whites, and this is larger than the sum of disparities for sexual/gender majority blacks and sexual/gender minority whites vs. sexual/gender majority whites. We have some evidence that the forms of marginalization for blacks and sexual/gender minorities combine in unique ways to impact psychological distress for those experiencing their intersection. Suppose that we carry out a decomposition analysis for each group vs. sexual/gender majority whites, where disparities in the explanatory mediating variable (e.g. educational opportunities) in each case are removed and find that the residual joint disparity is no longer greater than expected. We would have evidence that this variable plays a substantial role in explaining why multiple forms of marginalization uniquely pattern psychological distress for black sexual/gender minorities. Such intersectional mediation analyses can be insightful even when the initial joint disparity is not greater than expected (Jackson, 2017). Furthermore, if one's interest lies in understanding disparities between multiply marginalized groups, such as sexual/gender minority blacks, and other groups, such as sexual/gender majority blacks, one may of course pursue this line of inquiry.

## A Closer Look at the Differential Exposure vs. Differential Effects Paradigm

It is well recognized that explanatory mediating variables contribute to disparities when they are either unequally distributed across social characteristic categories (differential exposure) or when they have heterogeneous effects across social characteristic categories (differential effects) (Ward et al., 2018). While the nomenclature and conceptual mapping of differential effects varies across the literature, it could reflect differences in susceptibility or capacity of response by resources, coping, and adaptability, all of which can operate at individual and community levels (Diderichsen et al., 2018). In economics, differential effects have been conceptualized as reflecting disparate benefits in obtaining skills, qualifications, and other factors that affect, for example, employment and wages (Fortin et al., 2011). Thus, across fields, differential effects have been framed as unequal or differential vulnerability, susceptibility, or returns depending on the application (Diderichsen et al., 2018; Fortin et al., 2011). Along with Bauer & Scheim, there have been applied studies that leverage finer decompositions of disparities that pick out the contribution of differential effects to disparities (Hussein et al., 2018; Nordahl et al., 2014). These sorts of studies may be useful for enriching theory. Here, we will use the interventional effects framework to consider the utility of estimates of differential effects informing interventions to reduce disparities.

Before doing so, we take a moment to clarify how the construct of differential effects is expressed in the decomposition methods we reviewed so far, the Oaxaca-Blinder Decomposition and the various forms of causal mediation analysis. In the Oaxaca-Blinder Decomposition, the "unexplained portion" sums up the following components: 1) the disparity when all explanatory variables are observed to be at their referent values; 2) the difference in category-specific associations between each explanatory variable multiplied by its mean value among the referent social category. With causal mediation analysis under an interventional effects perspective, it follows from prior work that (Vanderweele, 2014) a disparity can be decomposed into: 1) a residual disparity under elimination of the explanatory variable from the population (or ubiquitously held to its baseline value), [analogue of the controlled direct effect (CDE)]; 2) the product between additive effect heterogeneity of the explanatory variable across social categories and its mean among the referent social category [analogue of the referent interaction ($INT_{ref}$)]; 3) the product between additive effect heterogeneity of the explanatory variable across social categories and the disparity in the explanatory variable [analogue of the mediated interaction ($INT_{med}$)]; 4) the change in the expected outcome among those in the referent (or privileged) social category if they had the explanatory variable distribution of those in the marginalized social category [analogue of the pure indirect effect (PIE)], each conditional on covariates not affected by the social categorization. Under two-way causal mediation analysis, the disparity reduction captures the sum of PIE and $INT_{med}$, and the disparity residual captures the sum of CDE and $INT_{ref}$. In spirit at least (Jackson and VanderWeele, 2018), the "unexplained portion" of the Oaxaca-Blinder Decomposition aligns with the "disparity residual" of the mediation analysis in that they both capture disparities when the explanatory variables are observed or set to zero or the baseline value (an analogue of the CDE) along with measures of additive association/effect heterogeneity multiplied by the explanatory variables' mean among the referent category (analogus to the $INT_{ref}$). Because Bauer and Scheim's proposal to use the 3-way decomposition, which decomposes the disparity into PDE, PIE, and $INT_{med}$, their construct of differential effects is not related to published work (Hussein et al., 2018) using the Oaxaca-Blinder Decomposition. Those papers focus on a quantity like $INT_{ref}$ which exists in the even in the absence of mediation, whereas Bauer and Scheim focus on $INT_{med}$ which only operates in the presence of mediation. Though Bauer and Scheim are not alone in choosing $INT_{med}$ (Nordahl et al., 2014), it is important to be precise about the construct under study, $INT_{ref}$ vs. $INT_{med}$, to aid comparison across studies and understand the implications of those results for reducing disparities, which we now turn to.

Researchers have rightly argued that a focus on differential exposure is an incomplete assessment of how an explanatory variable contributes to a disparity (Diderichsen et al., 2018; Ward et al., 2018). However, using $INT_{med}$ to measure the contribution of differential effects to disparities does not overcome this limitation. The reason for this is that part of differential effects' contribution to disparities occurs *through* differential exposure, and that part is represented by $INT_{med}$. It only captures how disparities in the explanatory variable exacerbate differences that arise through heterogeneous effects. A somewhat more technical explanation is that, for two-way decompositions, we ask how expected outcomes among the marginalized group would change and how the disparity itself would change if, contrary to fact, the distribution of the explanatory mediating variable for the marginalized group followed that of the privileged group. If we eliminate disparities in the explanatory variable, then we have as a side-effect eliminated the portion of differential effects that travels with it, $INT_{med}$. Put succinctly, it may be tempting to interpret estimates of $INT_{med}$ as how much a disparity would change if we addressed differential effects, but this is the wrong interpretation. $INT_{med}$ only reflects how much we would eliminate effects of differential effects by first addressing differential exposure. On this basis, then, it is the entire disparity reduction estimate and not $INT_{med}$ that is the preferred quantity for guiding intervention development.

Is there room for using the concept of differential effects and causal decomposition analyses to inform interventions? Certainly. It is important to understand why effects of explanatory variables differ for privileged and marginalized groups, as this may lead to a greater understanding of factors and conditions that interact with them to cause the outcome and should be accounted for when planning an intervention (VanderWeele, 2015). To some extent, the estimate of $\text{INT}_{\text{ref}}$ that does not depend on mediation can be helpful in evaluating additional gains if we could somehow eliminate differential effects. This suggests a different 3-way decomposition: CDE + $\text{INT}_{\text{ref}}$ + TIE or their interventional analogues. From causal theory in the potential outcomes framework (VanderWeele and Robins, 2007) we know that observing effect heterogeneity across levels of social characteristics could be due to many factors. It can be shown that such effect heterogeneity might be due to outright discrimination or due to the association between the social characteristic and another unmeasured factor that interacts with the explanatory variable to produce outcomes. This unmeasured variable's association with the social characteristic may be causal or the result of confounding or selection-bias. Either way, eliminating such effect heterogeneity would involve more empirical work to understand what that unmeasured factor is and to devise strategies to remove its association with the social characteristic or mitigate its impact on the outcome. A cautious and substantively rich interpretation of $\text{INT}_{\text{ref}}$ could provide the motivation to undertake this difficult but important work.

## Evaluating differential pathways and constructs as explanations

Bauer and Scheim's proposal to use perceived discrimination as an explanatory variable is important and provocative. Perceived discrimination is a salient example of how social exposures may "get under the skin" and affect health outcomes through psychosocial stress, allostatic load, perhaps also affecting health-related behaviors (Williams and Mohammed, 2013). However, working it into the analytic machinery of a decomposition analysis calls to attention some important limitations of that framework. These limitations are related to central tenets of intersectionality and must be reflected within quantitative investigations.

Intersectionality posits that marginalization may play out uniquely for those located at different intersections of social characteristics, identities, or positions (Crenshaw, 1991; Hill Collins, 2015). This could manifest as heterogeneous causal architectures for those at different intersections, where some causal pathways through certain explanatory variables are "switched on" for those at certain positions and "switched off" for others. This phenomenon, termed "switch intersectionality" (Bright et al., 2016), may depend on context, and is conceptually related to the construct of differential effects. Perceived discrimination may lead to psychological distress for some categories of intersected social characteristics, but perhaps not others. This differential effect could result in disparities in psychological distress in the absence of disparities in perceived discrimination (reflected in $\text{INT}_{\text{ref}}$), and its contribution to disparities in psychological distress could be exacerbated by the presence of disparities in perceived discrimination (reflected in $\text{INT}_{\text{med}}$). It represents an extreme form of differential effects because for some categories the effect of the explanatory mediating variable is non-null, whereas for others it is absent. Detecting and accounting for such heterogeneity of effects across intersections of social characteristic categories is entirely possible within decomposition analysis, but only when the construct has shared meaning.

Under another form of switch intersectionality where explanatory variables are only defined or manifested among certain intersected social characteristic categories, the conceptual meaning underlying a decomposition analysis disintegrates. Possibly because of this, Bauer and Scheim, in building their measures of discrimination, sought out questions that could plausibly be answered by persons at any intersection. By leaving out attributions specific to certain social categories, this will necessarily underestimate the explanatory role of perceived discrimination, but the constraint of shared constructs is required for meaningfully interpreting the results of a decomposition analysis. Bauer and Scheim do call for further validation of their measure, and it will be essential to demonstrate that even their attribution-blind measure of perceived discrimination is indeed shared across social categories. Decomposition analyses based on interventional effects ask us to envision what the residual disparity in psychological distress would be had say, black sexual/gender minorities, had the same distribution of perceived discrimination as say, white sexual/gender majorities. Experiencing the same levels of perceived discrimination on account of racial and sexual/gender identity micro- (or macro-) aggressions will likely have very different effects than perceived discrimination through anxiety over a loss of racial and/or sexual/gender identity privilege or suspicions of "reverse discrimination." While the decompositions we have considered here do allow for heterogeneous effects by accounting for statistical interactions between social categories and perceived discrimination in estimation procedures, it is not clear what the resulting quantities mean. Are we estimating outcomes for black sexual/gender minorities under the type of racial and sexual/gender identity anxiety that white sexual/gender majorities experience? Could we even do so? Perhaps for this reason so few have attempted mediation analyses with measures of perceived discrimination, even along the axis of a single social characteristic.

These issues of differential construct go beyond psychosocial measures. For example, with transgender persons undergoing hormone therapy, the quality of hormone therapy management will have a profound effect on health outcomes (Streed et al., 2017) and yet it is a construct that is only defined for this population. Similar issues with quality of care play out for persons with serious mental illness (Henderson, 2002). For cases of switch intersectionality via construct, one could restrict a decomposition analysis to subgroups where the construct is defined and shared, but this will not explain the entire joint disparity across social categories (Jackson et al., 2016). Another solution, one that would allow focus on the entire joint disparity, would be to use g-formula methods (Hernan and Robins, 2015) to estimate a form of population attributable effect: how would disparities in psychological distress change if we decreased the level of perceived discrimination among black sexual/gender minorities by a certain level. Certainly, one could envision reductions in perceived discrimination for one or more social categories, and even allow for different measures to be used across groups. The interpretations would be much more nuanced and careful than would be possible within the usual decomposition analysis framework.

## Conclusion

In summary, decomposition methods are a powerful analytic tool to understand how intersectional differences arise. The papers by Bauer and Scheim represent a bold and needed step to refine measures of perceived discrimination and other psychosocial stressors for comparative use across social categories and quantifying their contribution to disparities. As with most applications of causal methods, they represent but one expression of a larger set of possibilities. Clearly defining the causal question of substantive interest is essential for choosing among them. Whether or not investigators focus on a joint effect of social statuses or an intersectional disparity, which itself may be marginal or conditional, has implications for which confounders to adjust for and how they are handled in the analysis (Jackson and VanderWeele, 2018). Likewise, close attention to the form of differential effects, that which would disappear upon equalizing the explanatory variable ($INT_{med}$) vs. that which persists afterwards ($INT_{ref}$), has analytic implications as well, leading to alternative 3-way decompositions. Furthermore, great care must be taken to consider whether the explanatory variable's construct has shared meaning across the compared categories. Attending to these considerations can help produce estimates that are more interpretable and actionable for addressing intersectional disparities.

## Acknowledgements

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number K01HL145320 (Jackson) and the National Cancer Institute Award Number 1R01CA222147-01 (VanderWeele)

## References

Baron RM, Kenny DA, 1986 The moderator-mediator distinction in social psychological research: Conceptual, strategic and statistical considerations. J. Pers. Soc. Psychol 51, 1173–1182. [PubMed: 3806354]

Bauer GR, 2014 Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. Soc. Sci. Med 110, 10–17. 10.1016/j.socscimed.2014.03.022 [PubMed: 24704889]

Bauer GR, Scheim AI, 2019 Methods for analytic intercategorical intersectionality in quantitative research: Discrimination as a mediator of health disparities. Soc. Sci. Med (in press).

Blinder AS, 1973 Wage Discrimination: Reduced Form and Structural Estimates. J. Hum. Resour 8, 436 10.2307/144855

Bowleg L, 2012 The problem with the phrase women and minorities: Intersectionality-an important theoretical framework for public health. Am. J. Public Health 102, 1267–1273. 10.2105/AJPH.2012.300750 [PubMed: 22594719]

Bowleg L, Bauer G, 2016 Invited Reflection: Quantifying Intersectionality. Psychol. Women Q 40, 337–341. 10.1177/0361684316654282

Braveman P, 2006 Health disparities and health equity: concepts and measurement. Annu. Rev. Public Health 27, 167–94. 10.1146/annurev.publhealth.27.021405.102103 [PubMed: 16533114]

Bright LK, Malinsky D, Thompson M, 2016 Causally Interpreting Intersectionality Theory. Philos. Sci 83.

Cooper LA, Hill MN, Powe NR, 2002 Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. J. Gen. Intern. Med 17, 477–86. [PubMed: 12133164]

Crenshaw K, 1991 Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. Stanford Law Rev. 43, 1241 10.2307/1229039

Didelez V, Dawid AP, Geneletti S, 2006 Direct and Indirect Effects of Sequential Treatments, in: Detcher R, Richardson T (Eds.), Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence AUIA Press, pp. 138–146.

Diderichsen F, Hallqvist J, Whitehead M, 2018 Differential vulnerability and susceptibility: how to make use of recent development in our understanding of mediation and interaction to tackle health inequalities. Int. J. Epidemiol (in press). 10.1093/ije/dyy167

Duan N, Meng X-L, Lin JY, Chen C, Alegria M, 2008 Disparities in defining disparities: statistical conceptual frameworks. Stat. Med 27, 3941–56. 10.1002/sim.3283 [PubMed: 18626925]

Evans CR, Williams DR, Onnela JP, Subramanian SV, 2017 A multilevel approach to modeling health inequalities at the intersection of multiple social identities. Soc. Sci. Med 0–1 10.1016/j.socscimed.2017.11.011

Fortin N, Lemieux T, Firpo S, 2011 Decomposition Methods in Economics, in: Handbook of Labor Economics. pp. 1–102. 10.1016/S0169-7218(11)00407-2

Geneletti S, 2007 Identifying Direct and Indirect Effects in a Non-Counterfactual Framework. J. R. Stat. Soc. Ser. B 69, 199–215.

Glymour C, Glymour MR, 2014 Commentary: Race and sex are causes. Epidemiology 25, 488–490. 10.1097/EDE.0000000000000122 [PubMed: 24887161]

Glymour MM, Spiegelman D, 2017 Evaluating public health interventions: 5. Causal inference in public health research-do sex, race, and biological factors cause health outcomes? Am. J. Public Health 107, 81–85. 10.2105/AJPH.2016.303539 [PubMed: 27854526]

Goff PA, Kahn KB, 2014 How psychological science impedes intersectional thinking. Du Bois Rev. 10, 365–384.

Greiner DJ, Rubin DB, 2011 Causal Effects of Perceived Immutable Characteristics. Rev. Econ. Stat 93, 775–785. 10.1162/REST_a_00110

Gustafsson PE, Sebastián MS, Mosquera PA, 2016 Meddling with middle modalities: a decomposition approach to mental health inequalities between intersectional gender and economic middle groups in northern Sweden. Glob. Health Action 9, 32819 10.3402/gha.v9.32819 [PubMed: 27887668]

Henderson DC, 2002 Diabetes mellitus and other metabolic disturbances induced by atypical antipsychotic agents. Curr. Diab. Rep 2, 135–40. [PubMed: 12643133]

Hernán MA, Hernández-Díaz S, Robins JM, 2004 A structural approach to selection bias. Epidemiology 15, 615–25. [PubMed: 15308962]

Hernan MA, Robins JM, 2015 Longitudinal causal inference, in: International Encyclopedia of the Social & Behavioral Sciences. Elsevier, Oxford, England, pp. 340–344.

Hill Collins P, 2015 Intersectionality’s Definitional Dilemmas. Annu. Rev. Sociol 41, 1–20. 10.1146/annurev-soc-073014-112142

Holland PW, 1986 Statistics and Causal Inference. J. Am. Stat. Assoc 81, 945–960. 10.1080/01621459.1986.10478354

Hussein M, Diez Roux AV, Mujahid MS, Hastert TA, Kershaw KN, Bertoni AG, Baylin A, 2018 Unequal Exposure or Unequal Vulnerability? Contributions of Neighborhood Conditions and Cardiovascular Risk Factors to Socioeconomic Inequality in Incident Cardiovascular Disease in the Multi-Ethnic Study of Atherosclerosis. Am. J. Epidemiol 187, 1424–1437. 10.1093/aje/kwx363 [PubMed: 29186311]

Jackson JW, 2018 On the Interpretation of Path-specific Effects in Health Disparities Research. Epidemiology 29, 517–520. 10.1097/EDE.0000000000000843 [PubMed: 29642085]

Jackson JW, 2017 Explaining intersectionality through description, counterfactual thinking, and mediation analysis. Soc. Psychiatry Psychiatr. Epidemiol 52, 785–793. 10.1007/s00127-017-1390-0 [PubMed: 28540515]

Jackson JW, VanderWeele TJ, 2018 Decomposition analysis to identify intervention targets for reducing disparities. Epidemiology (in press). 10.1097/EDE.0000000000000901

Jackson JW, Williams DR, VanderWeele TJ, 2016 Disparities at the intersection of marginalized groups. Soc. Psychiatry Psychiatr. Epidemiol 51, 1349–1359. 10.1007/s00127-016-1276-6 [PubMed: 27531592]

Kaufman JS, 2008 Epidemiologic analysis of racial/ethnic disparities: Some fundamental issues and a cautionary example. Soc. Sci. Med 66, 1659–1669. 10.1016/j.socscimed.2007.11.046 [PubMed: 18248866]

Nordahl H, Lange T, Osler M, Diderichsen F, Andersen I, Prescott E, Tjønneland A, Frederiksen BL, Rod NH, 2014 Education and cause-specific mortality: The mediating role of differential exposure and vulnerability to behavioral risk factors. Epidemiology 25, 389–396. 10.1097/EDE.0000000000000080 [PubMed: 24625538]

Oaxaca R, 1973 Male-Female Wage Differentials in Urban Labor Markets. Int. Econ. Rev. (Philadelphia). 14, 693 10.2307/2525981

Pearl J, 2018 Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes. J. Causal Inference 6, e1–7. 10.1515/jci-2018-2001

Pearl J, 2012 The Causal Mediation Formula-A Guide to the Assessment of Pathways and Mechanisms. Prev. Sci. 13, 426–436. 10.1007/s11121-011-0270-1 [PubMed: 22419385]

Robins JM, Greenland S, 1992 Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143–55. [PubMed: 1576220]

Scheim AI, Bauer GR, 2019 The Intersectional Discrimination Index: Development and validation of measures of self-reported enacted and anticipated discrimination for intercategorical analysis. Soc. Sci. Med (in press).

Streed CG, Harfouch O, Marvel F, Blumenthal RS, Martin SS, Mukherjee M, 2017 Cardiovascular disease among transgender adults receiving hormone therapy: A narrative review. Ann. Intern. Med 167, 256–267. 10.7326/M17-0577 [PubMed: 28738421]

Vandenbroucke JP, Broadbent A, Pearce N, 2016 Causality and causal inference in epidemiology: the need for a pluralistic approach. Int. J. Epidemiol 45, 1776–1786. 10.1093/ije/dyv341 [PubMed: 26800751]

Vanderweele TJ, 2014 A unification of mediation and interaction: A 4-way decomposition. Epidemiology 25, 749–761. 10.1097/EDE.0000000000000121 [PubMed: 25000145]

Vanderweele TJ, 2013 A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology 24, 224–232. 10.1097/EDE.0b013e318281a64e [PubMed: 23354283]

VanderWeele TJ, 2016 Commentary: On Causes, Causal Inference, and Potential Outcomes. Int. J. Epidemiol 45, 1809–1816. 10.1093/ije/dyw230 [PubMed: 28130319]

VanderWeele TJ, 2015 Explanation in Causal Inference: Methods for Mediation and Interaction, 1st ed. Oxford Univeristy Press, New York, NY.

VanderWeele TJ, Hernán MA, 2012 Causal Effects and Natural Laws: Towards a Conceptualization of Causal Counterfactuals for Nonmanipulable Exposures, with Application to the Effects of Race and Sex, in: Berzuini C, David P, Bernardinelli L (Eds.), Causality: Statistical Perspectives and Applications. John Wiley & Sons, Ltd, pp. 101–113. 10.1002/9781119945710.ch9

VanderWeele TJ, Robins JM, 2007 Four types of effect modification: a classification based on directed acyclic graphs. Epidemiology 18, 561–8. 10.1097/EDE.0b013e318127181b [PubMed: 17700242]

VanderWeele TJ, Robinson WR, 2014 On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology 25, 473–84. 10.1097/EDE.0000000000000105 [PubMed: 24887159]

VanderWeele TJ, Tchetgen Tchetgen EJ, 2016 Mediation analysis with time varying exposures and mediators. J. R. Stat. Soc. Ser. B (Statistical Methodol 168, 1–22. 10.1111/rssb.12194

VanderWeele TJ, Vansteelandt S, Robins JM, 2014 Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder. Epidemiology 25, 300–306. 10.1097/EDE.0000000000000034 [PubMed: 24487213]

Ward JB, Gartner DR, Keyes KM, Fliss MD, McClure ES, Robinson WR, 2018 How do we assess a racial disparity in health? Distribution, interaction, and interpretation in epidemiological studies. Ann. Epidemiol 1–7. 10.1016/j.annepidem.2018.09.007

Wemrell M, Mulinari S, Merlo J, 2017 Intersectionality and risk for ischemic heart disease in Sweden: Categorical and anti-categorical approaches. Soc. Sci. Med 177, 213–222. 10.1016/j.socscimed.2017.01.050 [PubMed: 28189024]

Williams DR, Mohammed SA, 2013 Racism and Health I. Am. Behav. Sci 57, 1152–1173. 10.1177/0002764213487340

Yette EM, Ahern J, 2018 Health-related Quality of Life Among Black Sexual Minority Women. Am. J. Prev. Med 55, 281–289. 10.1016/j.amepre.2018.04.037 [PubMed: 30122211]

Zheng W, Van Der Laan M, 2017 Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes. J. Causal Infer 10.1515/jci-2016-0006

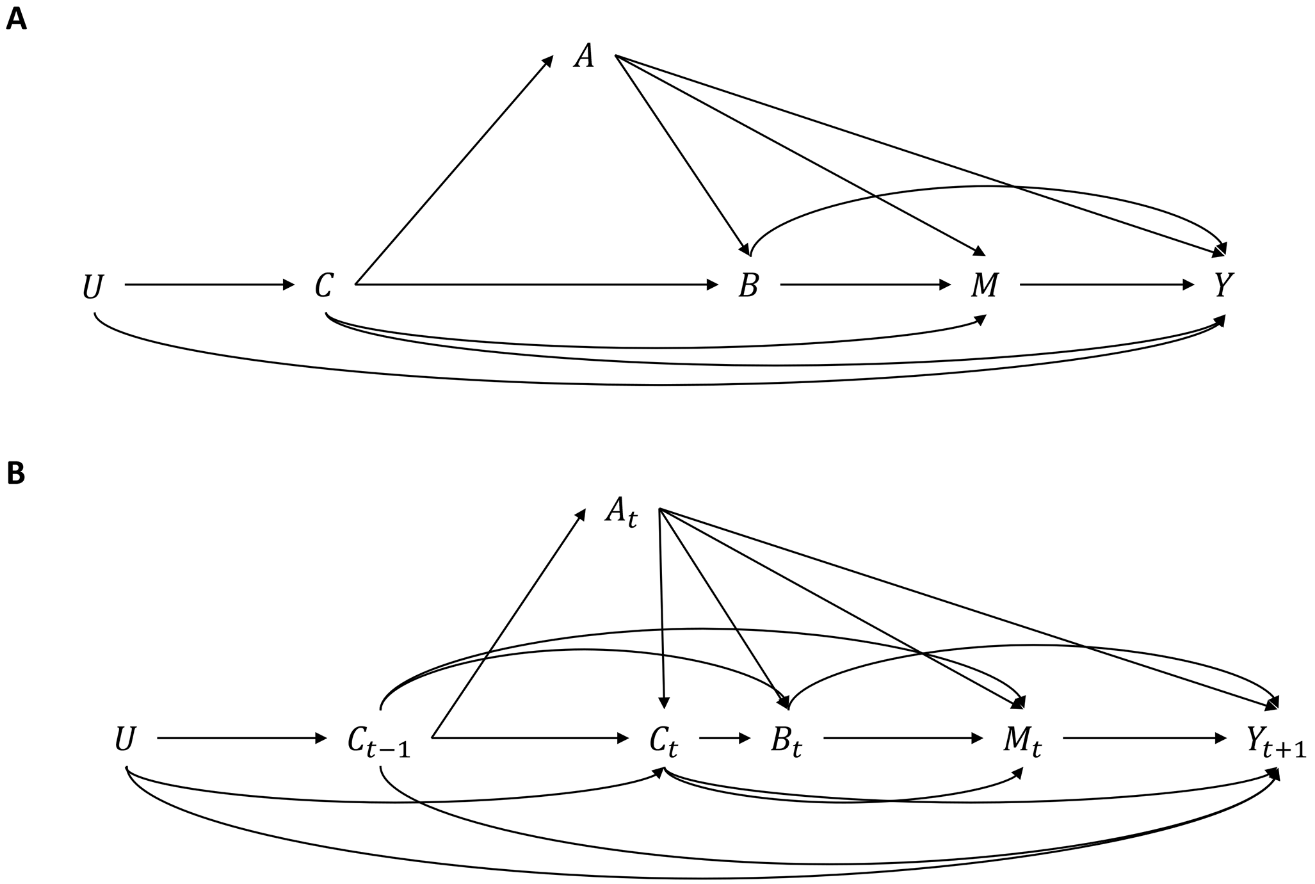

**Figure 1.**

Causal diagram depicting causal relationships between social characteristics $A$ and $B$ and an outcome $Y$ mediated by some construct $M$, where measured common causes (confounders) of these variables are denoted by $C$ and unmeasured common causes are denoted by $U$. Note that, more generally, $C$ can still be considered a confounder (by proxy) when it does not cause each of the other variables $A$, $B$, $M$ and $Y$ but is merely associated through another perhaps unmeasured variable that does. In panel A (top), the social characteristics are time-fixed or defined early in life, with $C$ representing factors in early life. Here, mediation analysis methods based on natural direct and indirect effects condition on $C$ to control for confounding of the $A – M$, $A – Y$, $B – M$, $B – Y$, and $M – Y$ relationships. However, with interventional effects, confounding of the $M – Y$ relationship is achieved by a form of standardization, and whether or not $C$ is conditioned on is optional (Jackson and VanderWeele, 2018). In panel B (bottom), the social characteristics are not necessarily fixed in early life and may be defined in adulthood, and this is represented by indexing all variables according to a time-specific measurement denoted by $t$. Here, mediation analysis methods based on natural direct and indirect effects condition the estimands (direct and indirect effects of $A$ and $B$) on $C_{t−1}$ and $C_t$ to adjust for all of the $A_t – M_t$, $A_t – Y_{t+1}$, $B_t – M_t$, $B_t – Y_{t+1}$, and $M_t – Y_{t+1}$ relationships. However, $C_t$ is a descendant of $A$ and $U$ and when it is conditioned on, a selection bias arises (Hernán et al., 2004) between $A_t$ and $Y_{t+1}$ via the path $A_t \rightarrow C_t \rightarrow U \leftarrow Y_{t+1}$. With interventional effects, control of $C_{t−2}$ and $C_t$ can be accomplished by a form of standardization without conditioning the estimand (i.e., a joint disparity (Jackson et al., 2016)) on $C_{t-1}$ or $C_t$, so that there is no selection-bias. Though formulae for non-parametric decompositions of this type have not been derived, it would be straightforward to do so following proposals from earlier work (Jackson, 2017; Jackson and VanderWeele, 2018). Finally, a subtle issue even with interventional effects may occur when one social identity affects another: the excess intersectional disparity (a statistical interaction) can suffer from selection-bias when $U$ affects or is associated with $B_t$. Estimating the excess intersectional disparity requires estimating disparities in $A_t$ conditional on $B_t$. This conditioning induces a selection-bias between $A_t$ and $Y_{t+1}$ through the path $A \rightarrow B_t \leftarrow U \rightarrow Y_{t+1}$.