Decomposition Analysis to Identify Intervention Targets for Reducing Disparities

John W. Jackson,[a,b,c] *and Tyler J. VanderWeele*[c,d]

Abstract: There has been considerable interest in using decomposition methods in epidemiology (mediation analysis) and economics (Oaxaca–Blinder decomposition) to understand how health disparities arise and how they might change upon intervention. It has not been clear when estimates from the Oaxaca–Blinder decomposition can be interpreted causally because its implementation does not explicitly address potential confounding of target variables. While mediation analysis does explicitly adjust for confounders of target variables, it typically does so in a way that effectively entails equalizing confounders across racial groups, which may not reflect the intended intervention. Revisiting prior analyses in the National Longitudinal Survey of Youth on disparities in wages, unemployment, incarceration, and overall health with test scores, taken as a proxy for educational attainment, as a target intervention, we propose and demonstrate a novel decomposition that controls for confounders of test scores (e.g., measures of childhood socioeconomic status [SES]) while leaving their association with race intact. We compare this decomposition with others that use standardization (to equalize childhood SES [the confounders] alone), mediation analysis (to equalize test scores within levels of childhood SES), and one that equalizes both childhood SES and test scores. We also show how these decompositions, including our novel proposals, are equivalent to implementations of the Oaxaca–Blinder decomposition but provide a more formal causal interpretation for these decompositions.

Keywords: Causal inference, Health disparity, Inequity, Interventional effects, Mediation analysis, Path analysis, Path-specific effect, Oaxaca Blinder decomposition, Standardization

(*Epidemiology* 2018;29: 825–835)

Submitted March 17, 2017; accepted July 24, 2018.

From the [a]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [b]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; [c]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; and [d]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

Data and Code Availability: https://github.com/jwjackson/SuppMaterials.

John Jackson was partly funded by the Alonzo Smythe Yerby Fellowship, and Tyler J. VanderWeele was funded by the National Institutes of health grant ES017876.

The authors report no conflicts of interest.

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, Departments of Epidemiology and Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. E-mail: john.jackson@jhu.edu.

ISSN: 1044-3983/18/2906-0825
DOI: 10.1097/EDE.0000000000000901

Health disparities are differences in health between socially advantaged versus disadvantaged groups that are considered unnecessary and unjust.[1] Although national and local efforts have sought to reduce and eliminate racial/ethnic disparities in health over the past few decades, they have often persisted.[2,3] Reducing racial/ethnic disparities requires that we understand how they arise and develop interventions to target the mechanisms that perpetuate them.[4]

Consider the seminal analyses by Fryer[5] in the National Longitudinal Survey of Youth (NLSY), patterned after analyses by Neal and Johnson,[6] which evaluated whether control for test score percentiles from the Armed Forces Qualifying Test, a measure of premarket skills, eliminated observed racial/ethnic disparities in wages, unemployment, incarceration, and health. Black–white differences in log wages decreased by 72%, and the disparity in self-reported health vanished. Though these analyses were originally motivated to statistically detect racial/ethnic discrimination, we are often tempted to interpret the results causally, that is, are these disparities driven by disparities in education? To answer this question, one would want to adjust for potential confounding by measures of familial socioeconomic status (SES) in childhood.[1,7] Otherwise, the disparity reductions might reflect the effect of equalizing childhood SES rather than just test scores. Such confounding, if present, could limit the value of these results for evidence-based policy to reduce disparities by targeting disparities in education.

More broadly, one might consider how disparities in wages and other outcomes would change by eliminating disparities in childhood SES versus disparities in test scores. While the queries involve ambitious, hypothetical interventions that could take many forms the results would help motivate and prioritize future work. Using the potential outcomes framework, one can examine how disparities might change by intervening on a target.[8] One might use standardization to examine how disparities might change upon equalizing the childhood SES distribution across race or use mediation analysis to examine how disparities might change upon equalizing the test score distribution across race among those with the same childhood SES.[9] One might even consider a joint intervention to equalize both childhood SES and test scores. But neither standardization nor mediation analysis examines how disparities in adult outcomes might change if we removed

disparities in educational test scores but not childhood SES.[10] If disparities could be considerably reduced through education, then even if we cannot intervene directly on early childhood SES, this could still lead to encouraging policy implications.

In a reanalysis of the Fryer data, we outline and demonstrate a novel decomposition method that estimates how well removing disparities in test scores, but not childhood SES, reduces disparities in adulthood outcomes. We compare these results to estimates under interventions to equalize childhood SES alone (standardization), to equalize test scores among children of the same SES (mediation analysis), and to equalize childhood SES and test scores together. Unlike these other approaches, our method appropriately adjusts for the confounding effects of childhood SES on test scores while leaving its association with race intact. Conceptually, it maps to a randomized trial of an intervention on test scores where the estimand is the association between race and the outcome.[11] In the main text, we implement the method using formulae derived under linear models for the outcome. In the Appendices, we provide formulae for nonlinear models and also for nonparametric formulae. There, we show how standardization, mediation analysis, and our decomposition can be expressed as Oaxaca–Blinder decompositions[12,13] under certain conditions.

We have kept the exposition in the text at a conceptual level. More technical readers may want to move directly to the formal results given in the Appendices where we describe how our results connect to the literature of causal mediation analysis and the Oaxaca–Blinder decomposition.

EXAMPLE DATA AND STATISTICAL ANALYSES

Our motivating example revisits analyses by Fryer.[5] Replicating those data, we extracted baseline and outcome data from the NLSY on black and white men in the United States who, at baseline, were of ages 14 to 22 years in the 1979 cohort[14] (NLSY79) and ages 12 to 16 years in the 1997 cohort[15] (NLSY97). We used baseline data to define indicators of ascertained gender and race (1, black; 0, white), ethnicity (1, Hispanic; 0, non-Hispanic), and mixed race (1, mixed race; 0, single race; NLSY97 only). Hourly wage in 2006 US dollars was calculated as a weighted average across all current jobs in 2006 or 2007 (with proportion of total hours/week per job as weights), excluding possibly implausible wages below $1 or above $115 per hour, and log-transformed. Unemployment was coded as a binary variable from current employment status in 2006, with individuals not in the labor market coded as missing. Incarceration was coded as a binary variable indicating self-reported residence in jail for any follow-up survey through 2006 or having been sentenced to a correctional institution before baseline. Self-reported health (only measured in NLSY79) was recoded from the physical component score from the 12-item Short Form Health Survey in 2006 and converted to a z-score. These represent outcomes in 2006 or 2007 at ages 42 to 44 years for the NLSY79 cohort and ages 22 to 27 years for the NLSY97 cohort. Test score percentiles were obtained for the NLSY79 cohort from the Armed Forces Qualification Test—the sum of the arithmetic reasoning score, the mathematics knowledge score, and two times the verbal composite score—which was administered as part of the Armed Services Vocational Aptitude Battery as reported in the 1981 survey year. For NLSY97, the Armed Forces Qualification Test percentiles obtained were based on a similarly constructed (but unofficial) score from the 1999 survey year. Scores were standardized by age within the NLSY79 cohort, and also within the NLSY97 cohort, as described elsewhere.[5] Total years of education before 2006 or 2007 was also extracted, bottom-coded at 8 years and top-coded at 16 years. Measures of childhood SES in the NLSY79 cohort included maternal educational attainment (highest grade completed), household income, and poverty status as assessed in 1979. For the NLSY97 cohort, we used the same measures of childhood SES (replacing poverty status with household net worth), which were assessed in 1997 or 1998. Missing indicators were constructed for test scores, total years of education, and all measures of childhood SES. See Fryer[5] for more details. The characteristics of the NLSY79 are described in Table 1 and the NLSY97 in eTable 1; http://links.lww.com/EDE/B387. Our study was a secondary analysis of de-identified publicly available data and thus did not undergo ethical review by an Institutional Review Board.

In the main text, we present formulae under models that rely on a single measure of childhood SES because the formulae are more intuitive. Nonetheless, the disparity estimates reported in Tables 2 and 3 were obtained under models that relied on three separate measures of childhood SES (see page 19 of the eAppendix; http://links.lww.com/EDE/B387). Replicating Fryer,[5] all models included mutually exclusive dummy variables for Hispanic ethnicity and mixed race (for NLSY97 only), as well as missing indicators for education and childhood SES variables. The nonparametric bootstrap with 1,000 replication samples was used to obtain standard errors. The proportion of the disparity reduced was estimated on the additive scale (see page 26 of the eAppendix; http://links.lww.com/EDE/B387). Note that for the NLSY79 cohort, we do not provide results for unemployment as there were insufficient cases for reliable estimates under the bootstrap.

THE STRUCTURE OF EXTANT DISPARITIES IN ADULTHOOD

Figure A portrays the structure of disparities as relationships between racial classification; henceforth race R (1, black; 0, white), childhood SES X, test scores M, an outcome Y, covariates gender and age C, and historical processes H such as slavery and Jim Crow that are responsible for black–white differences in SES and residence at conception.[16] The diagram could be detailed further by breaking the race node R into features that investigators might consider under their

TABLE 1. Characteristics of Males in the 1979 National Survey of American Youth Analytic Cohort, Mean (SD)

	White (n = 1,010)	Black (n = 597)
Age (years)	43.1 (0.8)	43.1 (0.8)
Adulthood outcomes		
Wage (dollars/hour)	$26.1 ($17.4)	$17.4 ($12.2)
Unemployed[a]	3.6 (18.6)	8.0 (27.2)
Incarceration, ever[a]	7.4 (26.2)	22.1 (41.5)
Health[b] (z score)	0.15 (0.8)	0.3 (1.0)
Measures of educational attainment		
AFQT (z score)	0.45 (1.0)	−0.58 (0.8)
Total years education (years)	13.3 (2.1)	12.6 (1.8)
Measures of childhood socioeconomic status		
Mother's highest grade level	11.9 (2.4)	10.9 (2.5)
Poverty status in childhood[a]	9.6 (29.5)	48.7 (50.0)
Household income in childhood (dollars)	$21,466 ($12,854)	$10,835 ($7799)
Missingness		
Missing AFQT[a]	3.9 (19.3)	2.2 (14.6)
Missing total years of education[a]	25.6 (43.7)	23.3 (42.3)
Missing mother's highest grade level[a]	5.3 (22.5)	10.2 (30.3)
Missing poverty status in childhood[a]	9.4 (39.1)	4.4 (20.4)
Missing household income in childhood[a]	18.8 (39.1)	17.3 (37.8)

[a]Binary variable (1 = yes, 0 = no), scaled by 100. E.g., 3.6% of NLSY79 whites were unemployed in 2006.

[b]Self-reported health assessed as a z-score from the physical component subdomain of the 12-item Short Form Health Survey.

AFQT indicates Armed Forces Qualifying Test; NLSY, National Longitudinal Survey of Youth.

study.[9] We retain a general race node R because our results apply to any definition of race that investigators use to operationalize the construct of race. A technical articulation of this diagram and its substantive interpretation are provided on page 7 of the eAppendix; http://links.lww.com/EDE/B387. The formal results in the Appendix are, however, given without reference to any particular causal diagram, so the diagram is presented here for intuition only.

In Figure A, the racial disparity in outcome *Y* arises in several ways. The disparity arises through backdoor paths involving history *H*: the effects of Jim Crow have been that blacks are more likely to be born into families with low SES who live in neighborhoods with lower quality schools[16] (a nonmediating path—it does not capture the effect of race on an intermediate variable *M*). Forward paths emanating from the race node, which include the mediating path through the intermediate variable *M*, could represent effects of discrimination. For example, blacks are more likely to be placed into less rigorous curriculum tracks in early education, and the effects of this accumulate, e.g., mathematics course choice in high school.[17,18] The direct path comprises all forward pathways that do not operate through the intermediate variable *M*.

DISPARITY REDUCTIONS UNDER ALTERNATIVE INTERVENTION STRATEGIES

We now describe results from decompositions that estimate how well certain interventions might reduce racial disparities in adulthood (wages, unemployment, incarceration, and health) by equalizing childhood SES and/or test scores across race. Each intervention is equivalent to deactivating certain paths linking race to adult outcomes. (For example, abolishing the effect of race *R* on test scores *M* would deactivate the path $R \rightarrow M \rightarrow Y$, among others.) Heterogeneous effects of test scores *M* across race could persist and contribute to the residual disparity.

To identify the disparity reductions with observational data, we assume (A1) the effect of childhood SES on an outcome is unconfounded given race and covariates (such as gender and age); (A2) the effect of test scores on an outcome is unconfounded given race, childhood SES, and covariates. We also assume covariate overlap among each racial group (positivity), and that one's observed outcome under the actual value of a target variable equals the outcome that would be observed upon intervening to set the target variable to that value (consistency). In the main text, we assume the absence of statistical interactions in the linear models for simplicity, though such expressions as Oaxaca–Blinder decompositions and nonparametric formulae in the Appendices allow for such interactions. See page 23 of the eAppendix; http://links.lww.com/EDE/B387 for formal statements.

We report estimates for the initial disparity, the residual after each intervention, and the corresponding reduction along with their standard errors in Table 2 (for test scores) and Table 3 (for total years of education) for the NLSY79 cohort. Our narrative focuses on log wages to introduce the method and later summarizes results for incarceration and self-reported health. Results for the NLSY97 cohort are provided in the Appendix. With the exception of Propositions 1 and 2, which were introduced in VanderWeele and Robinson 2014,[9] all others are novel.

Proposition 1: Intervene to Equalize Childhood SES Across Race

The first proposal is to randomly assign childhood SES among blacks such that it follows the distribution among whites of the same gender and age. We posit that childhood SES reflects conditions near the time of conception and thus link race with adult outcomes through a backdoor path, e.g., $R \leftarrow H \rightarrow X \rightarrow Y$ in Figure A. VanderWeele and Robinson[9] provided analytic formulae for the residual disparity under equalizing a nonmediating (i.e., a nonintermediate) variable such as childhood SES. These formulae require that assumption A1 holds. Provided this holds, we can estimate the residual disparity by fitting linear models that condition on race *R* and covariates gender and age *C* (1) and additionally childhood SES *X* (2):

$$E[Y \mid r,c] = \phi_0 + \phi_1 r + \phi_4' c \quad (1)$$

$$E[Y \mid r,x,c] = \gamma_0 + \gamma_1 r + \gamma_2 x + \gamma_4' c \quad (2)$$

TABLE 2. Estimates of Residual Disparities and Disparity Reductions in Adult Outcomes Under Hypothetical Intervention Strategies on Childhood SES Measures and/or Armed Forces Qualifying Test Scores in the 1979 NLSY Cohort[a]

	Proposition 1: Intervene to equalize the distribution of childhood SES measures across race but not AFQT scores	Proposition 2: Intervene to equalize the distribution of AFQT scores across race within levels of childhood SES	Proposition 3: Intervene to equalize the distribution of AFQT scores and childhood SES measures across race	Proposition 4: Intervene to equalize the distribution of AFQT scores across race but not childhood SES measures	Reanalysis of Fryer: Statistically equalize the distribution of AFQT scores across race without control for childhood SES
Log wages, mean difference (SE)					
Initial disparity	−0.41 (0.04)	−0.30 (0.05)	−0.41 (0.04)	−0.41 (0.04)	−0.41 (0.04)
Residual disparity	−0.30 (0.05)	−0.11 (0.05)	−0.11 (0.05)	−0.14 (0.10)	−0.13 (0.05)
% reduction	26	65	74	66	69
Incarceration (risk ratio, SE)					
Initial disparity	3.54 (1.17)	2.39 (1.21)	3.54 (1.17)	3.54 (1.17)	3.54 (1.17)
Residual disparity	2.39 (1.21)	1.49 (1.21)	1.49 (1.21)	1.86 (1.25)	1.76 (1.19)
% reduction	45	65	81	65	70
Health[b] (mean difference, SE)					
Initial disparity	−0.14 (0.05)	−0.04 (0.06)	−0.14 (0.05)	−0.14 (0.05)	−0.14 (0.05)
Residual disparity	−0.04 (0.06)	0.05 (0.07)	0.05 (0.07)	−0.02 (0.07)	0.02 (0.06)
% reduction	75	251	137	85	112

[a]The analytic sample size was 1154 for wages, 1988 for incarceration, and 1587 for health. All models included a mutually exclusive dummy variable for Hispanic ethnicity.
[b]Self-reported health assessed as a z score from the physical component subdomain of the 12-item Short Form Health Survey.
AFQT indicates Armed Forces Qualifying Test, SES socioeconomic status, SE standard error.

TABLE 3. Estimates of Residual Disparities and Disparity Reductions in Adult Outcomes Under Hypothetical Intervention Strategies on Childhood SES Measures and/or Total Years of Education in the 1979 NLSY Cohort[a]

	Proposition 1: Intervene to equalize the distribution of childhood SES measures across race but not total years of education	Proposition 2: Intervene to equalize the distribution of total years of education across race within levels of childhood SES	Proposition 3: Intervene to equalize the distribution of total years of education and childhood SES measures across race	Proposition 4: Intervene to equalize the distribution of total years of education across race but not childhood SES measures	Reanalysis of Fryer: Statistically equalize the distribution of total years of education across race without control for childhood SES
Log wages (mean difference, SE)					
Initial disparity	−0.41 (0.04)	−0.30 (0.05)	−0.41 (0.04)	−0.41 (0.04)	−0.41 (0.04)
Residual disparity	−0.30 (0.05)	−0.26 (0.05)	−0.26 (0.05)	−0.30 (0.04)	−0.30 (0.04)
% reduction	26	15	37	27	28
Incarceration (risk ratio, SE)					
Initial disparity	3.53 (1.17)	2.39 (1.21)	3.53 (1.17)	3.53 (1.17)	3.53 (1.17)
Residual disparity	2.39 (1.21)	2.36 (1.21)	2.36 (1.21)	3.22 (1.45)	3.06 (1.18)
% reduction	45	2	46	13	19
Health[b] (mean difference, SE)					
Initial disparity	−0.14 (0.05)	−0.03 (0.06)	−0.14 (0.05)	−0.14 (0.05)	−0.14 (0.05)
Residual disparity	−0.03 (0.06)	−0.03 (0.06)	−0.03 (0.06)	−0.11 (0.22)	−0.11 (0.05)
% reduction	75	12	78	16	24

[a]The analytic sample size was 1154 for wages, 1988 for incarceration, and 1587 for health. All models included a mutually exclusive dummy variable for Hispanic ethnicity.
[b]Self-reported health assessed as a z score from the physical component subdomain of the 12-item Short Form Health Survey.
SES indicates socioeconomic status, SE standard error.

where ϕ_1 represents the overall race disparity in log wages conditional on covariates, and γ_1 represents the race disparity given childhood SES, also conditional on covariates. Under an intervention to equalize the distribution of childhood SES across race, the residual disparity would equal γ_1 and the disparity reduction would be $\phi_1 - \gamma_1$. Thus, the initial disparity ϕ_1, which equals −0.41 (0.04), would under Proposition 1 decrease to γ_1, which equals −0.30 (0.05), a 26% reduction. Figure B shows

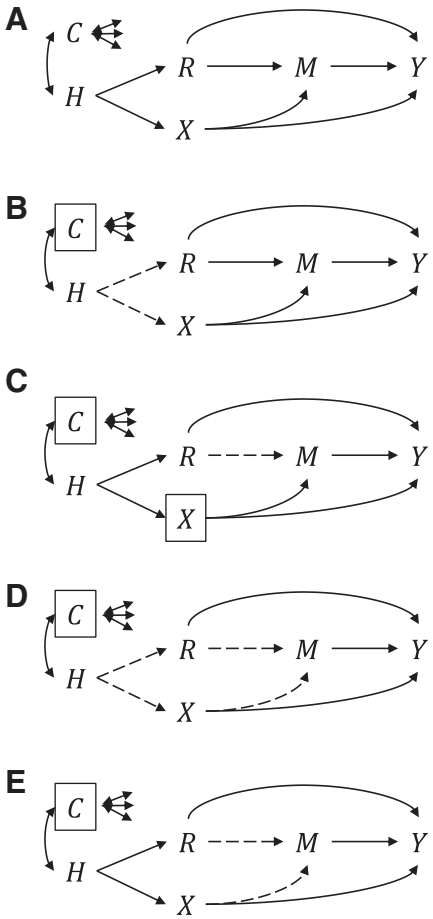

Figure. Diagram representing relationships between race R, an outcome Y, measures of characteristics in early life, for example, childhood SES X, measures of characteristics later in life, for example, test scores M, history H, and covariates gender, and age C in the population (A) and under alternative interventions (B)–(E), wherein solid arrows represent relationships that are preserved and dashed arrows represent relationships that are abolished. B, An intervention under Proposition 1 to equalize childhood SES across race deactivates the backdoor paths R ← H → X → M → Y and R ← H → X → Y. C, An intervention under Proposition 2 to equalize test scores within levels of childhood SES deactivates only the mediated path R → M → Y. D, An intervention to equalize both childhood SES and test scores across race deactivates the mediated path R → M → Y and also the backdoor paths R ← H → X → M → Y and R ← H → X → Y. E, An intervention to equalize test scores marginally across race deactivates the mediated path R → M → Y and the backdoor path R ← H → X → M → Y but leaves the backdoor path R ← H → X → Y intact. Note: we denote that a disparity is considered within levels of certain variables by placing a box around the conditioning variables. Bi-directional arrows out of C reflect association that may arise through causal relationships or selective forces, for example, early mortality. We omit this distinction here for simplicity. See page 7 of the eAppendix; http://links.lww.com/EDE/B387 for a more technical description of this graph.

that Proposition 1 corresponds to deactivating the backdoor path between race and childhood SES.

Proposition 2: Intervene to Equalize Test Scores Within Levels of Childhood SES Across Race

The second proposal is to randomly assign educational attainment (reflected in test scores, measures of premarket skills) among blacks such that it follows the same distribution among whites of the same gender, age, and childhood SES. This intervention attempts to remove disparities in test scores that cannot be attributed to disparities in childhood SES (a mediating path); it is not concerned with eliminating disparities that operate through childhood SES (a backdoor path). VanderWeele and Robinson[9] provided analytic formulae for the residual disparity under conditionally equalizing a mediating variable (i.e., an intermediate) such as test scores. These formulae require that assumption A2 holds. If this is so, we could estimate the residual disparity by fitting linear models that condition on race R, childhood SES X, covariates gender and age C (2), and additionally test scores M (3):

$$E[Y \mid r,x,c]=\gamma_0+\gamma_1 r+\gamma_2 x+\gamma_4{}' c \tag{2}$$

$$E[Y \mid r,x,m,c]=\theta_0+\theta_1 r+\theta_2 x+\theta_3 m+\theta_4{}' c \tag{3}$$

where γ_1 represents the race disparity given childhood SES, and θ_1 represents the disparity upon further stratifying on test scores, both of which condition on covariates. Under an intervention to take children with the same childhood SES and equalize their test score distribution across race, the residual disparity would equal θ_1 and the disparity reduction would be $\gamma_1 - \theta_1$. Thus, the initial disparity γ_1 which equals −0.30 (0.05) would decrease to θ_1, which equals −0.11 (0.05). Because this 65% reduction pertains to children who share the same childhood SES, much of the marginal disparity, without conditioning on childhood SES, between blacks and whites would remain. Figure C shows that the intervention eliminates the mediated path involving test scores but leaves backdoor paths involving childhood SES intact.

Proposition 3: Intervene to Equalize Both Test Scores and Childhood SES Across Race

The third proposal is to randomly assign childhood SES and educational attainment (reflected in test scores) among blacks such that they follow the same distribution among whites. This intervention targets mediated paths and backdoor paths by which disparities in log wages arise. In the Appendix, we extend the results of VanderWeele and Robinson[9] to provide formulae for the residual disparity under jointly equalizing a nonmediating variable such as childhood SES and also a mediating variable such as test scores. These formulae require that both assumptions A1 and A2 hold. Provided this is so, we

could estimate the residual disparity by fitting linear models that condition on race R, covariates gender and age C (1), and additionally childhood SES X and test scores M (3):

$$E[Y \mid r,c] = \phi_0 + \phi_1 r + \phi_4' c \quad (1)$$

$$E[Y \mid r,x,m,c] = \theta_0 + \theta_1 r + \theta_2 x + \theta_3 m + \theta_4' c \quad (3)$$

where ϕ_1 represents the overall race disparity in log wages, and θ_1 represents the disparity upon further stratifying on childhood SES and test scores, both of which condition on covariates. Under an intervention to equalize both the childhood SES and test score distributions across race, the residual disparity would equal θ_1 and the disparity reduction would be $\phi_1 - \theta_1$. Thus, the initial disparity γ_1 which equals −0.41 (0.04) would decrease to θ_1 which equals −0.11 (0.05), a 74% reduction. Figure D shows that the intervention eliminates the mediated path involving test scores as well as the backdoor paths involving childhood SES.

Proposition 4: Intervene to Equalize Test Scores Across Race

The previous proposals focused on targeting backdoor versus mediated paths that generate the disparity in log wages. But there is a conceptual issue with Proposition 2. Identifying the effect of eliminating a mediated path (through test scores) requires adjustment of confounders (of test scores, e.g., childhood SES, gender, and age). This may be problematic because achieving disparity reductions for children *of the same childhood SES* may constrain black children with low SES to a test score distribution that is already suboptimal. Test score disparities that arise through disparities in childhood SES would persist. These issues arise regardless of how one adjusts for childhood SES because methods for indirect effects will always statistically equalize the distribution of a nonmediating variable across race through stratification or standardization.

An alternative is to shift our focus away from eliminating mediated paths involving test scores and toward eliminating disparities in test scores entirely, regardless of whether they arise through mediated or backdoor paths. This describes the fourth proposition, which is to randomly assign educational attainment (reflected in test scores) among blacks such that they follow the same marginal distribution as among whites. In the Appendix, we provide formulae for the residual disparity under equalizing a variable such as test scores that lies along a mediating path and also a nonmediating path. These formulae adjust for confounding of test scores by childhood SES but preserve the relationship between race and childhood SES. They require that assumption A2 holds. Provided this is so, we could estimate the residual disparity by fitting linear models that condition on race R, covariates gender and age C (1), and additionally upon childhood SES X (2) and finally test scores M (3):

$$E[Y \mid r,c] = \phi_0 + \phi_1 r + \phi_4' c \quad (1)$$

$$E[Y \mid r,x,c] = \gamma_0 + \gamma_1 r + \gamma_2 x + \gamma_4' c \quad (2)$$

$$E[Y \mid r,x,m,c] = \theta_0 + \theta_1 r + \theta_2 x + \theta_3 m + \theta_4' c \quad (3)$$

where ϕ_1 represents the overall race disparity in log wages, γ_1 represents the disparity given childhood SES, θ_1 represents the disparity given childhood SES and test scores, γ_2 represents the race-specific total effect of childhood SES on log wages, and θ_2 represents the race-specific direct effect of childhood SES on log wages (with respect to test scores) which all condition on covariates. Under an intervention to equalize test scores alone across race, the residual disparity would equal $\theta_1 + (\theta_2/\gamma_2)(\phi_1 - \gamma_1)$, and the disparity reduction would be $(\gamma_1 - \theta_1) + (1 - \theta_2/\gamma_2)(\phi_1 - \gamma_1)$. Thus, the initial disparity, which equals −0.41 (0.04), would decrease to −0.14 (0.10), a 66% reduction. Conceptually, the disparity reduction estimate begins with the decrease that occurs under equalizing test scores within levels of childhood SES ($\gamma_1 - \theta_1$). To this amount, we add in the disparity reduction under equalizing childhood SES alone ($\phi_1 - \gamma_1$) but only scaled by the proportion that is mediated by test scores ($1 - \theta_2/\gamma_2$); this accounts for the extent to which an intervention on test scores would block the effect of childhood SES on the outcome. If test scores do not mediate the effect of childhood SES, such that $\theta_2 = \gamma_2$, none of the disparity reduction under equalizing childhood SES alone is added and the expression simplifies to the disparity reduction under Proposition 2, i.e., $\gamma_1 - \theta_1$. If test scores completely mediate the effect of childhood SES, such that $\theta_2 = 0$, then all of the disparity reduction under equalizing childhood SES alone is added, and the expression simplifies to the reduction under Proposition 3, i.e., $\phi_1 - \theta_1$. In Figure E, we see that the intervention removes a mediating path and a backdoor path that involves test scores, and that removing the backdoor path involves equalizing test scores across childhood SES.

A REANALYSIS OF FRYER

To illustrate our methods, we reconstructed the striking Fryer analyses where black–white disparities in log wages, unemployment, and health were substantially reduced upon controlling for test scores. We compared these results (where the term for X in model 3 was omitted) to those obtained under the four propositions described above that do account for childhood SES. In Table 2, the NLSY79 disparity in log wages was −0.41 (0.04), a quarter (26%) of which would be removed under equalizing childhood SES alone (Proposition 1). Equalizing childhood SES and test scores together (Proposition 3) would remove nearly three-quarters (74%) of the disparity. Interventions to equalize test scores alone would remove two-thirds (66%) of the disparity. For the disparity in incarceration, 3.54 (1.17), equalizing childhood SES would

remove just over two-fifths (45%) of the disparity, equalizing both childhood SES and test scores would reduce four-fifths (81%) of the disparity, and equalizing test scores alone would remove two-thirds of the disparity (65%). For the disparity in self-reported health −0.14 (0.05), equalizing childhood SES alone would remove all of the disparity (taking the standard errors into account), and this would be so for interventions to equalize childhood SES and test scores together and also interventions to equalize test scores alone.

In all, our analyses rule out confounding by our measures of childhood SES, supporting the results of Fryer that point towards educational disparities as a key target to reduce disparities. Importantly, while equalizing both childhood SES and test scores yielded the largest disparity reductions, equalizing test scores alone led to disparity reductions that were as large, and sometimes larger, than equalizing SES in early childhood alone. When equalizing total years of education rather than test scores (Table 3), we obtained much smaller reductions, possibly because the racial gaps were wider for test scores than for total years of education. Note also that Proposition 4 accomplishes its goal (in part) by equalizing educational attainment across childhood SES.

There are, however, several limitations in our illustrative example. Our results are subject to residual confounding from measurement error and missing data, and unmeasured confounding by dimensions of SES not included (e.g., parental occupation).[19,20] Gains in SES do not necessarily provide equal returns for blacks and whites and the present analysis does not include interactions.[21] Our analyses did not employ sampling weights to account for the NLSY design. The proposals describe goals, not detailed intervention protocols: much more difficult work is needed to formulate the best local and national interventions to address disparities in test scores and the enduring structural legacy of racism.

DISCUSSION

We have presented a way to decompose an extant disparity in adult outcomes (e.g., log wages) into a reduction and a residual portion upon equalizing disparities in a target variable that lies on a mediating path (e.g., test scores), even when that target is confounded by a variable that lies on a nonmediating path (e.g., childhood SES). This approach appropriately controls for confounding by a variable like childhood SES in a way that preserves its relationship with race.

This feature overcomes a conceptual constraint of current methods for indirect effects which can only estimate how well interventions reduce disparities after statistically equalizing confounding variables such as childhood SES across race. Substantively, interventions that intend to align educational outcomes within groups whose educational outcomes are already on average suboptimal (e.g., those with low SES) may not be effective.

Estimating disparity reductions while equalizing confounding variables could lead to misinterpretations. For example, consider years of education versus test scores. In Table 3, a mediation analysis that equalized total years of education within levels of childhood SES (Proposition 2) gave roughly the same numeric values for the residual disparity as an analysis that equalized childhood SES and test scores jointly (Proposition 3), and these were much smaller than the residual disparity under equalizing test scores marginally (Proposition 4). But the residual disparity under Proposition 2 ignores disparities that remain through race's association with childhood SES. Without careful interpretation of results from the mediation analysis, one might overinterpret the importance of total years of education for reducing disparities. Our decomposition method does not suffer from these limitations.

There has been considerable debate in the statistics, social science, and epidemiology literature as to whether socially defined characteristics such as race can be given causal attribution and whether their effects can be identified from observational data.[22–29] Our contribution is not meant to advance this debate. Our decomposition does not focus on the causal status of race or attempt to identify its effect. Rather it focuses causal inference on potentially manipulable targets and their ability to reduce the association between race and an outcome.[8] Mediation analysis methods—even when reframed to adopt this viewpoint[9,30]—requires epidemiologists to first equalize confounding variables across race and only then consider what targets can reduce that adjusted disparity.[31] Our approach sidesteps this restriction and opens a broader range of inquiry for reducing disparities.

Our contribution has implications for other approaches used to understand disparities. In economics, sources of disparities are often identified using the Oaxaca–Blinder decomposition. This method disaggregates the disparity into a portion due to statistical variation in the covariates—the explained portion—and an unexplained portion that is usually attributed to discrimination (as it captures both residual differences in the mean outcome at the reference levels of covariates and cases where adjusted covariate–outcome associations vary by race).[12,13] The Oaxaca–Blinder decomposition is increasingly appearing in the public health literature,[32,33] sometimes with causal interpretation, e.g., with the explained portion described as the disparity reduction under an intervention to equalize risk factors (targets). Such interpretations are highly questionable when they do not explicitly account for how those targets may be confounded. When all confounders of targets are adjusted for, and moreover when the confounders of targets in later life are not affected by race or targets in early life (i.e., no time-dependent confounding), each of Propositions 1 to 4 can be accomplished as an Oaxaca–Blinder decomposition (see Appendix). When there is time-dependent confounding, the Oaxaca–Blinder Decomposition methods would be vulnerable to selection bias[9,34] and should not be used. In the Appendix, we present nonparametric formulae that can be used to implement Propositions 2, 3, and 4 in the presence of a time-dependent confounder, effectively also generalizing Oaxaca–Blinder decomposition methods to this setting as well. On page 13 of the eAppendix; http://

links.lww.com/EDE/B387 we argue that, when causal interpretation is desired for the Oaxaca–Blinder decomposition, it will often be best to focus efforts on a single or small set of target/explanatory variables for which confounders are measured (and precede the explanatory variables), if one is indeed trying to control for confounding of each of these. The causal interpretations considered here pertain to interventions on the explanatory variables. Causal interpretations for interventions on race (or any group status),[35–41] among others,[42] have been explored in the economics literature (see page 11 of the eAppendix; http://links.lww.com/EDE/B387).

The results presented above require that the models be correctly specified; in the main text these formulas do not, for example, account for possible interaction between race, childhood SES, and test scores. However, the general nonparametric results given in the Appendix can be used to derive estimators for Propositions 2, 3, and 4 that allow not only for time-dependent confounding but also for interactions and less sensitivity to modeling assumptions, as has been done elsewhere.[43] The nonparametric formulae still require—along with those using linear models—that confounders of targets be measured and adjusted for, and also consistency and positivity with respect to the target(s) of interest. It will be important to develop intuitive sensitivity analyses that can quantify potential bias when some confounders are unmeasured.[44] Future work will also need to formalize interpretations of such interventions when they would plausibly vary across population subgroups, perhaps adapting guidance provided in the causal treatment effects and mediation literature.[45] Investigators will also have to consider how to deal with limited overlap in the covariates in each racial group through careful definition of the population of interest.[46] Future research could also expand this method to consider disparity reductions along multiple axes of disadvantage beyond race, i.e., questions framed with an intersectional focus.[47,48]

We have introduced a new perspective on how to use the potential outcomes framework to identify targets that appear attractive for reducing disparities. We hope these methods enable epidemiologists to help advance research priorities, policy initiatives, and intervention design to eliminate health disparities.

REFERENCES

1. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health*. 2006;27:167–194.
2. Ayanian JZ, Landon BE, Newhouse JP, Zaslavsky AM. Racial and ethnic disparities among enrollees in Medicare Advantage plans. *N Engl J Med*. 2014;371:2288–2297.
3. Sloan FA, Ayyagari P, Salm M, Grossman D. The longevity gap between Black and White men in the United States at the beginning and end of the 20th century. *Am J Public Health*. 2010;100:357–363.
4. 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.
5. Fryer RG. Racial inequality in the 21st century: the declining significance of discrimination. In: Card D, Ashenfelter O, eds. *Handbook of Labor Economics*, Vol. 4B. 1st ed. San Diego, CA: North Holland; 2011:855–971.
6. Neal DA, Johnson WR. The role of premarket factors in black–white wage differences. *J Polit Econ*. 1996;104(5):869–895.
7. Glymour MM, Avendaño M, Haas S, Berkman LF. Lifecourse social conditions and racial disparities in incidence of first stroke. *Ann Epidemiol*. 2008;18:904–912.
8. Greenland S. Epidemiologic measures and policy formulation: lessons from potential outcomes. *Emerg Themes Epidemiol*. 2005;2:5.
9. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25(4):473–484.
10. Duan N, Meng XL, Lin JY, Chen CN, Alegria M. Disparities in defining disparities: statistical conceptual frameworks. *Stat Med*. 2008;27:3941–3956.
11. Mackenbach JP, Gunning-Schepers LJ. How should interventions to reduce inequalities in health be evaluated? *J Epidemiol Community Health*. 1997;51:359–364.
12. Oaxaca R. Male-female wage differentials in urban labor markets. *Int Econ Rev (Philadelphia)*. 1973;14(3):693.
13. Blinder AS. Wage discrimination: reduced form and structural estimates. *J Hum Resour*. 1973;8(4):436.
14. Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1997 Cohort, 1997–2013 (Rounds 1–16). Produced by the National Opinion Research Center, the University of Chicago, and distributed by the Center for Human Resource.
15. Bureau of Labor Statistics, U.S. Department of Labor, and National Institute for Child Health and Human Development. Children of the NLSY79, 1979–2014. Produced and distributed by the Center for Human Resource Research, The Ohio State University, Columbus.
16. Reskin B. The race discrimination system. *Annu Rev Sociol*. 2012;38(1):17–35.
17. Oakes J, Ormseth T, Bell R, Camp P. *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: Rand Corp.; 1990.
18. Kelly S. The black–white gap in mathematics course taking. *Sociol Educ*. 2009;82(1):47–69.
19. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology*. 1997;8:621–628.
20. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. *JAMA*. 2005;294:2879–2888.
21. Williams DR, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. *Ann N Y Acad Sci*. 2010;1186:69–101.
22. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945–960.
23. Greiner DJ, Rubin DB. Causal effects of perceived immutable characteristics. *Rev Econ Stat*. 2011;93(3):775–785.
24. Glymour C. Statistics and metaphysics. *J Am Stat Assoc*. 1986;81(396):964–966.
25. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. *Am J Epidemiol*. 1999;150:113–120.
26. Krieger N, Smith GD. Re: "Seeking causal explanations in social epidemiology". *Am J Epidemiol*. 2000;151:831–833.
27. VanderWeele TJ, Hernán MA. 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; 2012:101–113.
28. Marcellesi A. Is race a cause? *Philos Sci*. 2013;80(5):650–659.
29. Sen M, Wasow O. Race as a bundle of sticks: designs that estimate effects of seemingly immutable characteristics. *Annu Rev Polit Sci*. 2016;19(1):499–522.
30. Naimi AI, Schnitzer ME, Moodie EE, Bodnar LM. Mediation analysis for health disparities research. *Am J Epidemiol*. 2016;184:315–324.
31. Morgenstern H. Defining and explaining race effects. *Epidemiology*. 1997;8:609–611.
32. Sen B. Using the Oaxaca-Blinder decomposition as an empirical tool to analyze racial disparities in obesity. *Obesity (Silver Spring)*. 2014;22:1750–1755.
33. Basu S, Hong A, Siddiqi A. Using decomposition analysis to identify modifiable racial disparities in the distribution of blood pressure in the United States. *Am J Epidemiol*. 2015;182(4):345–353.
34. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
35. Barsky R, Bound J, Charles KK, Lupton JP. Accounting for the black–white wealth gap: a nonparametric approach. *J Am Stat Assoc*. 2002;97(459):663–673.

36. Black D, Haviland A, Sanders S, Taylor L. Why do minority men earn less ? A study of wage differentials among the highly. *Rev Econ Stat*. 2006;88(2):300–313.
37. Fortin N, Lemieux T, Firpo S. Decomposition methods in economics. In: *Handbook of Labor Economics*. Vol 4.; 2011:1–102.
38. Kline P. Regression, reweighting, or both: Oaxaca–Blinder as a reweighting estimator. *Am Econ Rev*. 2011;101(3):532–537.
39. Słoczyński T. Average wage gaps and Oaxaca–Blinder decompositions. *IZA Discuss Pap No 9036*. 2015.
40. Słoczyński T. The Oaxaca–Blinder unexplained component as a treatment effects estimator. *Oxf Bull Econ Stat*. 2015;77(4):588–604.
41. Huber M. Causal pitfalls in the decomposition of wage gaps. *J Bus Econ Stat*. 2015;33(2):179–191.
42. Rothe C. Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes. *J Bus Econ Stat*. 2015;33(3):323–337.
43. Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann Stat*. 2012;40:1816–1845.
44. Ding P, Vanderweele TJ. Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika*. 2016;103:483–490.
45. VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1:1–20.
46. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*. 2009;96(1):187–199.
47. Jackson JW, Williams DR, VanderWeele TJ. Disparities at the intersection of marginalized groups. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:1349–1359.
48. Jackson JW. Explaining intersectionality through description, counterfactual thinking, and mediation analysis. *Soc Psychiatry Psychiatr Epidemiol*. 2017;52:785–793.

APPENDIX

Introduction and Notation

Consider a comparison of two race/ethnicity groups and let R denote a binary variable indicating race. Let X be a set of characteristics at birth or early childhood that are potentially manipulable (e.g., early SES measures), let M be one or more characteristics later in life or in adulthood that are potentially manipulable (e.g., educational attainment or adult SES), let Y be some outcome of interest, and let C be some other set of covariates at birth (e.g., gender, year of birth/age). The overall disparity measure within strata of covariates C (gender and age) would then be $E[Y \mid R=1,c]-E[Y \mid R=0,c]$. Unless noted otherwise, we will consider X to be a single measure of characteristics at birth.

Let $Y(x)$ be the value of the outcome that would have been observed for an individual had X been set to x. Likewise let $Y(m)$ be the value of the outcome that would have been observed for an individual had M been set to m. Finally, let $Y(x,m)$ be the value of the outcome that would have been observed for an individual had X been set to x and M to m.

Unless otherwise noted we will assume (see page 23 of the eAppendix; http://links.lww.com/EDE/B387 for formal statements):

A1: The effect of X on the outcome Y is unconfounded given (R,C)

A2: The effect of M on the outcome Y is unconfounded given (R,C,X)

Along with positivity and consistency for X and M.

Nonparametric Results in the Absence of Time-Dependent Confounding

Here we give nonparametric results for each of the various decompositions in the absence of time-dependent confounding. Estimates that are obtained from linear or logistic models from each of the decompositions are summarized in pages 17 through 19 of the eAppendix; http://links.lww.com/EDE/B387. Nonparametric results in the presence of time-dependent confounding can be found below.

Proposition 1 (VanderWeele and Robinson[9]). The disparity that would remain if the childhood distribution of X for black persons ($R=1$) with covariates $C=c$ were set equal to its distribution for white persons ($R=0$) with $C=c$ would be

$$\mu_x - E[Y \mid R=0,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R=1,c] - \mu_x$$

where

$$\mu_x = \sum_x E[Y \mid R=1,x,c]P(x \mid R=0,c).$$

Proposition 2 (VanderWeele and Robinson[9]). The disparity that would remain if the distribution of M for black persons ($R=1$) with covariates $C=c$ and $X=x$ were set equal to its distribution for white persons ($R=0$) with $C=c$ and $X=x$ would be

$$\mu_{m|x} - E[Y \mid R=0,x,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R=1,x,c] - \mu_{m|x}$$

where

$$\mu_{m|x} = \sum_m E[Y \mid R=1,x,m,c]P(m \mid R=0,x,c).$$

Proposition 3. The disparity that would remain if the distribution of (X,M) for black persons ($R=1$) with covariates $C=c$ were set equal to its distribution for white persons ($R=0$) with $C=$ c would be

$$\mu_{xm} - E[Y \mid R=0,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R=1,c] - \mu_{xm}$$

where

$$\mu_{xm} = \sum_{x,m} E[Y \mid R=1,x,m,c]P(m \mid R=0,x,c)P(x \mid R=0,c).$$

Proposition 4. The disparity that would remain if the distribution of M for black persons ($R = 1$) with covariates $C = c$ were set equal to its distribution for white persons ($R = 0$) with $C = c$ would be

$$\mu_m - E[Y \mid R = 0,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R = 1,c] - \mu_m$$

where

$$\mu_m = \sum_{x,m} E[Y \mid R = 1,x,m,c]P(m \mid R = 0,c)P(x \mid R = 1,c).$$

Nonparametric Results in the Presence of Time-Dependent Confounding

Let us additionally consider a variable L that may be affected by C,R,X and that affects both M and Y and an additional assumption A3: the effect of M on the outcome Y is unconfounded given (R,C,X,L).

Proposition 5. Under (A3), the disparity that would remain if the distribution of M for black persons ($R = 1$) with $X = x$ and covariates $C = c$ were set equal to its distribution for white persons ($R = 0$) with $X = x$ and $C = c$ would be

$$\mu_{m|x} - E[Y \mid R = 0,x,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R = 1,x,c] - \mu_{m|x}$$

where

$$\mu_{m|x} = \sum_{m,l|x} E[Y \mid R = 1,x,m,c,l]P(l \mid R = 1,x,c)P(m \mid R = 0,x,c).$$

Proposition 6. Under (A1) and (A3), the disparity that would remain if the distribution of (X,M) for black persons ($R = 1$) with covariates $C = c$ were set equal to its distribution for white persons ($R = 0$) with $C = c$ would be

$$\mu_{xm} - E[Y \mid R = 0,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R = 1,c] - \mu_{xm}$$

where

$$\mu_{xm} = \sum_{x,m,l} E[Y \mid R = 1,x,m,c,l]P(l \mid R = 1,x,c) P(m \mid R = 0,x,c)P(x \mid R = 0,c).$$

Proposition 7. Under (A3), the disparity that would remain if the distribution of M for black persons ($R = 1$) with covariates $C = c$ were set equal to its distribution for white persons ($R = 0$) with $C = c$ would be

$$\mu_m - E[Y \mid R = 0,c]$$

and the amount the disparity is reduced would be

$$E[Y \mid R = 1,c] - \mu_m$$

where

$$\mu_m = \sum_{x,m,l} E[Y \mid R = 1,x,m,c,l]P(l \mid R = 1,x,c) P(m \mid R = 0,c)P(x \mid R = 1,c).$$

Results Expressed as Oaxaca–Blinder Decompositions

Review of Oaxaca–Blinder Decomposition

The Oaxaca–Blinder decomposition[12,13] is often used in labor economics to understand how much differences in group characteristics explain disparities (or differences) in outcomes across groups. Below, we establish connections between Oaxaca–Blinder decomposition and our proposals 1 to 4 using linear models. However, connections with our nonparametric estimators for proposals 1 to 4 also extend to nonparametric Oaxaca–Blinder decomposition estimators for the mean from the economics literature.[35–37] Propositions 5 to 7 extend these nonparametric estimators to cases where there is time-dependent confounding. The causal interpretations in mind here are with respect to the explanatory variables. The economics literature has mostly explored causal inference with respect to race or group status.[35–42] Pages 2 to 6 of the eAppendix; http://links.lww.com/EDE/B387 contain an expanded version of the introduction below and an overview of where our results fit with the economics literature on causal inference for OBD decompositions.

Marginal Decompositions

A typical Oaxaca–Blinder decomposition with linear regression fits race-specific models for Y given the explanatory variables X and M:

$$E[Y \mid R = 1,m,x,c] = \alpha_0^* + \alpha_1^* x + \alpha_2^* m$$

$$E[Y \mid R = 0,m,x,c] = \beta_0^* + \beta_1^* x + \beta_2^* m$$

An "aggregate decomposition" examines how much the marginal disparity $E[Y \mid R = 1] - E[Y \mid R = 0]$ is explained by differences in means for the explanatory variables X and M, which is estimated by $\beta_1^*\{E[X \mid R = 1] - E[X \mid R = 0]\} + \beta_2^* E\{[M \mid R = 1] - E[X \mid M = 0]\}$ (the "explained portion" or the "composition effect"). It also examines how much marginal disparity is explained by differential associations of the explanatory variables (by race), which is captured in $(\alpha_1^* - \beta_1^*)E[X = 0] + (\alpha_2^* - \beta_2^*)E[M = 0] + (\alpha_0^* - \beta_0^*)$ (the "unexplained portion" or "structure effect"). In a detailed decomposition, the explained and unexplained portion are further decomposed for each explanatory variable. If the effects of X and M are confounded, these quantities have statistical but not causal interpretations.

Conditional Decompositions

Suppose now that we refit the models above to include C:

$$E[Y \mid R=1,m,x,c] = \alpha_0 + \alpha_1 x + \alpha_2 m + \alpha_3' c$$

$$E[Y \mid R=0,m,x,c] = \beta_0 + \beta_1 x + \beta_2 m + \beta_3' c$$

With a conditional disparity, $E[Y \mid R=1,c] - E[Y \mid R=0,c]$, the explained portion is $\beta_1\{E[X \mid R=1,c] - E[X \mid R=0,c]\} + \beta_2\{E[M \mid R=1,c] - E[X \mid M=0,c]\}$ and an unexplained portion is $(\alpha_1 - \beta_1)E[X=0,c] + (\alpha_2 - \beta_2)E[M=0,c] + (\alpha_2 - \beta_2)E[M=0,c] + (\alpha_0 - \beta_0)$. We have not seen these quantities considered formally in the economics or epidemiology literature. This minor extension has important implications for the Oaxaca-Blinder Decomposition's causal interpretation. If the effect of X on Y is unconfounded given R,C and the effect of M on Y is unconfounded given X,R,C, then the Oaxaca-Blinder Decompositions have causal interpretations, as we present below. The causal interpretation here is with respect to interventions to set the explanatory variables instead of race (see page 23 of the eAppendix; http://links.lww.com/EDE/B387 for proofs).

Propositions 1 to 4 Recast as Causal Oaxaca–Blinder Decompositions Under Linear Models

Suppose we fit the following sets of models, assuming no statistical interactions between C and any other variable in the model (to simplify the formulae), that the effect of X is unconfounded given $\{R,C\}$, and that the effect of M is unconfounded given $\{R,M,C\}$.

Set1: $E[Y \mid R=1,x,c] = \omega_0 + \omega_1 x + \omega_3' c$

$$E[Y \mid R=0,x,c] = \pi_0 + \pi_1 x + \pi_3' c$$

Set2: $E[Y \mid R=1,m,x,c] = \alpha_0 + \alpha_1 x + \alpha_2 m + \alpha_3' c$

$$E[Y \mid R=0,m,x,c] = \beta_0 + \beta_1 x + \beta_2 m + \beta_3' c$$

Set3: $E[Y \mid r,x,m,c] = \theta_0 + \theta_1 r + \theta_2 x + \theta_3 m + \theta_4 rx + \theta_5 rm + \theta_6' c$

$$E[Y \mid r,x,c] = \gamma_0 + \gamma_1 r + \gamma_2 x + \gamma_4 rx + \gamma_6' c$$

$$E[Y \mid r,c] = \phi_0 + \phi_1 r + \phi_6' c$$

Goal of Proposition 1: equalize childhood SES across race given covariates, i.e., standardization. Consider the disparity $E[Y \mid R=1,c] - E[Y \mid R=0,c]$. Under Proposition 1, the residual disparity equals $(\omega_0 - \pi_0) + (\omega_1 - \pi_1)E[X \mid R=0,c]$ with set 1 and equals $\gamma_1 + \gamma_4 E[X \mid R=0,c]$ with set 3. The disparity reduction equals $\omega_1\{E[X \mid R=1,c] - E[X \mid R=0,c]\}$ with set 1 and equals $(\gamma_2 + \gamma_4)\{E[X \mid R=1,c] - E[X \mid R=0,c]\}$ with set 3. These are the unexplained and explained portions from an aggregate decomposition of $E[Y \mid R=1,c] - E[Y \mid R=0,c]$.

Goal of Proposition 2: equalize test scores across race given childhood SES and covariates, i.e., mediation-analysis. Consider the disparity $E[Y \mid R=1,x,c] - E[Y \mid R=0,x,c]$. Under Proposition 2, the residual disparity equals $(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x + (\alpha_2 - \beta_2)E[M \mid R=0,x,c]$ with set 2 and equals $\theta_1 + \theta_4 x + \theta_5 E[M \mid R=0,x,c]$ with set 3. The disparity reduction equals $\alpha_2\{E[M \mid R=1,x,c] - E[M \mid R=0,x,c]\}$ with set 2 and equals $(\theta_3 + \theta_5)\{E[M \mid R=1,x,c] - E[M \mid R=0,x,c]\}$ with set 3. These are the unexplained and explained portions from an aggregate decomposition of $E[Y \mid R=1,x,c] - E[Y \mid R=0,x,c]$.

Goal of Proposition 3: equalize childhood SES and test scores across race given covariates. Consider the disparity $E[Y \mid R=1,c] - E[Y \mid R=0,c]$. Under Proposition 3, the residual disparity equals $(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)E[X \mid R=0,c] + (\alpha_2 - \beta_2)E[M \mid R=0,c]$ with set 2 and equals $\theta_1 + \theta_4 E[X \mid R=0,c] + \theta_5 E[M \mid R=0,c]$ with set 3. The disparity reduction equals $\alpha_1\{E[X \mid R=1,c] - E[X \mid R=0,c]\} + \alpha_2\{E[M \mid R=1,c] - E[M \mid R=0,c]\}$ with set 2 and equals $(\theta_2 + \theta_4)\{E[X \mid R=1,c] - E[X \mid R=0,c]\} + (\theta_3 + \theta_5)\{E[M \mid R=1,c] - E[M \mid R=0,c]\}$ with set 3. These are the unexplained and explained portions from an aggregate decomposition of $E[Y \mid R=1,c] - E[Y \mid R=0,c]$.

Goal of Proposition 4: equalize test scores across race given covariates. Under Proposition 4, the residual disparity equals $(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)E[X \mid R=0,c] + (\alpha_2 - \beta_2)E[M \mid R=0,c] + \alpha_1\{E[X \mid R=1,c] - E[X \mid R=0,c]\}$ with set 2 and equals $\theta_1 + \theta_4 E[X \mid R=1,c] + \theta_2\{E[X \mid R=1,c] - E[X \mid R=0,c]\} + \theta_5 E[M \mid R=0,c]$ with set 3. The disparity reduction equals $\alpha_2\{E[M \mid R=1,c] - E[M \mid R=0,c]\}$ with set 2 and equals $(\theta_3 + \theta_5)\{E[M \mid R=1,c] - E[M \mid R=0,c]\}$ with set 3. This corresponds to a detailed decomposition wherein the explained portion specific to M equals the disparity reduction, and the sum of the explained portion specific to X along with the entire unexplained portion equals the residual disparity. Note that unless M does not mediate the effect of X on Y, the "explained" detailed decomposition term for X does not on its own have a clear causal interpretation (see page 39 of the eAppendix; http://links.lww.com/EDE/B387).