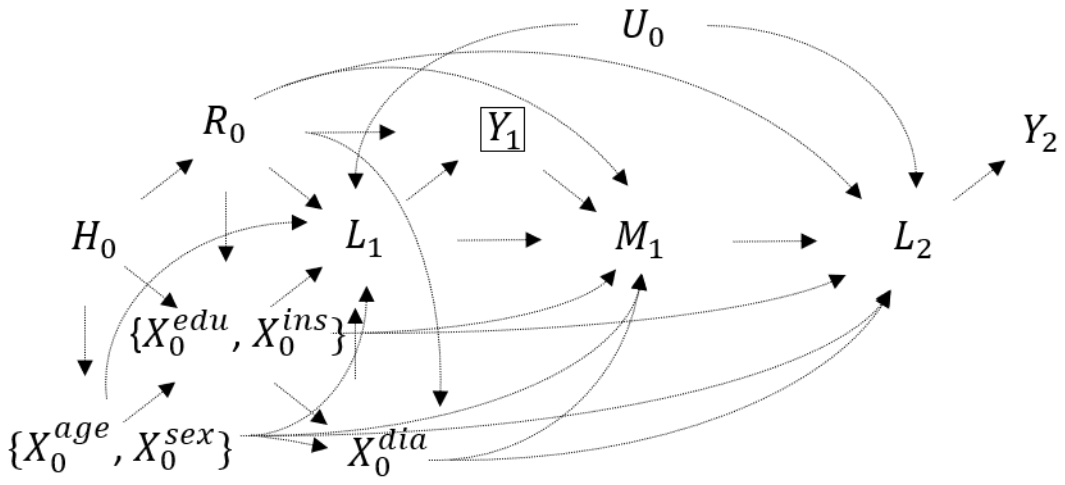
eAppendix.

Table of Contents

Proofs

Notation.

Let the subscript t index the timing of measurement for variable V_t (0=pre-baseline, 1=baseline, 2=follow-up). Let L_1 and L_2 equal, respectively, a patient's outcome (e.g., blood pressure) at the baseline and follow-up visit. Let Y_1 and Y_2 equal, respectively, a patient's diagnosis based on L (e.g., uncontrolled hypertension; 1=yes, 0=no) at the baseline and follow-up visit. Let M_1 equal a determinant of L_2 that we want to intervene upon to alter the distribution of Y_2 (e.g., decision to intensify antihypertensive treatment; 1=yes, 0=no). Let $X_0^{edu}, X_0^{ins}, X_0^{dia}$ (educational attainment, private health insurance, and diabetes, respectively) be measured common causes of L_1, M_1, and L_2, let X_0^{age} and X_0^{sex} (age and sex, respectively) equal common causes of all these variables, let R_0 equal a binary variable that defines a socially marginalized population (e.g., race), let H_0 equal sociopolitical forces (e.g., racism) that creates association between R_0 and X. Let U_0 equal an unmeasured source of correlation between L_1 and L_2. (Our results still hold even if this unmeasured cause affects the covariates X). For intuition, see eFigure 1 for a causal graph relating these variables. Let $V(w)$ equal the value that V would take (i.e. potential outcome, counterfactual) had W been set to value w. Let the notation $V \coprod W|Z$ denote statistical independence between V and W given Z.



eFigure 1. Causal graph describing the the R_0—Y_2 association through H,X,L_1,Y_1,M_1, and L_2

Definition.

General formulation.

As defined above, the variable R represents the social status across which the disparity will be measured (e.g. race). $R_0 = r_0$ will represent a marginalized group (e.g. blacks) and $R_0 = r_0'$ the privileged group (e.g. whites). (It is entirely possible to consider the following proposition with these values switched). The population of interest consists of all patients with uncontrolled hypertension at baseline ($Y_1 = 1$). Consider an intervention to set the distribution of a target variable M_1 (antihypertensive treatment intensification) to affect disparities in the outcome Y_2, uncontrolled hypertension. We will define three non-overlapping sets. The first variable set $\boldsymbol{A_1^y}$ defines the covariates that are considered both outcome- and target-allowable. The second variable set $\boldsymbol{A_1^m}$ defines covariates that are additionally considered target-allowable but not outcome-allowable. The third variable set $\boldsymbol{N_1}$ defines covariates that, in addition to those in $\boldsymbol{A_1^y}$ and $\boldsymbol{A_1^m}$, are needed for causal identification but are nonetheless considered non-allowable. It is permissible to partition the covariates such that some sets remain empty. For example, if $\boldsymbol{A_1^y}$ contains all covariates, then by definition $\boldsymbol{A_1^m}$ and $\boldsymbol{N_1}$ are empty. All expressions that follow condition on the population of interest, patients with hypertension at baseline.

Proposition.

Consider an intervention $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ among those with $R_0 = r_0$ to set the distribution of M_1 according to the observed distribution $P\left(m_1 \middle| R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)$. The observed disparity prior to intervention, and the reduced and residual disparity after intervention are given, respectively, as:

i) $\sum_{\boldsymbol{a_1^y}} E[Y_2|R_0 = r_0, \boldsymbol{a_1^y}]P(\boldsymbol{a_1^y}) - \sum_{\boldsymbol{a_1^y}} E\left[Y_2 \middle| R_0 = r_0', \boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})$
ii) $\sum_{\boldsymbol{a_1^y}} E\left[Y_2 \middle| R_0 = r_0, \boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y}) - \sum_{\boldsymbol{a_1^y}} E[Y_2(G_{m_1|\boldsymbol{a_1^m a_1^y}})|R_0 = r_0, \boldsymbol{a_1^y}]P(\boldsymbol{a_1^y})$
iii) $\sum_{\boldsymbol{a_1^y}} E\left[Y_2(G_{m_1|\boldsymbol{a_1^m a_1^y}}) \middle| R_0 = r_0, \boldsymbol{a_1^y}\right] P(\boldsymbol{a_1^y}) - \sum_{\boldsymbol{a_1^y}} E\left[Y_2 \middle| R_0 = r_0', \boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})$

These definitions require that $P\left(R_0 = r_0 \middle| \boldsymbol{a_1^y}\right) > 0$ and $P\left(R_0 = r_0' \middle| \boldsymbol{a_1^y}\right) > 0$ for all $\boldsymbol{a_1^y}$ with $P\left(\boldsymbol{a_1^y}\right) > 0$.

Remark 1. Alternate definitions of (i)-(iii) that replace the pooled distribution of the outcome-allowable covariates $P(\boldsymbol{a_1^y})$ can be used. For example, if the distribution the outcome-allowable covariates among blacks were used, we would replace $P(\boldsymbol{a_1^y})$ with $P\left(\boldsymbol{a_1^y} \middle| R_0 = r_0\right)$, under the weaker assumption that $P\left(R_0 = r_0' \middle| \boldsymbol{a_1^y}\right) > 0$ for all $\boldsymbol{a_1^y}$ with $P\left(\boldsymbol{a_1^y} | R_0 = r_0\right) > 0$. Likewise, if the distribution of the outcome-allowable covariates among whites were used, we would replace $P(\boldsymbol{a_1^y})$ with $P\left(\boldsymbol{a_1^y} \middle| R_0 = r_0'\right)$, under the weaker assumption that $P\left(R_0 = r_0 \middle| \boldsymbol{a_1^y}\right) > 0$ for all $\boldsymbol{a_1^y}$ with $P\left(\boldsymbol{a_1^y} | R_0 = r_0'\right) > 0$. Alternatively, one could restrict the population of interest to a region where common support holds, perhaps through eligibility criteria. Then, each component of the formulae above would implicitly condition on the eligible population of interest.

Henceforth, we will develop results using the pooled distribution of outcome-allowable covariates $P\left(\boldsymbol{a_1^y}\right)$ to measure disparities. The alternatives we outlined will produce diverging estimates when measures of disparity within levels of the outcome-allowable covariates $\boldsymbol{A_1^y}$ vary across levels of these covariates.

Identification.

As stated, let $\boldsymbol{N_1}$ denote additional variables needed for conditional exchangeability beyond $\boldsymbol{A_1^m}$ and $\boldsymbol{A_1^y}$.

Assumptions.

A) Conditional exchangeability among $R_0 = r_0$:

$Y_2(m) \coprod m_1 | R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}$ for all values m_1 with $P\left(m_1 \middle| R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) > 0$

B1) Positivity among $R_0 = r_0$:

$P\left(m_1 \middle| R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) > 0$ for $\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}$ with $P(\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y} | R_0 = r_0) > 0$ for all values m_1 with $P\left(m_1 \middle| R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) > 0$

B2) Common support across R_0:

$P(\boldsymbol{a_1^m}, \boldsymbol{a_1^y} | R_0 = r_0) > 0$ if $P(\boldsymbol{a_1^m}, \boldsymbol{a_1^y} | R_0 = r_0') > 0$ and $P\left(m_1 \middle| R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) > 0$ if $P\left(m_1 \middle| R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) > 0$ for $\boldsymbol{a_1^m}, \boldsymbol{a_1^y}$ with $P(\boldsymbol{a_1^m}, \boldsymbol{a_1^y} | R_0 = r_0') > 0$

C) Consistency:

$M_{1,i} = m_{1,i} \Rightarrow Y_{2,i} = Y_{2,i}\left(m_{1,i}\right)$ for all individuals i

Remark 2. Assumption "A" is a form of "partial" conditional exchangeability: within levels of allowables and non-allowables among blacks, only potential outcomes indexed by certain target values (those values observed among whites with identical values for allowables) are assumed to be independent of the observed target value. Assumption B1 is a form of "partial" positivity: within levels of allowables and non-allowables among blacks, only certain values of the target (those values observed among whites with identical values for allowables) are assumed to be observed with positive probability. These assumptions allow for identification when, for some values of allowables, blacks' treatment is always intensified if whites' treatment is as well. These conditional exchangeability and positivity assumptions are weaker than their standard versions.

Following Jackson & VanderWeele 2018 and Jackson 2018 we have among those with $R_0 = r_0$:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G_{m_1|\boldsymbol{a_1^m a_1^y}} = m_1\right)\middle|R_0 = r_0, \boldsymbol{a_1^y}\right] P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0 = r_0, G_{m_1|\boldsymbol{a_1^m a_1^y}} = m_1, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P\left(G_{m_1|\boldsymbol{a_1^m a_1^y}} = m_1\middle|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P\left(G_{m_1|\boldsymbol{a_1^m a_1^y}} = m_1\middle|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2(m_1)\middle|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2(m_1)\middle|R_0 = r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2\middle|R_0 = r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \qquad (1)
\end{aligned}$$

Where the first (and fourth) equality follow by the total law of probability, the second by definition of $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ as random among $R_0 = r_0$ given $\boldsymbol{a_1^m}$ and $\boldsymbol{a_1^y}$, the third by definition of $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ among $R_0 = r_0$ as a random draw from the distribution $P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ under assumption B2, the fifth by A and B1, and the sixth by C.

Note that among those with $R_0 = r_0$:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E[Y_2|R_0 = r_0, \boldsymbol{a_1^y}] P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \qquad (2)
\end{aligned}$$

And likewise, among those with $R_0 = r_0'$:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E[Y_2|R_0 = r_0', \boldsymbol{a_1^y}] P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2|R_0 = r_0', m_1, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0', \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \qquad (3)
\end{aligned}$$

Alternatively, among those with $R_0 = r_0'$:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E[Y_2|R_0 = r_0', \boldsymbol{a_1^y}] P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0', m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_1|R_0 = r_0', \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{n_1}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}) P(\boldsymbol{a_1^m}|R_0 = r_0', \boldsymbol{a_1^y}) P(\boldsymbol{a_1^y}) \qquad (3^*)
\end{aligned}$$

Remark 3. Equations (1), (2), and (3) represent g-formulae for decomposition with time-fixed interventions.

Remark 4. The distribution $P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ in (1) could be viewed as a marginalization of $P(M_1 = m_1|R_0 = r_0', \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ over the distribution $P(\boldsymbol{n_1}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ rather than conditional independence between M_1 and $\boldsymbol{N_1}$ given $R_0 = r_0'$, $\boldsymbol{a_1^m}$, and $\boldsymbol{a_1^y}$, as can be seen by contrasting the expressions (3) and (3*). Nonetheless, this distribution $P(M_1 = m_1|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ defines the intervention $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ which, when applied to blacks $R_0 = r_0$, produces independence between M_1 and $\boldsymbol{N_1}$ given $\boldsymbol{a_1^m}$, and $\boldsymbol{a_1^y}$.

Remark 5. The comparison to existing estimators on pages 9 to 13 are derived under the alternate identification formulae (1), (2), and (3*).

Estimation.

Decomposition using Ratio of Mediator Probability Weights (RMPW)

Let M_1 be categorical with j levels m_{1j}.

Among those with $R_0 = r_0$ we have that:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E[Y_2(G_{m_1|\boldsymbol{a_1^m a_1^y}})|R_0 = r_0, \boldsymbol{a_1^y}]P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y})P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y})P(\boldsymbol{a_1^y}) \\
&\qquad \times \frac{P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})} \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y})P(\boldsymbol{a_1^y}|R_0 = r_0) \\
&\qquad \times \frac{P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})} \times \frac{P(\boldsymbol{a_1^y})}{P(\boldsymbol{a_1^y}|R_0 = r_0)} \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right] P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0 = r_0, \boldsymbol{a_1^y})P(\boldsymbol{a_1^y}|R_0 = r_0) \\
&\qquad \times \frac{P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})} \times \frac{P(r_0)}{P(r_0|\boldsymbol{a_1^y})} \\
&= E\left[E\left[Y_2 \times w_{r_0}^{rmpw}|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right]|r_0\right] \qquad (4a)
\end{aligned}$$

$$\text{where } w_{r_0}^{rmpw} = \frac{P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})} \times \frac{P(r_0)}{P(r_0|\boldsymbol{a_1^y})} \qquad (4b)$$

The first equality is identified via eqn 1.

Note that, among those with $R_0 = r_0$ we have:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2|r_0, \boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y}) \\
&= E\left[E\left[Y_2 \times w_{r_0}|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right]|r_0\right] \qquad (5a)
\end{aligned}$$

$$\text{where } w_{r_0} = \frac{P(r_0)}{P(r_0|\boldsymbol{a_1^y})} \qquad (5b)$$

And among those with $R_0 = r_0'$:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2|r_0', \boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y}) \\
&= E\left[E\left[Y_2 \times w_{r_0'}|r_0', m_1, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right]|r_0'\right] \qquad (6a)
\end{aligned}$$

$$\text{where } w_{r_0'} = \frac{P(r_0')}{P(r_0'|\boldsymbol{a_1^y})} \qquad (6b)$$

Thus, under the expressions and weights defined above, we have the general result:

The observed disparity

$$\psi^{obs} = \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E[Y_2|R_0 = r_0, \boldsymbol{a}_1^{\boldsymbol{y}}]P(\boldsymbol{a}_1^{\boldsymbol{y}}) - \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E\big[Y_2\big|R_0 = r_0', \boldsymbol{a}_1^{\boldsymbol{y}}\big]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \quad (7a)$$
$$= E\big[E\big[Y_2 \times w_{r_0}\big|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0\big] - E\big[E\big[Y_2 \times w_{r_0'}\big|r_0', m_1, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0'\big]$$

The reduced disparity

$$\psi^{red} = \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E[Y_2|R_0 = r_0, \boldsymbol{a}_1^{\boldsymbol{y}}]P(\boldsymbol{a}_1^{\boldsymbol{y}}) - \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E[Y_2(G_{m_1|\boldsymbol{a}_1^{\boldsymbol{m}}\boldsymbol{a}_1^{\boldsymbol{y}}})|R_0 = r_0, \boldsymbol{a}_1^{\boldsymbol{y}}]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \quad (7b)$$
$$= E\big[E\big[Y_2 \times w_{r_0}\big|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0\big] - E\big[E\big[Y_2 \times w_{r_0}^{rmpw}\big|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0\big]$$

The residual disparity

$$\psi^{res} = \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E[Y_2(G_{m_1|\boldsymbol{a}_1^{\boldsymbol{m}}\boldsymbol{a}_1^{\boldsymbol{y}}})|R_0 = r_0, \boldsymbol{a}_1^{\boldsymbol{y}}]P(\boldsymbol{a}_1^{\boldsymbol{y}}) - \sum_{\boldsymbol{a}_1^{\boldsymbol{y}}} E\big[Y_2\big|R_0 = r_0', \boldsymbol{a}_1^{\boldsymbol{y}}\big]P(\boldsymbol{a}_1^{\boldsymbol{y}}) \quad (7c)$$
$$= E\big[E\big[Y_2 \times w_{r_0}^{rmpw}\big|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0\big] - E\big[E\big[Y_2 \times w_{r_0'}\big|r_0', m_1, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big]|r_0'\big]$$

With weights defined as

$$w_{r_0} = \frac{P(r_0)}{P\big(r_0|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}$$

$$w_{r_0'} = \frac{P(r_0')}{P\big(r_0'|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}$$

$$w_{r_0}^{rmpw} = \frac{P\big(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)} \times \frac{P(r_0)}{P\big(r_0|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}$$

Contrast (7a) can be estimated as β_1 in the weighted regression model with the observed data:

$E[Y_2|R_0] = \beta_0 + \beta_1 R_0$ fit with weights w_{r_0} for those with $R_0 = r_0$ and $w_{r_0'}$ for those with $R_0 = r_0'$.

Constrast (7b) can be estimated as β_1 in the weighted regression model with a stacked dataset consisting of the original subset $R_0 = r_0$ (labelled as $D_0 = d_0$) and a copy of the subset $R_0 = r_0$ (labelled as ($D_0 = d_0'$).

$E[Y_2|D_0] = \beta_0 + \beta_1 D_0$ fit with weights w_{r_0} for those with $D_0 = d_0$ and $w_{r_0}^{rmpw}$ for those with $D_0 = d_0'$.

Contrast (7c) can be estimated as β_1 in the weighted regression model with the observed data:

$E[Y_2|R_0] = \beta_0 + \beta_1 R_0$ fit with weights $w_{r_0}^{rmpw}$ for those with $R_0 = r_0$ and $w_{r_0'}$ for those with $R_0 = r_0'$.

Remark 6. (7a), (7b), and (7c) are based on disparity measures that use the pooled distribution $P(\boldsymbol{a}_1^{\boldsymbol{y}})$ to standardize the outcome-allowable covariates. With $P(\boldsymbol{a}_1^{\boldsymbol{y}}|R_0 = r_0)$ as the standard the weights would be:

$$w_{r_0} = 1 \qquad w_{r_0'} = \frac{P\big(r_0|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(r_0'|\boldsymbol{a}_1^{\boldsymbol{y}}\big)} \times \frac{P(r_0')}{P(r_0)} \qquad w_{r_0}^{rmpw} = \frac{P\big(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)}$$

With $P(\boldsymbol{a}_1^{\boldsymbol{y}}|R_0 = r_0')$ as the standard the weights would be:

$$w_{r_0} = \frac{\big(r_0'|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(r_0|\boldsymbol{a}_1^{\boldsymbol{y}}\big)} \times \frac{P(r_0)}{P(r_0')} \qquad w_{r_0'} = 1 \qquad w_{r_0}^{rmpw} = \frac{P\big(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(M_1 = m_{1j}|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a}_1^{\boldsymbol{m}}, \boldsymbol{a}_1^{\boldsymbol{y}}\big)} \times \frac{\big(r_0'|\boldsymbol{a}_1^{\boldsymbol{y}}\big)}{P\big(r_0|\boldsymbol{a}_1^{\boldsymbol{y}}\big)} \times \frac{P(r_0)}{P(r_0')}$$

Remark 7. The conditionality of the intervention $G_{m_1|\boldsymbol{a}_1^{\boldsymbol{m}}\boldsymbol{a}_1^{\boldsymbol{y}}}$ appears through the numerator in (4b). Any non-allowable confounders $\boldsymbol{N_1}$ beyond the allowable variables defined in $\boldsymbol{A}_1^{\boldsymbol{m}}$ and $\boldsymbol{A}_1^{\boldsymbol{y}}$ appear only in the denominator. Thus, the conditionality of the numerator will differ from the denominator whenever the intervention $G_{m_1|\boldsymbol{a}_1^{\boldsymbol{m}}\boldsymbol{a}_1^{\boldsymbol{y}}}$ does not condition on all of the confounders of M_1.

Decomposition using Inverse Odds Ratio Weights (IORW)

Let M_1 be categorical with j levels m_{1j}.

Among those with $R_0 = r_0$ we have that:

$$\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E[Y_2(G_{m_1|\boldsymbol{a_1^m a_1^y}})|R_0 = r_0, \boldsymbol{a_1^y}]P(\boldsymbol{a_1^y}) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\big[Y_2\big|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big]\, P\big(M_1 = m_{1j}\big|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{n_1}\big|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^m}\big|R_0 = r_0, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^y}\big) \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\big[Y_2\big|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big]\, P\big(M_1 = m_{1j}\big|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{n_1}\big|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^m}\big|R_0 = r_0, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^y}\big|R_0 = r_0\big) \\
&\qquad \times \frac{P\big(M_1 = m_{1j}\big|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)}{P\big(M_1 = m_{1j}\big|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)} \times \frac{P(r_0)}{P\big(r_0\big|\boldsymbol{a_1^y}\big)} \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\big[Y_2\big|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big]\, P\big(M_1 = m_{1j}\big|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{n_1}\big|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^m}\big|R_0 = r_0, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^y}\big|R_0 = r_0\big) \\
&\qquad \times \frac{\frac{P\left(R_0 = r_0', M_1 = m_{1j}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}}{\frac{P\left(R_0 = r_0, M_1 = m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}} \times \frac{P(r_0)}{P\big(r_0\big|\boldsymbol{a_1^y}\big)} \\
&= \textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\big[Y_2\big|R_0 = r_0, m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big]\, P\big(M_1 = m_{1j}\big|R_0 = r_0, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{n_1}\big|R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^m}\big|R_0 = r_0, \boldsymbol{a_1^y}\big)P\big(\boldsymbol{a_1^y}\big|R_0 = r_0\big) \\
&\qquad \times \frac{\frac{P\left(R = r_0'\middle|m_{1j}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}}{\frac{P\left(R = r_0'\middle|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}} \times \frac{P\big(M_1 = m_{1j}\big|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)}{P\big(M_1 = m_{1j}\big|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)} \times \frac{P(r_0)}{P\big(r_0\big|\boldsymbol{a_1^y}\big)} \\
&= E\big[E\big[Y_2 \times w_{r_0}^{iorw}\big|r_0, m_1, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big]|r_0\big] \qquad (8a)
\end{aligned}$$

$$\text{where } w_{r_0}^{iorw} = \frac{\frac{P\left(R = r_0'\middle|m_{1j}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}}{\frac{P\left(R = r_0'\middle|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}} \times \frac{P\big(M_1 = m_{1j}\big|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)}{P\big(M_1 = m_{1j}\big|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)} \times \frac{P(r_0)}{P\big(r_0\big|\boldsymbol{a_1^y}\big)} \qquad (8b)$$

The first equality is identified via eqn 1.

The second and sixth equalities show that $w_{r_0}^{iorw} = w_{r_0}^{rmpw}$ non-parametrically. Thus, we can implement IORW approach by following the procedure outlined with RMPW, replacing $w_{r_0}^{rmpw}$ (4b) by $w_{r_0}^{iorw}$ (8b).

Remark 8. The conditionality of the intervention $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ appears through the numerators in (8b). Any non-allowable confounders $\boldsymbol{N_1}$ beyond the allowable variables defined in $\boldsymbol{A_1^m}$ and $\boldsymbol{A_1^y}$ appear only in the denominators. Thus, the conditionality of the numerators will differ from the denominators whenever the intervention $G_{m_1|\boldsymbol{a_1^m a_1^y}}$ does not condition on all of the confounders of M_1.

Remark 9. The weight in (8b) is based on disparity measures that use the pooled distribution $P(\boldsymbol{a_1^y})$ to standardize the outcome-allowable covariates. With $P(\boldsymbol{a_1^y}|R_0 = r_0)$ as the standard the weight would be:

$$w_{r_0}^{iorw} = \frac{\frac{P\left(R = r_0'\middle|m_{1j}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}}{\frac{P\left(R = r_0'\middle|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}{P\left(R = r_0\middle|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)}} \times \frac{P\big(M_1 = m_{1j}\big|\boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)}{P\big(M_1 = m_{1j}\big|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\big)}$$

With $P(\boldsymbol{a_1^y}|R_0 = r_0')$ as the standard the weight would be:

$$\frac{\frac{P(R = r_0'|m_{1j}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(R = r_0|m_{1j}, \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}}{\frac{P(R = r_0'|\boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(R = r_0|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})}} \times \frac{P(M_1 = m_{1j}|\boldsymbol{a_1^m}, \boldsymbol{a_1^y})}{P(M_1 = m_{1j}|\boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})} \times \frac{P(r_0'|\boldsymbol{a_1^y})}{P(r_0|\boldsymbol{a_1^y})} \times \frac{P(r_0)}{P(r_0')}$$

Implementation

The sketch for parametric g-computation in the main-text was based on models using the factorizations (1), (2), and (3). However, one can replace (3) with (3*). If that is done, the outcome models for blacks and whites would always condition on allowable and non-allowable covariates. Also, the target factor models would always condition on the allowable and non-allowable covariates among the observed scenarios for blacks and whites, but only condition on the allowables in the counterfactual scenario for blacks. This alternate specification can lead to issues with non-compatibility, as it may be difficult to specify models for $P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ (used for estimating the counterfactual scenario for blacks under (3*)) and $P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{n_1}, \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ (used for estimating the observed scenario for whites under (3*)) that are compatible with one another. This challenge does not arise when (3) is used because then only $P(M_1 = m_{1j}|R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y})$ must be specified.

Ratio of Mediator Probability Weighting

The first component of the weight $w_{r_0}^{rmpw}$ is a ratio of two probabilities. The numerator could be estimated by fitting, among whites, a logistic regression model for the probability of treatment intensification M_1 given the allowable covariates $\boldsymbol{A_1^m}$ and $\boldsymbol{A_1^y}$. The denominator could be estimated by fitting an analogous model among blacks that further conditions on non-allowable confounders $\boldsymbol{N_1}$. These models need not be compatible when fitted this way, separately for blacks and whites. The second component of the weight $w_{r_0}^{rmpw}$ could be obtained with logistic regression models for race R_0 that do and do not control for the outcome-allowable covariates $\boldsymbol{A_1^y}$. The predicted values from these four models are used to obtain the weight $w_{r_0}^{rmpw}$ for each individual.

A stacking procedure can be used to estimate the effects of interest. To obtain the disparity reduction (2) minus (1), the data from blacks with weight w_{r_0} (5b) are stacked onto a copy from blacks with weight $w_{r_0}^{rmpw}$ (4b) and labelled with a new variable called data origin (D_0; 1=original, 0=copy). The weighted mean difference in Y_2 across data origin D_0 estimates the disparity reduction. To obtain the disparity residual (1) minus (3), the data from blacks with weight $w_{r_0}^{rmpw}$ (4b) are stacked onto the data from whites with weight $w_{r_0'}$ (6b). The weighted mean difference in Y_2 across race R_0 estimates the disparity residual. For inference, the non-parametric bootstrap could be used to obtain 95% confidence intervals.

Inverse Odds Ratio Weighting

The first component of $w_{r_0}^{iorw}$ is a ratio of two odds. The numerator odds can be estimated by fitting logistic regressions for race R_0 given treatment intensification M_1, allowable covariates $\boldsymbol{A_1^m}$ and $\boldsymbol{A_1^y}$ with and without further control for non-allowable confounders $\boldsymbol{N_1}$. For the denominator odds one can use similar models but without control for treatment intensification M_1. For the second and third components one can adapt what was described for the RMPW-style estimator, with the caveat that the models for treatment intensification M_1 do not condition on race R_0. As noted in the main text, the estimation procedure is valid if all models are specified correctly, and here special care should be taken to ensure that models are compatible with one another. For guidance, see the procedure proposed by Miles et al. Once all necessary models are fit, their predicted values are used to form individual weights. The stacking procedure described above is used but replacing $w_{r_0}^{rmpw}$ weights (4b) with $w_{r_0}^{iorw}$ weights (8b).

Relation to Existing Estimators (under identifying formulae (1), (2), and (3*))

In what follows we make the notation more compact as follows: $X_0^{age} = X_0^g$, $X_0^{sex} = X_0^s$, $X_0^{edu} = X_0^e$, $X_0^{ins} = X_0^i$, $X_0^{dia} = X_0^d$, *with sets notated as, e.g.,* $X_0^g, X_0^s = X_0^{g,s}$. *In weight expressions,* M_1 *is categorical with* j *levels* m_{1j}.

Interventional Analogue of the Natural Indirect Effect

Suppose we estimate the disparity reduction where $\boldsymbol{A_1^y}$, the covariates deemed both outcome- and target-allowable, includes all covariates. This leaves $\boldsymbol{A_1^m}$ empty because we have exhausted the potential covariates that could be deemed target-allowable. This leaves $\boldsymbol{N_1}$ empty because we have exhausted the covariates needed to establish conditional exchangeability for M_1. The disparity reduction is identified by the non-parametric expression of Pearl and the weighting estimators of Hong (2010) and (2015), Huber, Lange et al., and Tchetgen Tchetgen.

Non-parametric

$$\begin{aligned}\psi^{red} = \sum_{m_1,l_1,x_0^{g,s,e,i,d}} & E\left[Y_2 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \\ & \times \{P\left(M_1 = m_1 \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right) - P\left(M_1 = m_1 \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)\} \\ & \times P(l_1, x_0^{g,s,e,i,d})\end{aligned}$$

A conditional expression is obtained by removing the integration over $l_1, x_0^{g,s,e,i,d}$. This expression is equivalent to the mediation formula of Pearl, which underlies the regression-based estimators of Valeri and VanderWeele, as well as the simulation-based estimators of Imai et al. and Wang et al. The marginal expression serves as the basis for the imputation estimator of Albert, and VanderWeele and Vansteelandt.

Ratio of Mediator Probability Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \middle| r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{rmpw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \middle| r_0\right]$$

Where

$$w_{r_0} = \frac{P(r_0)}{P\left(r_0 \middle| l_1, x_0^{g,s,e,i,d}\right)} \qquad w_{r_0}^{rmpw} = \frac{P\left(M_1 = m_{1j} \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)}{P\left(M_1 = m_{1j} \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right)} \times w_{r_0}$$

A conditional expression is obtained by removing the outer expectation and setting $w_{r_0} = \mathrm{w}_{\mathrm{r}_0'} = 1$. The marginal and conditional versions are equivalent to the weighting approaches of Hong (2010) and (2015) and those used in the natural effect models of Lange et al.

Inverse Odds Ratio Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \middle| r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{iorw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \middle| r_0\right]$$

Where

$$w_{r_0} = \frac{P(r_0)}{P\left(r_0 \middle| l_1, x_0^{g,s,e,i,d}\right)} \qquad w_{r_0}^{iorw} = \frac{\frac{P\left(R = r_0' \middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0 \middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}}{\frac{P\left(R = r_0' \middle| l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0 \middle| l_1, x_0^{g,s,e,i,d}\right)}} \times w_{r_0}$$

This is equivalent to the approach of Huber. A conditional expression is obtained by removing the outer expectation and setting $\mathrm{w_r} = \mathrm{w_r'} = 1$ which is related to a variant of the approach of Tchetgen Tchetgen.

Interventional Analogue of the Path-Specific Indirect Effect I

Suppose we estimate the disparity reduction where $\boldsymbol{A_1^y}$, the covariates deemed both outcome- and target-allowable, include $X^{g,s}$. $\boldsymbol{A_1^m}$ is left empty so that no additional variables are considered target-allowable, and $\boldsymbol{N_1}$ includes all other variables needed to establish conditional exchangeability for M_1 (i.e., $L_1, X_0^{e,i,d}$). The disparity reduction is identified by the non-parametric expression and weighting estimator for the interventional indirect effect of VanderWeele, Vansteelandt, and Robins.

Non-parametric

$$\begin{aligned}\psi^{red} = \sum_{m_1,l_1,x_0^{g,s,e,i,d}} & E\left[Y_2 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \\ & \times \{P\left(M_1 = m_1 \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right) - P\left(M_1 = m_1 \middle| R_0 = r_0', x_0^{g,s}\right)\} \\ & \times P(l_1, x_0^{e,i,d} | R_0 = r_0, x_0^{g,s}) \\ & \times P(x_0^{g,s})\end{aligned}$$

A conditional expression is obtained by removing the integration over $x_0^{g,s}$. This is equivalent to the expression of VanderWeele, Vansteelandt, and Robins under a stochastic intervention.

Ratio of Mediator Probability Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{rmpw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right]$$

Where

$$w_{r_0} = \frac{P(r_0)}{P\left(r_0 \middle| x_0^{g,s}\right)} \qquad w_{r_0}^{rmpw} = \frac{P\left(M_1 = m_{1j} \middle| R_0 = r_0', x_0^{g,s}\right)}{P\left(M_1 = m_{1j} \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right)} \times w_{r_0}$$

A conditional expression is obtained by conditioning the outer expectation on $x_0^{g,s}$ and setting $\mathrm{w_r} = \mathrm{w_r'} = 1$. This is equivalent to the approach of VanderWeele, Vansteelandt, and Robins under a stochastic intervention. They express $P\left(M_1 = m_{1j} \middle| R_0 = r_0', x_0^{g,s}\right)$ as $\sum_{l_1, x_0^{e,i,d}} P\left(M_1 = m_{1j} \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right) P\left(l_1, x_0^{e,i,d} \middle| R_0 = r_0', x_0^{g,s}\right)$ to emphasize that $P\left(M_1 = m_{1j} \middle| R_0 = r_0', x_0^{g,s}\right)$ represents a marginalization of $P\left(M_1 = m_{1j} \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)$ over $P\left(l_1, x_0^{e,i,d} \middle| R_0 = r_0', x_0^{g,s}\right)$ rather than conditional independence of M_1 and $\{l_1, x_0^{e,i,d}\}$ given $R = r_0'$ and $x_0^{g,s}$.

Interventional Analogue of the Path-Specific Indirect Effect II

The simulation-based estimator of the interventional indirect effect of Vansteelandt and Daniel does not generally estimate the disparity reduction, but rather a contrast of two interventions.

To see this, consider two stochastic interventions. The first intervention, $G^{r_0}_{m_1|\boldsymbol{a_1^m a_1^y}}$, assigns treatment intensification, the targeted factor, according to its conditional distribution among blacks given the allowables $\boldsymbol{A_1^y}$ and $\boldsymbol{A_1^m}$, defined as $P\left(M_1 = m_1 \middle| R_0 = r_0, \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)$. The second intervention, $G^{r_0'}_{m_1|\boldsymbol{a_1^m a_1^y}}$, assigns treatment intensification, the targeted factor, according to its conditional distribution among whites given the allowables $\boldsymbol{A_1^y}$ and $\boldsymbol{A_1^m}$, defined as $P\left(M_1 = m_1 \middle| R_0 = r_0', \boldsymbol{a_1^m}, \boldsymbol{a_1^y}\right)$.

According to assumptions A, a slight variant of B1, and C, under the first intervention, the proportion of blacks with uncontrolled hypertension, standardized by the outcome-allowable covariates, is:

$$
\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\right)\middle|R_0=r_0,\boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0=r_0,G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1,\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P\left(G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\middle|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right)P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P\left(G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\middle|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right)P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m}} E\left[Y_2(m_1)\middle|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P(M_1=m_1|R_0=r_0,\boldsymbol{a_0^m},\boldsymbol{a_0^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2(m_1)\middle|R_0=r_0,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P(M_1=m_1|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2(m_1)\middle|R_0=r_0,m_1,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P(M_1=m_1|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0=r_0,m_1,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P(M_1=m_1|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})
\end{aligned}
\tag{9}
$$

According to equation (1), under assumptions A, B1, B2, and C, under the second intervention, the proportion of blacks with uncontrolled hypertension, standardized by the outcome-allowable covariates, is:

$$
\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0'}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\right)\middle|R_0=r_0,\boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0=r_0,m_1,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]P(M_1=m_1|R_0=r_0',\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{n_1}|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})
\end{aligned}
$$

The difference in uncontrolled hypertension among blacks comparing the two interventions is:

$$
\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\right)\middle|R_0=r_0,\boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})-\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0'}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}=m_1\right)\middle|R_0=r_0,\boldsymbol{a_1^y}\right]P(\boldsymbol{a_1^y})\\
&\quad=\textstyle\sum_{m_1,\boldsymbol{a_1^y},\boldsymbol{a_1^m},\boldsymbol{n_1}} E\left[Y_2|R_0=r_0,m_1,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y}\right]\\
&\qquad\qquad\times\{P(M_1=m_1|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})-P(M_1=m_1|R_0=r_0',\boldsymbol{a_1^m},\boldsymbol{a_1^y})\}\\
&\qquad\qquad\times P(\boldsymbol{n_1}|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})P(\boldsymbol{a_1^m}|R_0=r_0,\boldsymbol{a_1^y})P(\boldsymbol{a_1^y})
\end{aligned}
\tag{10}
$$

Note that (10) does not generally equal to the disparity reduction (2) minus (1) because the first intervention $G^{r_0}_{m_1|\boldsymbol{a_1^m}\boldsymbol{a_1^y}}$, within levels of the allowables $\boldsymbol{A_1^y}$ and $\boldsymbol{A_1^m}$ breaks any dependence of treatment intensification M_1 on non-allowables $\boldsymbol{N_1}$, whereas in the observed scenario (2) this dependence is present. They are equivalent when
$P(M_1=m_1|R_0=r_0,\boldsymbol{a_1^m},\boldsymbol{a_1^y})=P(M_1=m_1|R_0=r_0,\boldsymbol{n_1},\boldsymbol{a_1^m},\boldsymbol{a_1^y})$.

Suppose we estimate this difference where $\boldsymbol{A_1^y}$, the covariates deemed both outcome- and target-allowable include $X_0^{g,s}$. $\boldsymbol{A_1^m}$ is left empty so that no additional variables are considered target-allowable, and $\boldsymbol{N_1}$ includes all other variables needed to establish conditional exchangeability for M_1 (i.e., $L_1, X_0^{e,i,d}$). Now, under these allowability choices, the difference between the first and second interventions is:

$$
\begin{aligned}
&\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0}_{m_1|x_0^{g,s}}=m_1\right)\middle|R_0=r_0,x_0^{g,s}\right]P(x_0^{g,s})-\textstyle\sum_{\boldsymbol{a_1^y}} E\left[Y_2\left(G^{r_0'}_{m_1|x_0^{g,s}}=m_1\right)\middle|R_0=r_0,x_0^{g,s}\right]P(x_0^{g,s})\\
&\quad=\textstyle\sum_{m_1,l_1,x_0^{g,s,e,i,d}} E\left[Y_2|R_0=r_0,m_1,l_1,x_0^{g,s,e,i,d}\right]\\
&\qquad\qquad\times\{P(M_1=m_1|R_0=r_0,x_0^{g,s})-P(M_1=m_1|R_0=r_0',x_0^{g,s})\}\\
&\qquad\qquad\times P(l_1,x_0^{e,i,d}|R_0=r_0,x_0^{g,s})P(x_0^{g,s})
\end{aligned}
$$

This last expression is equivalent to the identification formula for the interventional indirect effect of Vansteelandt and Daniel for the terminal mediator when applied to our motivating example. Again, this does not estimate the disparity reduction because $P(M_1=m_1|R_0=r_0,x_0^{g,s})\neq P(M_1=m_1|R_0=r_0,l_1,x_0^{g,s,e,i,d})$.

Interventional Analogue of the Path-Specific Indirect Effect III

Suppose we estimate the disparity reduction where $\boldsymbol{A_1^y}$, the covariates deemed both outcome- and target-allowable, include $X^{g,s}$. $\boldsymbol{A_1^m}$, the additional covariates deemed target allowable, includes all other covariates (i.e.,$L_1, X_0^{e,i,d}$). $\boldsymbol{N_1}$ is left empty since in this specific case conditional exchangeability among blacks has been established for M_1 given $\boldsymbol{A_1^m}$ and $\boldsymbol{A_1^y}$. The disparity reduction is identified by the non-parametric expressions and weighting approaches of Zheng and van der Laan and also Miles et al.

Non-parametric

$$\begin{aligned}\psi^{red} = \sum_{m_1,l_1,x_0^{g,s,e,i,d}} & E\left[Y_2 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \\ & \times \{P\left(M_1 = m_1 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right) - P\left(M_1 = m_1 \middle| R_0 = r_0', m_1, l_1, x_0^{g,s,e,i,d}\right)\} \\ & \times P(l_1, x_0^{e,i,d} | R_0 = r_0, x_0^{g,s}) \\ & \times P(x_0^{g,s})\end{aligned}$$

A conditional expression is obtained by removing the integration over $x_0^{g,s}$. This is equivalent to the non-parametric expression of a path-specific effect discussed in Jackson 2018.

Ratio of Mediator Probability Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{rmpw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right]$$

Where

$$w_{r_0} = \frac{P(r_0)}{P\left(r_0 \middle| x_0^{g,s}\right)} \qquad w_{r_0}^{rmpw} = \frac{P\left(M_1 = m_{1j} \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)}{P\left(M_1 = m_{1j} \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right)} \times w_{r_0}$$

A conditional expression is obtained by conditioning the outer expectation on $x_0^{g,s}$ and setting $\mathrm{w_r} = \mathrm{w_r'} = 1$. This is related to the weighting approach of Zheng and Van der Laan (2017).

Inverse Odds Ratio Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{iorw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right]$$

Where

$$w_{r_0} = \frac{P(r_0)}{P\left(r_0 \middle| x_0^{g,s}\right)} \qquad w_{r_0}^{iorw} = \frac{\frac{P\left(R = r_0' \middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0 \middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}}{\frac{P\left(R = r_0' \middle| l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0 \middle| l_1, x_0^{g,s,e,i,d}\right)}} \times w_{r_0}$$

A conditional expression is obtained by conditioning the outer expectation on $x_0^{g,s}$ and setting $\mathrm{w_r} = \mathrm{w_r'} = 1$. This is equivalent to the "m-ratio" weighting approach proposed by Miles et al. albeit under an alternate coding for race R_0. (Note that the specification by Miles et al. would code blacks as $R_0 = r_0'$ and whites as $R_0 = r_0$, mapping to a path-specific effect whose analog imagines an intervention upon whites by fixing the conditional distribution of the target to match that of blacks). Our coding scheme maps to an identification formula for a path-specific effect discussed in Jackson 2018, wherein blacks are intervened upon by fixing the conditional distribution of the target to match that of whites.

"Detailed" Oaxaca-Blinder Decomposition

Suppose we estimate the disparity reduction where no covariates are deemed outcome- or target-allowable. All covariates are included in $\boldsymbol{N_1}$ to establish exchangeability for M_1. The disparity reduction is identified by a "detailed" Oaxaca-Blinder Decomposition implemented with linear models.

The non-parametric formula is:

$$\psi^{red} = \sum_{m_1, l_1, x_0^{g,s,e,i,d}} E\left[Y_2 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \times \{P\left(M_1 = m_1 \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right) - P(M_1 = m_1 | R_0 = r_0')\} \times P(l_1, x_0^{g,s,e,i,d} | R_0 = r_0)$$

Consider the following linear models:

$$E[Y_2 | R_0 = r_0, m_{1j}, l_1, x_0^{g,s,e,i,d}] = \beta_0^{r_0} + \sum_{j \neq ref} \beta_{1j}^{r_0} I\left(M_1 = m_{1j}\right) + \beta_2^{r_0} L_1 + \sum_k \beta_3^k X^k$$

$$E[Y_2 | R_0 = r_0', m_{1j}, l_1, x_0^{g,s,e,i,d}] = \beta_0^{r_0'} + \sum_{j \neq ref} \beta_{1j}^{r_0'} I\left(M_1 = m_{1j}\right) + \beta_2^{r_0'} L_1 + \sum_k \beta_3^k X^k$$

where, with a slight abuse of notation, X^k is the kth element of $X_0^{g,s,e,i,d}$.

It follows from the arguments of Jackson and VanderWeele 2018 we that:

$$\psi^{red} = \sum_{j \neq ref} \beta_{1j}^{r_0} \{P\left(M = m_{1j} \middle| R_0 = r_0\right) - P\left(M = m_{1j} \middle| R_0 = r_0'\right)\}$$

This is the typical formulation of a detailed Oaxaca-Blinder Decomposition under linear models. Alternate implementations of the Oaxaca-Blinder Decomposition make different allowability choices. For example, suppose we estimate the disparity reduction where no covariates are deemed outcome-allowable but all covariates are considered target-allowable, leaving $\boldsymbol{N_1}$ empty. The disparity reduction is identified by the following non-parametric formula which leads to adaptations of the weighting estimators of Dinardo et al. (a form of ratio of mediator probability weighting) and Barsky et al. (a form of inverse odds ratio weighting).

Non-parametric

$$\psi^{red} = \sum_{m_1, l_1, x_0^{g,s,e,i,d}} E\left[Y_2 \middle| R_0 = r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] \times \{P\left(M_1 = m_1 \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right) - P\left(M_1 = m_1 \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)\} \times P(l_1, x_0^{g,s,e,i,d} | R_0 = r_0)$$

Ratio of Mediator Probability Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{rmpw} \middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right] | r_0\right]$$

Where

$$w_{r_0} = 1 \qquad w_{r_0}^{rmpw} = \frac{P\left(M_1 = m_{1j} \middle| R_0 = r_0', l_1, x_0^{g,s,e,i,d}\right)}{P\left(M_1 = m_{1j} \middle| R_0 = r_0, l_1, x_0^{g,s,e,i,d}\right)} \times w_{r_0}$$

This is equivalent to an extension of the weighting approach proposed by Dinardo, Fortin and Lemieux where the conditioning events of the numerator and denominator of $w_{r_0}^{rmpw}$ include all covariates.

Inverse Odds Ratio Weighting

$$\psi^{red} = E\left[E\left[Y_2 \times w_{r_0}\middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right]|r_0\right] - E\left[E\left[Y_2 \times w_{r_0}^{iorw}\middle| r_0, m_1, l_1, x_0^{g,s,e,i,d}\right]|r_0\right]$$

Where

$$w_{r_0} = 1 \qquad w_{r_0}^{iorw} = \frac{\frac{P\left(R = r_0'\middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0\middle| m_{1j}, l_1, x_0^{g,s,e,i,d}\right)}}{\frac{P\left(R = r_0'\middle| l_1, x_0^{g,s,e,i,d}\right)}{P\left(R = r_0\middle| l_1, x_0^{g,s,e,i,d}\right)}} \times w_{r_0}$$

This is equivalent to an extension of the weighting approach proposed by Barsky et al., and also one discussed by Dinardo, Fortin and Lemieux, where the conditioning events of the numerator and denominator of $w_{r_0}^{iorw}$ include all covariates.

eAppendix References

Albert JM. Distribution-free mediation analysis for nonlinear models with confounding. Epidemiology 2012;23(6):879-88.

Barsky R, Bound J, Charles KK, Lupton JP. Accounting for the Black-White Wealth Gap: A Nonparametric Approach. Journal of the American Statistical Association 2002;97(459):663-673.

Blinder A. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources* 1973;8(4):436.

Dinardo J, Fortin N, Lemieux T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica 1996;64(5):1001-1044.

Hong G. Ratio of mediator probability weighting for estimating natural direct and indirect effects. *Joint Statistical Meeting*. Alexandria, Virginia: American Statistical Association, 2010;2401-2415.

Hong G, Deutsch J, Hill HD. Ratio-of-Mediator-Probability Weighting for Causal Mediation in the Presence of Treatment-by-Mediator Interaction. *Journal of Educational and Behavioral Statistics* 2015;40(3):307-340.

Huber M. Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics* 2014;29(6):920-943.

Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010;15(4):309-34.

Jackson JW. On the Interpretation of Path-specific Effects in Health Disparities Research. Epidemiology 2018;29(4):517-520.

Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012;176(3):190-5.

Machada J, Mata JM. Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. Journal of Applied Econometrics. J Appl Econ. 2005;20:445-465.

Miles CH, Shpitser I, Kanki P, Meloni S, Tchetgen Tchetgen EJ. On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. Biometrika 2019;107(1):159-172.

Oaxaca R. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 1973;14(3):693-709.

Pearl J. Direct and indirect effects. In: Breese K, Koller D, eds. Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann, 2001;411-420.

Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Stat Med* 2013;32(26):4567-80.

Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013;18(2):137-50.

VanderWeele TJ, Vansteelandt S. Mediation Analysis with Multiple Mediators. Epidemiol Methods. 2014;2(1):95-115.

Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* 2014;25(2):300-6.

Vansteelandt S, Daniel RM. Interventional Effects for Mediation Analysis with Multiple Mediators. Epidemiology 2017;28(2):258-265.

Wang A, Arah OA. G-computation demonstration in causal mediation analysis. Eur J Epidemiol 2015;30(10):1119-27.

Zheng W, van der Laan M. Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes. *J Causal Inference* 2017;5(2).