Biometrika (2017), **xx**, x, *pp*. 1–13
Printed in Great Britain

Supplementary material for "Instrumental variables as bias amplifiers with general outcome and confounding"

BY P. DING
Department of Statistics, University of California, Berkeley, California, USA.
pengdingpku@berkeley.edu

T. J. VANDERWEELE AND J. M. ROBINS
Departments of Epidemiology and Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA.
tvanderw@hsph.harvard.edu robins@hsph.harvard.edu

APPENDIX 1. LEMMAS AND THEIR PROOFS

In order to prove the main results, we need to invoke the following lemmas. Some of them are from the literature, and some of them are new and of independent interest.

Lemma S1 is from Esary et al. (1967, Theorem 2.1).

LEMMA S1. *Let $f(\cdot)$ and $g(\cdot)$ be functions with K real-valued arguments, which are both non-decreasing in each of their arguments. If $U = (U_1, \ldots, U_K)$ is a multivariate random variable with K mutually independent components, then* $\text{cov}\{f(U), g(U)\} \geq 0$.

Lemma S2 is from VanderWeele (2008), and Lemmas S3 and S4 are from Chiba (2009).

LEMMA S2. *For a univariate U or a multivariate U with mutually independent components, if for $a = 1$ and 0, $Y(a) \perp\!\!\!\perp A \mid U$, $E(Y \mid A = a, U = u)$ is non-decreasing in each component of u, and* $\text{pr}(A = 1 \mid U = u)$ *is non-decreasing in each component of u, then $E(Y \mid A = 1) \geq E\{Y(1)\}$ and $E(Y \mid A = 0) \leq E\{Y(0)\}$.*

LEMMA S3. *For a univariate U and a multivariate U with mutually independent components, if $Y(0) \perp\!\!\!\perp A \mid U$, $E(Y \mid A = 0, U = u)$ is non-decreasing in each component of u, and* $\text{pr}(A = 1 \mid U = u)$ *is non-decreasing in each component of u, then $E(Y \mid A = 0) \leq E\{Y(0) \mid A = 1\}$.*

LEMMA S4. *For a univariate U and a multivariate U with mutually independent components, if $Y(1) \perp\!\!\!\perp A \mid U$, $E(Y \mid A = 1, U = u)$ is non-decreasing in each component of u, and* $\text{pr}(A = 1 \mid U = u)$ *is non-decreasing in each component of u, then $E(Y \mid A = 1) \geq E\{Y(1) \mid A = 0\}$.*

Lemma S5, extending Rothman et al. (2008), states that under monotonicity, no additive interaction implies non-positive multiplicative interactions for both presence and absence of the outcome.

LEMMA S5. *If $p_{11} \geq \max(p_{10}, p_{01})$, $\min(p_{10}, p_{01}) \geq p_{00} > 0$, and $p_{11} - p_{10} - p_{01} + p_{00} = 0$, then*

$$\frac{p_{11}p_{00}}{p_{10}p_{01}} \leq 1, \quad \frac{(1-p_{11})(1-p_{00})}{(1-p_{10})(1-p_{01})} \leq 1. \tag{S1}$$

Proof of Lemma S5. Define $\text{RR}_{11} = p_{11}/p_{00} \geq 1$, $\text{RR}_{10} = p_{10}/p_{00} \geq 1$ and $\text{RR}_{01} = p_{01}/p_{00} \geq 1$. Then $p_{11} - p_{10} - p_{01} + p_{00} = 0$ implies $\text{RR}_{11} = \text{RR}_{10} + \text{RR}_{01} - 1$, which further implies

$$\frac{p_{11}p_{00}}{p_{10}p_{01}} = \frac{\mathrm{RR}_{11}}{\mathrm{RR}_{10}\mathrm{RR}_{01}} = 1 + \frac{1}{\mathrm{RR}_{10}\mathrm{RR}_{01}}(\mathrm{RR}_{11} - \mathrm{RR}_{10}\mathrm{RR}_{01})$$
$$= 1 + \frac{1}{\mathrm{RR}_{10}\mathrm{RR}_{01}}(\mathrm{RR}_{10} + \mathrm{RR}_{01} - 1 - \mathrm{RR}_{10}\mathrm{RR}_{01})$$
$$= 1 - \frac{1}{\mathrm{RR}_{10}\mathrm{RR}_{01}}(\mathrm{RR}_{10} - 1)(\mathrm{RR}_{01} - 1) \leq 1.$$

The second inequality of (S1) follows from

$$\frac{(1-p_{11})(1-p_{00})}{(1-p_{10})(1-p_{01})} = 1 + \frac{(1-p_{11})(1-p_{00}) - (1-p_{10})(1-p_{01})}{(1-p_{10})(1-p_{01})}$$
$$= 1 + \frac{1}{(1-p_{10})(1-p_{01})}\{(1 - p_{11} - p_{00} + p_{11}p_{00}) - (1 - p_{10} - p_{01} + p_{10}p_{01})\}$$
$$= 1 + \frac{1}{(1-p_{10})(1-p_{01})}(p_{11}p_{00} - p_{10}p_{01})$$
$$= 1 + \frac{p_{10}p_{01}}{(1-p_{10})(1-p_{01})}\left(\frac{p_{11}p_{00}}{p_{10}p_{01}} - 1\right) \leq 1. \qquad \square$$

Lemma S5 is about interaction between two binary causes, and for our discussion we need to extend it to interaction between two general causes. Lemma S6 extends Piegorsch et al. (1994) and Yang et al. (1999) by relating the conditional association between two independent causes given the outcome to the interaction between the two causes on the outcome.

LEMMA S6. *If* $Z \perp\!\!\!\perp U$, *and* $\mathrm{pr}(A = 1 \mid Z = z, U = u) = \beta(z) + \gamma(u)$ *with* $\beta(z)$ *and* $\gamma(u)$ *non-decreasing in* z *and* u, *then for both* $a = 1$ *and* 0 *and for all values of* u *and* z,

$$\frac{\partial F(u \mid A = a, Z = z)}{\partial z} \geq 0,$$

i.e., U has non-positive distributional dependence on Z, given A.

Proof of Lemma S6. For a fixed u and $z_1 > z_0$, we define

$$p_{11} = \mathrm{pr}(A = 1 \mid U > u, Z = z_1) = \int_u^\infty \{\beta(z_1) + \gamma(u')\}F(\mathrm{d}u')/\{1 - F(u)\},$$
$$p_{10} = \mathrm{pr}(A = 1 \mid U > u, Z = z_0) = \int_u^\infty \{\beta(z_0) + \gamma(u')\}F(\mathrm{d}u')/\{1 - F(u)\},$$
$$p_{01} = \mathrm{pr}(A = 1 \mid U \leq u, Z = z_1) = \int_{-\infty}^u \{\beta(z_1) + \gamma(u')\}F(\mathrm{d}u')/F(u),$$
$$p_{00} = \mathrm{pr}(A = 1 \mid U \leq u, Z = z_0) = \int_{-\infty}^u \{\beta(z_0) + \gamma(u')\}F(\mathrm{d}u')/F(u),$$

following from the additive model of A and $Z \perp\!\!\!\perp U$.

Because $\beta(z_1) \geq \beta(z_0)$, it is straightforward to show that $p_{11} \geq p_{10}$ and $p_{01} \geq p_{00}$. Because $\gamma(u)$ is increasing in u, we have

$$p_{11} \geq \beta(z_1) + \gamma(u), \quad p_{10} \geq \beta(z_0) + \gamma(u), \quad p_{01} \leq \beta(z_1) + \gamma(u), \quad p_{00} \leq \beta(z_0) + \gamma(u),$$

which imply $p_{11} \geq p_{01}$ and $p_{10} \geq p_{00}$. We further have

$$
\begin{aligned}
& p_{11} - p_{10} - p_{01} + p_{00} \\
= & \int_u^{\infty} \{\beta(z_1) - \beta(z_0)\} F(\mathrm{d}u')/\{1 - F(u)\} - \int_{-\infty}^{u} \{\beta(z_1) - \beta(z_0)\} F(\mathrm{d}u')/F(u) \\
= & \, 0.
\end{aligned}
$$

The four probabilities $(p_{11}, p_{10}, p_{01}, p_{00})$ satisfy the conditions in Lemma S5, Therefore, (2) holds. Replacing the probabilities in (2) by their definitions above, we have

$$
\begin{aligned}
& \frac{\mathrm{pr}(A = 1 \mid U > u, Z = z_1)\mathrm{pr}(A = 1 \mid U \leq u, Z = z_0)}{\mathrm{pr}(A = 1 \mid U > u, Z = z_0)\mathrm{pr}(A = 1 \mid U \leq u, Z = z_1)} \leq 1 \\
\Longleftrightarrow & \frac{\mathrm{pr}(A = 1 \mid U > u, z_1)}{\mathrm{pr}(A = 1 \mid U \leq u, z_1)} \leq \frac{\mathrm{pr}(A = 1 \mid U > u, z_0)}{\mathrm{pr}(A = 1 \mid U \leq u, z_0)},
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{\mathrm{pr}(A = 0 \mid U > u, Z = z_1)\mathrm{pr}(A = 0 \mid U \leq u, Z = z_0)}{\mathrm{pr}(A = 0 \mid U > u, Z = z_0)\mathrm{pr}(A = 0 \mid U \leq u, Z = z_1)} \leq 1 \\
\Longleftrightarrow & \frac{\mathrm{pr}(A = 0 \mid U > u, z_1)}{\mathrm{pr}(A = 0 \mid U \leq u, z_1)} \leq \frac{\mathrm{pr}(A = 0 \mid U > u, z_0)}{\mathrm{pr}(A = 0 \mid U \leq u, z_0)}.
\end{aligned}
$$

Therefore, for both $a = 1$ and 0 and for all values of u,

$$
\frac{\mathrm{pr}(A = a \mid U > u, Z = z)}{\mathrm{pr}(A = a \mid U \leq u, Z = z)} \tag{S2}
$$

is non-increasing in z. Because of the independence of Z and U, we have

$$
\begin{aligned}
& F(u \mid A = a, Z = z) \\
= & \frac{\mathrm{pr}(U \leq u, A = a \mid Z = z)}{\mathrm{pr}(A = a \mid Z = z)} \\
= & \frac{\mathrm{pr}(U \leq u)\mathrm{pr}(A = a \mid U \leq u, Z = z)}{\mathrm{pr}(U \leq u)\mathrm{pr}(A = a \mid U \leq u, Z = z) + \mathrm{pr}(U > u)\mathrm{pr}(A = a \mid U > u, Z = z)} \\
= & \left\{1 + \frac{\mathrm{pr}(U > u)}{\mathrm{pr}(U \leq u)} \times \frac{\mathrm{pr}(A = a \mid U > u, Z = z)}{\mathrm{pr}(A = a \mid U \leq u, Z = z)}\right\}^{-1}.
\end{aligned}
$$

Therefore, $F(u \mid A = a, Z = z)$ is a non-increasing function of (S2), and the conclusion holds.$\square$

Lemmas S5 and S6 above hold under the assumption of no additive interaction, and the following two lemmas state similar results under the assumption of no multiplicative interaction.

LEMMA S7. *If* $p_{11} \geq \max(p_{10}, p_{01}), \min(p_{10}, p_{01}) \geq p_{00}$, *and* $p_{11}p_{00} = p_{10}p_{01}$, *then*

$$
p_{11} - p_{10} - p_{01} + p_{00} \geq 0, \quad \frac{(1 - p_{11})(1 - p_{00})}{(1 - p_{10})(1 - p_{01})} \leq 1.
$$

Proof of Lemma S7. Using the same notation in the proof of Lemma S5, $p_{11}p_{00} = p_{10}p_{01}$ implies $\mathrm{RR}_{11} = \mathrm{RR}_{10}\mathrm{RR}_{01}$, with $\mathrm{RR}_{10} \geq 1, \mathrm{RR}_{01} \geq 1$, and $\mathrm{RR}_{11} \geq 1$. Therefore,

$$
p_{11} - p_{10} - p_{01} + p_{00} = p_{00}(\mathrm{RR}_{10}\mathrm{RR}_{01} - \mathrm{RR}_{10} - \mathrm{RR}_{01} + 1) = p_{00}(\mathrm{RR}_{10} - 1)(\mathrm{RR}_{01} - 1) \geq 0,
$$

which further implies that

$$\frac{(1-p_{11})(1-p_{00})}{(1-p_{10})(1-p_{01})} = 1 + \frac{1}{(1-p_{10})(1-p_{01})}\{(1-p_{11})(1-p_{00}) - (1-p_{10})(1-p_{01})\}$$
$$= 1 - \frac{p_{11}-p_{10}-p_{01}+p_{00}}{(1-p_{10})(1-p_{01})} \leq 1. \qquad \square$$

LEMMA S8. *If* $Z \perp\!\!\!\perp U$, *and* $\mathrm{pr}(A=1 \mid Z=z, U=u) = \beta(z)\gamma(u)$ *with* $\beta(z)>0$ *and* $\gamma(u)>0$ *non-decreasing in* z *and* u*, then* $Z \perp\!\!\!\perp U \mid A=1$*, and for all values of* u *and* z*,*

$$\frac{\partial F(u \mid A=0, Z=z)}{\partial z} \geq 0,$$

i.e., U *has non-positive distributional dependence on* Z*, given* $A=0$*.*

*Proof of Lemma S*8. For a fixed u and $z_1 > z_0$, we define

$$p_{11} = \mathrm{pr}(A=1 \mid U>u, Z=z_1) = \beta(z_1)\int_u^\infty \gamma(u')F(\mathrm{d}u')/\{1-F(u)\},$$
$$p_{10} = \mathrm{pr}(A=1 \mid U>u, Z=z_0) = \beta(z_0)\int_u^\infty \gamma(u')F(\mathrm{d}u')/\{1-F(u)\},$$
$$p_{01} = \mathrm{pr}(A=1 \mid U\leq u, Z=z_1) = \beta(z_1)\int_{-\infty}^u \gamma(u')F(\mathrm{d}u')/F(u),$$
$$p_{00} = \mathrm{pr}(A=1 \mid U\leq u, Z=z_0) = \beta(z_0)\int_{-\infty}^u \gamma(u')F(\mathrm{d}u')/F(u),$$

following from the multiplicative model of A and $Z \perp\!\!\!\perp U$. Because $\beta(z_1)\geq\beta(z_0)$, we have $p_{11}\geq p_{10}$ and $p_{01}\geq p_{00}$. Because $\gamma(u)$ is increasing in u, we have

$$p_{11}\geq \beta(z_1)\gamma(u), \quad p_{10}\geq\beta(z_0)\gamma(u), \quad p_{01}\leq\beta(z_1)\gamma(u), \quad p_{00}\leq\beta(z_0)\gamma(u),$$

which imply $p_{11}\geq p_{01}$ and $p_{10}\geq p_{00}$. We can further verify $(p_{11}p_{00})/(p_{10}p_{01})=1$. Because the four probabilities $(p_{11},p_{10},p_{01},p_{00})$ satisfy the conditions in Lemma S7, we have $\{(1-p_{11})(1-p_{00})\}/\{(1-p_{10})(1-p_{01})\}\leq 1$. Replacing the probabilities by their definitions, we have

$$\frac{\mathrm{pr}(A=1 \mid U>u, Z=z_1)\mathrm{pr}(A=1 \mid U\leq u, Z=z_0)}{\mathrm{pr}(A=1 \mid U>u, Z=z_0)\mathrm{pr}(A=1 \mid U\leq u, Z=z_1)} = 1,$$
$$\frac{\mathrm{pr}(A=0 \mid U>u, Z=z_1)\mathrm{pr}(A=0 \mid U\leq u, Z=z_0)}{\mathrm{pr}(A=0 \mid U>u, Z=z_0)\mathrm{pr}(A=0 \mid U\leq u, Z=z_1)} \leq 1.$$

Following the same logic of the proof of Lemma S6, we can prove that $Z \perp\!\!\!\perp U \mid A=1$, and Z has non-positive distributional association on U, given $A=0$. $\square$

Define $f=\mathrm{pr}(A=1)$ to be the proportion of the population under treatment. The average causal effect for the whole population can be written as a convex combination of the average causal effects for the treated and control populations:

$$\mathrm{ACE}^{\mathrm{true}} = E\{Y(1)\} - E\{Y(0)\} = f\mathrm{ACE}_1^{\mathrm{true}} + (1-f)\mathrm{ACE}_0^{\mathrm{true}}.$$

Analogously, with a scalar instrumental variable, the adjusted estimator for the whole population can be written as

$$\mathrm{ACE}^{\mathrm{adj}} = \int \mu_1(z)F(\mathrm{d}z) - \int \mu_0(z)F(\mathrm{d}z) = f\mathrm{ACE}_1^{\mathrm{adj}} + (1-f)\mathrm{ACE}_0^{\mathrm{adj}},$$

and with a general instrumental variable,

$$\mathrm{ACE}^{\mathrm{adj}} = \int \nu_1(\pi)F(\mathrm{d}\pi) - \int \nu_0(\pi)F(\mathrm{d}\pi) = f\mathrm{ACE}_1^{\mathrm{adj}} + (1-f)\mathrm{ACE}_0^{\mathrm{adj}}.$$

LEMMA S9. *With a scalar instrumental variable Z, the differences between the adjusted and unadjusted estimators are* 65

$$\mathrm{ACE}_1^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}} = -\frac{\mathrm{cov}\{\Pi(Z),\mu_0(Z)\}}{f(1-f)},$$
$$\mathrm{ACE}_0^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}} = -\frac{\mathrm{cov}\{\Pi(Z),\mu_1(Z)\}}{f(1-f)},$$
$$\mathrm{ACE}^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}} = -\frac{\mathrm{cov}\{\Pi(Z),\mu_0(Z)\}}{1-f} - \frac{\mathrm{cov}\{\Pi(Z),\mu_1(Z)\}}{f}.$$

With a general instrumental variable Z, the above formulas hold if we replace $\Pi(Z)$ by Π and $\mu_a(Z) = E(Y \mid A=a, Z)$ by $\nu_a(\Pi) = E(Y \mid A=a,\Pi)$.

Proof of Lemma S9. The difference $\mathrm{ACE}_1^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}}$ is equal to

$$\begin{aligned}
&\mathrm{ACE}_1^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}}\\
&= E(Y \mid A=0) - \int \mu_0(z)F(\mathrm{d}z \mid A=1)\\
&= \int \mu_0(z)F(\mathrm{d}z \mid A=0) - \int \mu_0(z)F(\mathrm{d}z \mid A=1)\\
&= \frac{\int \mu_0(z)\{1-\Pi(z)\}F(\mathrm{d}z)}{1-f} - \frac{\int \mu_0(z)\Pi(z)F(\mathrm{d}z)}{f}\\
&= \frac{1}{f(1-f)}\Big[E\{\mu_0(Z)(1-\Pi(Z))\}E\{\Pi(Z)\} - E\{\mu_0(Z)\Pi(Z)\}E\{1-\Pi(Z)\}\Big]\\
&= \frac{1}{f(1-f)}\Big[E\{\mu_0(Z)\}E\{\Pi(Z)\} - E\{\mu_0(Z)\Pi(Z)\}\Big]\\
&= -\frac{\mathrm{cov}\{\Pi(Z),\mu_0(Z)\}}{f(1-f)}.
\end{aligned}$$

Similarly, the difference $\mathrm{ACE}_0^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}}$ is equal to

$$\begin{aligned}
\mathrm{ACE}_0^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}} &= \int \mu_1(z)F(\mathrm{d}z \mid A=0) - \int \mu_1(z)F(\mathrm{d}z \mid A=1)\\
&= -\frac{\mathrm{cov}\{\Pi(Z),\mu_1(Z)\}}{f(1-f)}.
\end{aligned}$$

Therefore, the difference $\mathrm{ACE}^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}}$ is equal to 70

$$\begin{aligned}
\mathrm{ACE}^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}} &= f(\mathrm{ACE}_1^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}}) + (1-f)(\mathrm{ACE}_0^{\mathrm{adj}} - \mathrm{ACE}^{\mathrm{unadj}})\\
&= -\frac{\mathrm{cov}\{\Pi(Z),\mu_0(Z)\}}{1-f} - \frac{\mathrm{cov}\{\Pi(Z),\mu_1(Z)\}}{f}.
\end{aligned}$$

Analogously, we can prove the results for general instrumental variables. □

APPENDIX 2. PROOFS OF THEOREMS AND COROLLARIES IN THE MAIN TEXT

Proof of Theorem 1. Because $\Pi(z) = \text{pr}(A = 1 \mid Z = z)$ and $\text{pr}(A = 1 \mid U = u)$ are non-decreasing in z and u, and $E(Y \mid A = a, U = u)$ is non-decreasing in u for both $a = 0$ and 1, the unadjusted estimator, $\text{ACE}^{\text{unadj}}$, is larger than or equal to ACE^{true}, $\text{ACE}_1^{\text{true}}$ and $\text{ACE}_0^{\text{true}}$, according to Lemmas S2–S4.

Because $\Pi(Z)$ is non-decreasing and $\mu_a(Z)$ is non-increasing in Z for both $a = 0$ and 1, their covariance is non-positive according to Lemma S1, i.e., $\text{cov}\{\Pi(Z), \mu_a(Z)\} \leq 0$.

Because the differences between all the adjusted estimators, $\text{ACE}_1^{\text{adj}}$, $\text{ACE}_0^{\text{adj}}$ and ACE^{adj}, and the unadjusted estimator, $\text{ACE}^{\text{unadj}}$, are negative constants multiplied by $\text{cov}\{\Pi(Z), \mu_a(Z)\}$, according to Lemma S9 all of $\text{ACE}_1^{\text{adj}}$, $\text{ACE}_0^{\text{adj}}$, and ACE^{adj} are larger or equal to $\text{ACE}^{\text{unadj}}$. $\square$

Proof of Theorem 2. The independence of Z and U implies that

$$\text{pr}(A = 1 \mid Z = z) = \int \text{pr}(A = 1 \mid Z = z, U = u) F(\mathrm{d}u) = \beta(z) + E\{\gamma(U)\},$$
$$\text{pr}(A = 1 \mid U = u) = \int \text{pr}(A = 1 \mid Z = z, U = u) F(\mathrm{d}z) = E\{\beta(Z)\} + \gamma(u)$$

are non-decreasing in z and u. Therefore, according to Theorem 1 we need only to verify that $E(Y \mid A = a, Z = z)$ in non-increasing in z for both $a = 0$ and 1.

Because $Z \perp\!\!\!\perp U$ and $\text{pr}(A = 1 \mid Z = z, U = u) = \beta(z) + \gamma(u)$ with non-decreasing $\beta(z)$ and $\gamma(u)$, we can apply Lemma S6, and conclude that $\partial F(u \mid A = a, Z = z)/\partial z \geq 0$.

Write the essential infimum and supremum of U given $(A = a, Z = z)$ as $\underline{u}(a, z)$ and $\overline{u}(a)$, with the later depending only on a according to Condition (c) of Theorem 2. Because $Y \perp\!\!\!\perp Z \mid (A, U)$, integration or summation by parts gives

$$\begin{aligned}
& E(Y \mid A = a, Z = z) \\
&= \int E(Y \mid A = a, Z = z, U = u) F(\mathrm{d}u \mid A = a, Z = z) \\
&= \int m_a(u) F(\mathrm{d}u \mid A = a, Z = z) \\
&= m_a(u) F(u \mid A = a, Z = z)\big|_{u=\underline{u}(a,z)}^{u=\overline{u}(a)} - \int \left\{\frac{\partial m_a(u)}{\partial u}\right\} F(u \mid A = a, Z = z)\, \mathrm{d}u \\
&= m_a\{\overline{u}(a)\} - \int \left\{\frac{\partial m_a(u)}{\partial u}\right\} F(u \mid A = a, Z = z)\, \mathrm{d}u.
\end{aligned}$$

Therefore, its derivative with respect to z,

$$\begin{aligned}
\frac{\partial E(Y \mid A = a, z)}{\partial z} &= -\frac{\partial}{\partial z} \int \left\{\frac{\partial m_a(u)}{\partial u}\right\} F(u \mid A = a, Z = z)\, \mathrm{d}u \\
&= -\int \left\{\frac{\partial m_a(u)}{\partial u}\right\} \left\{\frac{\partial F(u \mid A = a, Z = z)}{\partial z}\right\} \mathrm{d}u,
\end{aligned}$$

is smaller than or equal to zero, because $\partial m_a(u)/\partial u \geq 0$ for both $a = 0$ and 1 and for all u. $\square$

Proof of Corollary 1. According to Theorem 1 we need only to verify that $\mu_a(z) = E(Y \mid A = a, Z = z)$ is non-increasing in z for both $a = 0$ and 1. Following Lemma S5, for binary and independent Z and U, monotonicity and no additive interaction imply (S1), which, according to

Bayes' Theorem, is equivalent to

$$\frac{\text{pr}(A=1 \mid Z=1, U=1)\text{pr}(A=1 \mid Z=0, U=0)}{\text{pr}(A=1 \mid Z=1, U=0)\text{pr}(A=1 \mid Z=0, U=1)} = \text{OR}_{ZU|A=1} \leq 1, \tag{S3}$$

$$\frac{\text{pr}(A=0 \mid Z=1, U=1)\text{pr}(A=0 \mid Z=0, U=0)}{\text{pr}(A=0 \mid Z=1, U=0)\text{pr}(A=0 \mid Z=0, U=1)} = \text{OR}_{ZU|A=0} \leq 1. \tag{S4}$$

The above inequalities (S3) and (S4) state that Z and U have negative association given each level of A, and therefore pr$(U=1 \mid A=a, Z=z)$ is non-increasing in z for both $a=1$ and 0.

Because $m_a(1) \geq m_a(0)$ and

$$\begin{aligned}
\mu_a(z) &= E(Y \mid A=a, Z=z) \\
&= \sum_{u=0,1} E(Y \mid A=a, Z=z, U=u)\text{pr}(U=u \mid A=a, Z=z) \\
&= m_a(1)\text{pr}(U=1 \mid A=a, Z=z) + m_a(0)\{1-\text{pr}(U=1 \mid A=a, Z=z)\} \\
&= \{m_a(1)-m_a(0)\}\text{pr}(U=1 \mid A=a, Z=z) + m_a(0),
\end{aligned}$$

we know that $\mu_a(z)$ is non-decreasing in pr$(U=1 \mid A=a, Z=z)$. Therefore, $\mu_a(z)$ is non-increasing in z for both $a=1$ and 0. □

Proof of Theorem 3. Because of the independence of Z and U, we have pr$(A=1 \mid Z=z) = \beta(z)E\{\gamma(U)\}$ and pr$(A=1 \mid U=u) = E\{\beta(Z)\}\gamma(u)$ are non-decreasing in z and u. According to Lemma S8, the multiplicative model of A also implies that for both $a=1$ and 0 and for all z and u, $\partial F(u \mid A=a, Z=z)/\partial z \geq 0$. Following exactly the same steps of the proof of Theorem 2, we can prove Theorem 3. □

Proof of Corollary 2. For binary and independent Z and U, monotonicity, no multiplicative interaction, and Lemma S7 imply

$$\frac{p_{11}p_{00}}{p_{10}p_{01}} = 1 \leq 1, \quad \frac{(1-p_{11})(1-p_{00})}{(1-p_{10})(1-p_{01})} \leq 1. \tag{S5}$$

With the above results in (S5), the rest of the proof is the same as the proof of Corollary 1. □

Proof of Theorem 4. First, we consider the treatment effect on the population under treatment. Taking $U = Y(0)$ in Lemma S3, we have $\text{ACE}^{\text{unadj}} \geq \text{ACE}_1^{\text{true}}$, because $A \perp\!\!\!\perp Y(0) \mid Y(0)$, pr$\{A=1 \mid Y(0)\}$ is non-decreasing in $Y(0)$, and $E\{Y \mid A=0, Y(0)\} = Y(0)$ is non-decreasing in $Y(0)$. The condition $\text{cov}\{\Pi, E(Y \mid A=0, \Pi)\} \leq 0$ implies that $\text{ACE}_1^{\text{adj}} \geq \text{ACE}^{\text{unadj}}$ according to Lemma S9. Therefore, $\text{ACE}_1^{\text{adj}} \geq \text{ACE}^{\text{unadj}} \geq \text{ACE}_1^{\text{true}}$.

Second, we take $U = Y(1)$ in Lemma S4, and by a similar argument as above we have $\text{ACE}_0^{\text{adj}} \geq \text{ACE}^{\text{unadj}} \geq \text{ACE}_0^{\text{true}}$.

The conclusion holds because $\text{ACE}^{\text{true}} = f\text{ACE}_1^{\text{true}} + (1-f)\text{ACE}_0^{\text{true}}$ and $\text{ACE}^{\text{adj}} = f\text{ACE}_1^{\text{adj}} + (1-f)\text{ACE}_0^{\text{adj}}$. □

Proof of Theorem 5. Under the additive model of A given Π and $U = \{Y(1), Y(0)\}$, we have the following results. First, pr$(A=1 \mid \Pi) = \Pi$ is increasing in Π. Second, $\Pi \perp\!\!\!\perp \{Y(1), Y(0)\}$

implies

$$\begin{aligned}\mathrm{pr}\{A=1\mid \Pi, Y(1)=y_1\} &= \int \mathrm{pr}(A=1\mid \Pi, U) F(\mathrm{d}y_0\mid y_1)\\ &= \int \{\Pi+\delta(y_1)+\eta(y_0)\} F(\mathrm{d}y_0\mid y_1)\\ &= \Pi+\delta(y_1)+\int \eta(y_0)F(\mathrm{d}y_0\mid y_1) \equiv \Pi+\widetilde{\delta}(y_1).\end{aligned}$$

Denote the infimum and supremum of $Y(0)$ given $Y(1)=y_1$ by $\underline{y}_0(y_1)$ and $\overline{y}_0$, with the later not depending on y_1 according to Condition (c) of Theorem 5. Applying integration or summation by parts, we have

$$\begin{aligned}\widetilde{\delta}(y_1) &= \delta(y_1)+\eta(y_0)F(y_0\mid y_1)\big|_{y_0=\underline{y}_0(y_1)}^{y_0=\overline{y}_0} - \int \left\{\frac{\mathrm{d}\eta(y_0)}{\mathrm{d}y_0}\right\} F(y_0\mid y_1)\,\mathrm{d}y_0\\ &= \delta(y_1)+\eta(\overline{y}_0) - \int \left\{\frac{\mathrm{d}\eta(y_0)}{\mathrm{d}y_0}\right\} F(y_0\mid y_1)\,\mathrm{d}y_0.\end{aligned}$$

The function $\widetilde{\delta}(y_1)$ is non-decreasing in y_1, because

$$\frac{\mathrm{d}\,\widetilde{\delta}(y_1)}{\mathrm{d}y_1} = \frac{\mathrm{d}\delta(y_1)}{\mathrm{d}y_1} - \int \left\{\frac{\mathrm{d}\eta(y_0)}{\mathrm{d}y_0}\right\}\left\{\frac{\partial F(y_0\mid y_1)}{\partial y_1}\right\}\mathrm{d}y_0 \geq 0.$$

Third, following the same reasoning as the second argument, we have $\mathrm{pr}\{A=1\mid \Pi, Y(1)=y_0\}=\Pi+\widetilde{\eta}(y_0)$, with $\widetilde{\eta}(y_0)$ being a non-decreasing function of y_0. Fourth, $\Pi \perp\!\!\!\perp Y(1)$ implies $\mathrm{pr}\{A=1\mid Y(1)=y_1\}=f+\widetilde{\delta}(y_1)$, which is non-decreasing in y_1. Fifth, $\Pi \perp\!\!\!\perp Y(0)$ implies $\mathrm{pr}\{A=1\mid Y(0)=y_0\}=f+\widetilde{\eta}(y_0)$, which is non-decreasing in y_0.

According the fourth and fifth arguments above, Condition (a) in Theorem 4 holds. Therefore, we need only to verify Condition (b) in Theorem 4 to complete the proof.

We have shown that $\mathrm{pr}\{A=1\mid \Pi, Y(1)\}=\Pi+\widetilde{\delta}\{Y(1)\}$, which is additive and non-decreasing in Π and $Y(1)$. According to Lemma S6, we know that

$$\frac{\partial \mathrm{pr}\{Y(1)\leq y_1\mid A=1, \Pi=\pi\}}{\partial \pi} \geq 0 \tag{S6}$$

for all y_1 and π. We have also shown that $\mathrm{pr}\{A=1\mid \Pi, Y(0)\}=\Pi+\widetilde{\eta}\{Y(0)\}$, which is additive and non-decreasing in Π and $Y(0)$. Again according to Lemma S6, we know that

$$\frac{\partial \mathrm{pr}\{Y(0)\leq y_0\mid A=0, \Pi=\pi\}}{\partial \pi} \geq 0 \tag{S7}$$

for all y_0 and π. According to Xie et al. (2008), the above negative distributional associations in (S6) and (S7) imply the negative associations in expectation between $Y(0)$ and Π given A, as required by condition (b) of Theorem 4. □

Proof of Corollary 3. As shown in the proof of Theorem 5, the conclusion follows immediately from the five ingredients. We will show that they hold even if there is non-negative interaction between binary $Y(1)$ and $Y(0)$. The following proof is in parallel with the proof of Theorem 5.

First, $\text{pr}(A = 1 \mid \Pi) = \Pi$ is increasing in Π. Second,

$$\begin{aligned}
&\text{pr}\{A = 1 \mid \Pi, Y(1) = y_1\} \\
&= E\left[\text{pr}\{A = 1 \mid \Pi, Y(1) = y_1, Y(0)\} \mid \Pi, Y(1) = y_1\right] \\
&= E\left\{\alpha + \Pi + \delta y_1 + \eta Y(0) + \theta y_1 Y(0) \mid \Pi, Y(1) = y_1\right\} \\
&= \alpha + \Pi + \delta y_1 + \eta \text{pr}\{Y(0) = 1 \mid Y(1) = y_1\} + \theta y_1 \text{pr}\{Y(0) = 1 \mid Y(1) = y_1\} \quad \text{(S8)} \\
&\equiv \Pi + \widetilde{\delta}[y_1 - E\{Y(1)\}]. \quad \text{(S9)}
\end{aligned}$$

The last equation in (S9) follows from the fact that $Y(1)$ is binary and the functional form must be linear in y_1, where the coefficient is

$$\begin{aligned}
\widetilde{\delta} &= \text{pr}\{A = 1 \mid \Pi, Y(1) = 1\} - \text{pr}\{A = 1 \mid \Pi, Y(1) = 0\} \\
&= \delta + \eta[\text{pr}\{Y(0) = 1 \mid Y(1) = 1\} - \text{pr}\{Y(0) = 1 \mid Y(1) = 0\}] + \theta \text{pr}\{Y(0) = 1 \mid Y(1) = 1\} \quad \text{(S10)} \\
&\geq \eta[\text{pr}\{Y(0) = 1 \mid Y(1) = 1\} - \text{pr}\{Y(0) = 1 \mid Y(1) = 0\}], \quad \text{(S11)}
\end{aligned}$$

where (S10) follows from (S8), and (S11) follows from $\delta \geq 0$ and $\theta \geq 0$. Because $\text{OR}_Y \geq 1$, the potential outcomes have non-negative association, implying that their risk difference $\text{RD}_Y = \text{pr}\{Y(0) = 1 \mid Y(1) = 1\} - \text{pr}\{Y(0) = 1 \mid Y(1) = 0\} \geq 0$. Therefore, $\widetilde{\delta} \geq 0$, and $\text{pr}\{A = 1 \mid \Pi, Y(1)\}$ is additive and non-decreasing in Π and $Y(1)$.

Third, similar to the second argument, we have $\text{pr}\{A = 1 \mid \Pi, Y(0) = y_0\} = \Pi + \widetilde{\eta}[y_0 - E\{Y(0)\}]$ with $\widetilde{\eta} \geq 0$. Therefore, $\text{pr}\{A = 1 \mid \Pi, Y(0)\}$ is additive and non-decreasing in Π and $Y(0)$. Fourth, $\Pi \perp\!\!\!\perp Y(1)$ implies that $\text{pr}\{A = 1 \mid Y(1)\} = f + \widetilde{\delta} Y(1)$ is increasing in $Y(1)$. Fifth, $\Pi \perp\!\!\!\perp Y(0)$ implies that $\text{pr}\{A = 1 \mid Y(0)\} = f + \widetilde{\eta} Y(0)$ is increasing in $Y(0)$.

With these five ingredients, the rest of the proof is exactly the same as the proof of Theorem 5. □

Proof of Theorem 6. First, $\text{pr}(A = 1 \mid \Pi) = \Pi$ is non-decreasing in Π. Second,

$$\text{pr}\{A = 1 \mid \Pi, Y(1) = y_1\} = \Pi \delta(y_1) \int \delta(y_0) F(\mathrm{d}y_0 \mid y_1) \equiv \Pi \widetilde{\delta}(y_1)$$

is multiplicative and non-decreasing in Π and y_1, following the same argument as the proof of Theorem 5. Third, $\text{pr}\{A = 1 \mid \Pi, Y(0) = y_0\} = \Pi \widetilde{\eta}(y_0)$ is multiplicative and non-decreasing in Π and y_0. Fourth, $\text{pr}\{A = 1 \mid Y(1) = y_1\} = f \widetilde{\delta}(y_1)$ is non-decreasing in y_1. Fifth, $\text{pr}\{A = 1 \mid Y(0) = y_0\} = f \widetilde{\eta}(y_0)$ is non-decreasing in y_0.

The multiplicative models and Lemma S8 imply that for all π, y_1 and y_0,

$$\frac{\partial \text{pr}\{Y(1) \leq y_1 \mid A = 1, \Pi = \pi\}}{\partial \pi} = 0 \leq 0, \quad \frac{\partial \text{pr}\{Y(0) \leq y_0 \mid A = 0, \Pi = \pi\}}{\partial \pi} \geq 0. \quad \text{(S12)}$$

The rest part is the same as the proof of Theorem 5. □

Proof of Corollary 4. First, $\text{pr}(A = 1 \mid \Pi) = \Pi$ is non-decreasing in Π. Second,

$$\text{pr}\{A = 1 \mid \Pi, Y(1) = y_1\} = \alpha \Pi \delta^{y_1} E\{\eta^{Y(0)} \theta^{y_1 Y(0)} \mid Y(1) = y_1\} \equiv \alpha \Pi \widetilde{\delta}^{Y(1)},$$

where the functional form must be multiplicative because of binary $Y(0)$, and the parameter $\widetilde{\delta}$ is

$$\begin{aligned}\widetilde{\delta} &= \frac{\text{pr}\{A=1 \mid \Pi, Y(1)=1\}}{\text{pr}\{A=1 \mid \Pi, Y(1)=0\}} \\ &= \delta \times \frac{E\{\eta^{Y(0)}\theta^{Y(0)} \mid Y(1)=1\}}{E\{\eta^{Y(0)} \mid Y(1)=0\}} \\ &= \delta \times \frac{\eta\theta \text{pr}\{Y(0)=1 \mid Y(1)=1\} + \text{pr}\{Y(0)=0 \mid Y(1)=1\}}{\eta \text{pr}\{Y(0)=1 \mid Y(1)=0\} + \text{pr}\{Y(0)=0 \mid Y(1)=0\}} \\ &= \delta \times \frac{(\eta\theta-1)\text{pr}\{Y(0)=1 \mid Y(1)=1\}+1}{(\eta-1)\text{pr}\{Y(0)=1 \mid Y(1)=0\}+1}.\end{aligned}$$

Because $\text{OR}_Y \geq 1$, we have $\text{pr}\{Y(0)=1 \mid Y(1)=1\} \geq \text{pr}\{Y(0)=1 \mid Y(1)=0\}$, which implies that $\widetilde{\delta} \geq 1$. Therefore, $\text{pr}\{A=1 \mid \Pi, Y(1)\}$ is multiplicative and non-decreasing in Π and $Y(1)$. Third, we can similarly show that $\text{pr}\{A=1 \mid \Pi, Y(0)\}$ is multiplicative and non-decreasing in Π and $Y(0)$. Fourth, $\text{pr}\{A=1 \mid Y(1)=y_1\} = \alpha f \widetilde{\delta}^{y_1}$ is non-decreasing in y_1. Fifth, $\text{pr}\{A=1 \mid Y(0)=y_0\} = \alpha f \widetilde{\eta}^{y_0}$ is non-decreasing in y_0.

The rest part is the same as the proof of Theorem 6. □

Proof of Theorem 7.. In Figure 4, Z and U are two independent confounders for the relationship between A and Y. Because $\text{pr}(A=1 \mid Z=z, U=u)$ and $E(Y \mid A=a, Z=z, U=u)$ are non-decreasing in z and u for both $a=0$ and 1, Lemmas S2–S4 imply that the unadjusted estimator, $\text{ACE}^{\text{unadj}}$, is larger than or equal to ACE^{true}, $\text{ACE}_1^{\text{true}}$ and $\text{ACE}_0^{\text{true}}$.

The independence between Z and U implies $\text{pr}(A=1 \mid Z=z) = \int \text{pr}(A=1 \mid Z=z, U=u)F(\text{d}u)$, and the monotonicity of $\text{pr}(A=1 \mid Z=z, U=u)$ in z implies that $\text{pr}(A=1 \mid Z=z)$ is non-decreasing in z. The rest of the proof is identical to the proof of Theorem 1. □

APPENDIX 3. EXTENSIONS TO OTHER CAUSAL MEASURES

Appendix 3·1. Distributional Causal Effects

Sometimes we are also interested in estimating the distributional causal effects (Ju & Geng, 2010) for the treatment, control and whole populations:

$$\begin{aligned}\text{DCE}_1^{\text{true}}(y) &= \text{pr}\{Y(1)>y \mid A=1\} - \text{pr}\{Y(0)>y \mid A=1\}, \\ \text{DCE}_0^{\text{true}}(y) &= \text{pr}\{Y(1)>y \mid A=0\} - \text{pr}\{Y(0)>y \mid A=0\}, \\ \text{DCE}^{\text{true}}(y) &= \text{pr}\{Y(1)>y\} - \text{pr}\{Y(0)>y\}.\end{aligned}$$

The unadjusted estimator is

$$\text{DCE}^{\text{unadj}}(y) = \text{pr}(Y>y \mid A=1) - \text{pr}(Y>y \mid A=0).$$

The adjusted estimators for the treatment, control and whole populations are

$$\begin{aligned}\text{DCE}_1^{\text{adj}}(y) &= \text{pr}(Y>y \mid A=1) - \int \text{pr}(Y>y \mid A=0, z)F(\text{d}z \mid A=1), \\ \text{DCE}_0^{\text{adj}}(y) &= \int \text{pr}(Y>y \mid A=1, z)F(\text{d}z \mid A=0) - \text{pr}(Y>y \mid A=0), \\ \text{DCE}^{\text{adj}}(y) &= \int \text{pr}(Y>y \mid A=1, z)F(\text{d}z) - \int \text{pr}(Y>y \mid A=0, z)F(\text{d}z).\end{aligned}$$

If the outcome is binary, then the distributional causal effects at $y < 1$ are the average causal effects, and zero at $y \geq 1$. All results about distributional causal effects reduce to average causal effects for binary outcome. For a general outcome, the distributional causal effects are the average causal effects on the dichotomized outcome $I_y = I(Y > y)$. Therefore, if we replace the outcome Y by I_y in Theorems 1–3, the results about Z-bias hold for distributional effects. For instance, the condition that $\text{pr}(Y > y \mid A = a, U = u)$ is non-decreasing in u for all a is the same as requiring a non-negative sign on the arrow $U \to Y$, according to the theory of signed directed acyclic graphs (VanderWeele & Robins, 2010). The following theorem states the results analogous to Theorems 4–6.

COROLLARY S1. *In the causal diagram of Figure 2, if for all y and for both $a = 1$ and 0,*

(a) $\text{pr}\{Y(a) > y \mid A = 1\} \geq \text{pr}\{Y(a) > y \mid A = 0\}$*;*
(b) $\text{cov}\{\Pi, \text{pr}(Y > y \mid A = a, \Pi)\} \leq 0$*;*

then

$$\begin{pmatrix} \text{DCE}_1^{\text{adj}}(y) \\ \text{DCE}_0^{\text{adj}}(y) \\ \text{DCE}^{\text{adj}}(y) \end{pmatrix} \geq \begin{pmatrix} \text{DCE}^{\text{unadj}}(y) \\ \text{DCE}^{\text{unadj}}(y) \\ \text{DCE}^{\text{unadj}}(y) \end{pmatrix} \geq \begin{pmatrix} \text{DCE}_1^{\text{true}}(y) \\ \text{DCE}_0^{\text{true}}(y) \\ \text{DCE}^{\text{true}}(y) \end{pmatrix}. \tag{S13}$$

Under the conditions of Theorems 5 *and* 6*,* (S13) *holds.*

*Proof of Corollary S*1. Condition (a) of Corollary S1 is equivalent to $\text{pr}\{A = 1 \mid I_y(a) = 1\} \geq \text{pr}\{A = 1 \mid I_y(a) = 0\}$, and Condition (b) of Corollary S1 is equivalent to $\text{cov}\{\Pi, E(I_y \mid A = a, \Pi)\} \leq 0$. Therefore, the conclusion follows from Theorem 4.

According to the proofs of Theorems 5 and 6, we have

$$\begin{aligned} &\text{pr}\{A = 1 \mid I_y(a) = 1\} = \text{pr}\{A = 1 \mid Y(a) > y\} \geq \text{pr}\{A = 1 \mid Y(a) = y\} \\ &\geq \text{pr}\{A = 1 \mid Y(a) \leq y\} = \text{pr}\{A = 1 \mid I_y(a) = 0\}, \end{aligned}$$

because of monotonicity of $\text{pr}\{A = 1 \mid Y(a)\}$ in $Y(a)$. Therefore, Condition (a) of Theorem S1 holds. Under the conditions of Theorems 5 and 6, we have also shown in (S6)–(S12) that for all a, y and π, $\partial \text{pr}(Y \leq y \mid A = a, \Pi = \pi)/\partial \pi \geq 0$, which implies that $E(I_y \mid A = a, \Pi = \pi)$ is non-increasing in π. Therefore, Condition (b) of Theorem S1 holds. The proof is complete. □

Appendix 3·2. Ratio Measures

In many applications with binary or positive outcomes, we are also interested in assessing causal effects on the ratio scale for the treatment, control and whole populations, defined as

$$\text{RR}_1^{\text{true}} = \frac{E\{Y(1) \mid A = 1\}}{E\{Y(0) \mid A = 1\}}, \quad \text{RR}_0^{\text{true}} = \frac{E\{Y(1) \mid A = 0\}}{E\{Y(0) \mid A = 0\}}, \quad \text{RR}^{\text{true}} = \frac{E\{Y(1)\}}{E\{Y(0)\}}.$$

The unadjusted estimator on the ratio scale is

$$\text{RR}^{\text{unadj}} = \frac{E(Y \mid A = 1)}{E(Y \mid A = 0)}.$$

The adjusted estimators on the ratio scale for the treatment, control and whole populations are

$$\mathrm{RR}_1^{\mathrm{adj}} = \frac{E(Y \mid A=1)}{\int E\{Y \mid A=0, Z=z\}F(\mathrm{d}z \mid A=1)},$$
$$\mathrm{RR}_0^{\mathrm{adj}} = \frac{\int E\{Y \mid A=1, Z=z\}F(\mathrm{d}z \mid A=0)}{E(Y \mid A=0)},$$
$$\mathrm{RR}^{\mathrm{adj}} = \frac{\int E\{Y \mid A=1, Z=z\}F(\mathrm{d}z)}{\int E\{Y \mid A=0, Z=z\}F(\mathrm{d}z)}.$$

With a general instrumental variable Z, we can replace Z by Π in the definitions of the adjusted estimators.

COROLLARY S2. *All the theorems and corollaries in* §§3 *and* 4 *hold on the ratio scale, i.e., under their conditions,*

$$\begin{pmatrix} \mathrm{RR}_1^{\mathrm{adj}} \\ \mathrm{RR}_0^{\mathrm{adj}} \\ \mathrm{RR}^{\mathrm{adj}} \end{pmatrix} \geq \begin{pmatrix} \mathrm{RR}^{\mathrm{unadj}} \\ \mathrm{RR}^{\mathrm{unadj}} \\ \mathrm{RR}^{\mathrm{unadj}} \end{pmatrix} \geq \begin{pmatrix} \mathrm{RR}_1^{\mathrm{true}} \\ \mathrm{RR}_0^{\mathrm{true}} \\ \mathrm{RR}^{\mathrm{true}} \end{pmatrix}.$$

*Proof of Corollary S*2. First, $\mathrm{RR}^{\mathrm{true}}$ is a convex combination of $\mathrm{RR}_1^{\mathrm{true}}$ and $\mathrm{RR}_0^{\mathrm{true}}$, and $\mathrm{RR}^{\mathrm{adj}}$ is a convex combination of $\mathrm{RR}_1^{\mathrm{adj}}$ and $\mathrm{RR}_0^{\mathrm{adj}}$, which are formally stated in Ding & VanderWeele (2016, eAppendix). Then the conclusion follows from the proofs of the theorems above. □

Appendix 3·3. Average Over Observed Covariates

In practice, we need to adjust for the observed covariates X that are confounders affecting both the treatment and outcome. The discussion in previous sections is conditional on or within strata of observed covariates X, and the causal effects and their estimators are given X. For example,

$$\begin{aligned} \mathrm{ACE}^{\mathrm{true}}(x) &= E\{Y(1) \mid X=x\} - E\{Y(0) \mid X=x\}, \\ \mathrm{ACE}^{\mathrm{unadj}}(x) &= E(Y \mid A=1, X=x) - E(Y \mid A=0, X=x), \\ \mathrm{ACE}^{\mathrm{adj}}(x) &= \int E(Y \mid A=1, Z=z, X=x)F(\mathrm{d}z \mid X=x) \\ &\quad - \int E(Y \mid A=0, Z=z, X=x)F(\mathrm{d}z \mid X=x), \end{aligned}$$

and other conditional quantities can be analogously defined. If the conditions in the theorems and corollaries in §§3 and 4 hold within each level of X, then the conclusions in (1) and (S13) hold not only within each level of X but also averaged over X. For example, for the average causal effects, we have

$$\begin{aligned} \begin{pmatrix} \int \mathrm{ACE}_1^{\mathrm{adj}}(x)F(\mathrm{d}x \mid A=1) \\ \int \mathrm{ACE}_0^{\mathrm{adj}}(x)F(\mathrm{d}x \mid A=0) \\ \int \mathrm{ACE}^{\mathrm{adj}}(x)F(\mathrm{d}x) \end{pmatrix} &\geq \begin{pmatrix} \int \mathrm{ACE}^{\mathrm{unadj}}(x)F(\mathrm{d}x \mid A=1) \\ \int \mathrm{ACE}^{\mathrm{unadj}}(x)F(\mathrm{d}x \mid A=0) \\ \int \mathrm{ACE}^{\mathrm{unadj}}(x)F(\mathrm{d}x) \end{pmatrix} \\ &\geq \begin{pmatrix} \int \mathrm{ACE}_1^{\mathrm{true}}(x)F(\mathrm{d}x \mid A=1) \\ \int \mathrm{ACE}_0^{\mathrm{true}}(x)F(\mathrm{d}x \mid A=0) \\ \int \mathrm{ACE}^{\mathrm{true}}(x)F(\mathrm{d}x) \end{pmatrix}. \end{aligned}$$

References

Chiba, Y. (2009). The sign of the unmeasured confounding bias under various standard populations. *Biometrical Journal* **51**, 670–676.

Ding, P. & VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology* **27**, 368–377.

Esary, J. D., Proschan, F. & Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics* **38**, 1466–1474.

Ju, C. & Geng, Z. (2010). Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 129–142.

Piegorsch, W. W., Weinberg, C. R. & Taylor, J. A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* **13**, 153–162.

Rothman, K. J., Greenland, S. & Lash, T. L. (2008). *Modern Epidemiology (3rd Edition)*. Lippincott Williams & Wilkins.

VanderWeele, T. J. (2008). The sign of the bias of unmeasured confounding. *Biometrics* **64**, 702–706.

VanderWeele, T. J. & Robins, J. M. (2010). Signed directed acyclic graphs for causal inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 111–127.

Xie, X., Ma, Z. & Geng, Z. (2008). Some association measures and their collapsibility. *Statistica Sinica* **18**, 1165–1183.

Yang, Q., Khoury, M. J., Sun, F. & Flanders, W. D. (1999). Case-only design to measure gene-gene interaction. *Epidemiology* **10**, 167–170.