Environ Health Perspect

DOI: 10.1289/EHP2450

Supplemental Material

Bias Amplification in Epidemiologic Analysis of Exposure to Mixtures

Marc G. Weisskopf, Ryan M. Seals, and Thomas F. Webster

Table of Contents

Results are organized by DAG (refer to figures). DAG-specific assumptions are listed at the beginning of each section. Results are given for expected values of the regression coefficients b_i (for exposure X_i) under the usual assumptions for linear regression. Some results are elementary.

For simplicity, assume variables are standardized (centered and variance=1). For DAG 1, we provide results relaxing this assumption. A similar procedure can be followed to generalize results for other cases.

The results for inequalities can depend on the signs of coefficients as multiplying or dividing by negative numbers reverses the direction of inequalities.

DAG in Figure 1

$$Y = \beta_0 + c_1 X_1 + \varepsilon$$

General case allowing for exposure variables with variance not equal to one (i.e. not standardized)

Single exposure analyses:

$$b_1 = \frac{\text{cov}[X_1, Y]}{\text{var}[X_1]} = \frac{\text{cov}[X_1, \beta_0 + c_1 X_1 + \varepsilon]}{\text{var}[X_1]} = \frac{c_1 \text{var}[X_1]}{\text{var}[X_1]} = c_1$$

$$b_2 = \frac{\text{cov}[X_2, Y]}{\text{var}[X_2]} = \frac{\text{cov}[X_2, \beta_0 + c_1 X_1 + \varepsilon]}{\text{var}[X_2]} = \frac{c_1 \text{cov}[X_1, X_2]}{\text{var}[X_2]}$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{var}[X_2]\text{cov}[X_1, Y] - \text{cov}[X_1, X_2]\text{cov}[X_2, Y]}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1, X_2]} = \frac{\text{var}[X_2]\left(c_1 \text{var}[X_1]\right) - \text{cov}[X_1, X_2]\left(c_1 \text{cov}[X_1, X_2]\right)}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1, X_2]} = c_1$$

$$b_2 = \frac{\text{var}[X_1]\text{cov}[X_2, Y] - \text{cov}[X_1, X_2]\text{cov}[X_1, Y]}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1, X_2]} = \frac{\text{var}[X_1]\left(c_1 \text{cov}[X_1, X_2]\right) - \text{cov}[X_1, X_2]\left(c_1 \text{var}[X_1]\right)}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1, X_2]} = 0$$

When exposure variances equal one (i.e. standardized), the above equations simplify with $\text{cov}[X_1, X_2] = r_{12}$

Single exposure analyses:

$$b_1 = \frac{\text{cov}[X_1, Y]}{\text{var}[X_1]} = \text{cov}[X_1, \beta_0 + c_1 X_1 + \varepsilon] = c_1$$

$$b_2 = \frac{\text{cov}[X_2, Y]}{\text{var}[X_2]} = \text{cov}[X_2, \beta_0 + c_1 X_1 + \varepsilon] = c_1 \text{cov}[X_1, X_2] = c_1 r_{12}$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{var}[X_2]\text{cov}[X_1,Y] - \text{cov}[X_1,X_2]\text{cov}[X_2,Y]}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1,X_2]} = \frac{c_1 - r_{12}\left(c_1 r_{12}\right)}{1-r_{12}^2} = c_1$$

$$b_2 = \frac{\text{var}[X_1]\text{cov}[X_2,Y] - \text{cov}[X_1,X_2]\text{cov}[X_1,Y]}{\text{var}[X_1]\text{var}[X_2] - \text{cov}^2[X_1,X_2]} = \frac{c_1 r_{12} - r_{12}\left(c_1\right)}{1-r_{12}^2} = 0$$

DAG in Figure 2a

$$Y = \beta_0 + c_1 X_1 + c_3 U' + \varepsilon$$

$$\text{cov}[X_1,U'] = c_2, \quad \text{cov}[X_2,U'] = 0 \ \text{ since } \mathrm{X}_1 \text{ is a collider on } \mathrm{X}_2 \leftarrow \mathrm{U} \rightarrow \mathrm{X}_1 \leftarrow \mathrm{U}'$$

Single exposure analyses:

$$b_1 = \text{cov}[X_1,Y] = \text{cov}[X_1, \beta_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 + c_3\,\text{cov}[X_1,U'] = c_1 + c_3 c_2$$

$$b_2 = \text{cov}[X_2,Y] = \text{cov}[X_2, \beta_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1\,\text{cov}[X_1,X_2] + c_3\,\text{cov}[X_2,U'] = c_1 r_{12}$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{cov}[X_1,Y] - r_{12}\,\text{cov}[X_2,Y]}{1-r_{12}^2} = \frac{\left(c_1 + c_3 c_2\right) - r_{12}\left(c_1 r_{12}\right)}{1-r_{12}^2} = c_1 + \frac{c_3 c_2}{1-r_{12}^2}$$

$$b_2 = \frac{\text{cov}[X_2,Y] - r_{12}\,\text{cov}[X_1,Y]}{1-r_{12}^2} = \frac{\left(c_1 r_{12}\right) - r_{12}\left(c_1 + c_3 c_2\right)}{1-r_{12}^2} = -\frac{r_{12} c_3 c_2}{1-r_{12}^2}$$

The sign of b_2 will flip when adjusting for X_1 if c_1 and $c_2 c_3$ have the same sign.
The absolute value of the adjusted b_2 exceeds the crude estimate when:

$$\left| -\frac{r_{12} c_2 c_3}{1-r_{12}^2} \right| = \frac{r_{12} c_2 c_3}{1-r_{12}^2} > r_{12} c_1 \leftrightarrow \frac{c_2 c_3}{1-r_{12}^2} > c_1 \ \text{ provided } r_{12}>0,\ c_1>0,\ c_2 c_3>0$$

Other cases can be derived when r_{12}, c_1, or $c_2 c_3<0$, recalling that inequalities reverse direction when multiplying or dividing by negative numbers.

DAG in Figure 2b

The derivation is identical to 2a with Z replacing X_2, c_4 replacing r_{12}.

DAG in Figure 3a

This is a special case of 2a with $c_1=0$.

DAG in Figure 3b

$$Y = \beta_0 + c_1 X_1 + c_4 X_2 + c_3 U' + \varepsilon$$
$$\mathrm{cov}[X_1, U'] = c_2, \quad \mathrm{cov}[X_2, U'] = 0$$

Single exposure analyses:

$$b_1 = \mathrm{cov}[X_1, Y] = \mathrm{cov}[X_1, \beta_0 + c_1 X_1 + c_4 X_2 + c_3 U' + \varepsilon] = c_1 + c_4 r_{12} + c_3\,\mathrm{cov}[X_1, U'] = c_1 + c_4 r_{12} + c_3 c_2$$
$$b_2 = \mathrm{cov}[X_2, Y] = \mathrm{cov}[X_2, \beta_0 + c_1 X_1 + c_4 X_2 + c_3 U' + \varepsilon] = c_1 r_{12} + c_4 + c_3\,\mathrm{cov}[X_2, U'] = c_1 r_{12} + c_4$$

Mutually adjusted analysis:

$$b_1 = \frac{\mathrm{cov}[X_1, Y] - r_{12}\,\mathrm{cov}[X_2, Y]}{1 - r_{12}^2} = \frac{\left(c_1 + c_4 r_{12} + c_3 c_2\right) - r_{12}\left(c_1 r_{12} + c_4\right)}{1 - r_{12}^2} = c_1 + \frac{c_3 c_2}{1 - r_{12}^2}$$
$$b_2 = \frac{\mathrm{cov}[X_2, Y] - r_{12}\,\mathrm{cov}[X_1, Y]}{1 - r_{12}^2} = \frac{\left(c_1 r_{12} + c_4\right) - r_{12}\left(c_1 + c_4 r_{12} + c_3 c_2\right)}{1 - r_{12}^2} = c_4 - \frac{c_2 c_3 r_{12}}{1 - r_{12}^2}$$

b_1: the bias of the crude estimate exceeds bias of the adjusted estimate when:

$$r_{12} c_4 + c_2 c_3 > \frac{c_2 c_3}{1 - r_{12}^2} \leftrightarrow r_{12} c_4 > c_2 c_3 \left(\frac{r_{12}^2}{1 - r_{12}^2}\right) \leftrightarrow c_4 > c_2 c_3 \left(\frac{r_{12}}{1 - r_{12}^2}\right) \text{ if } r_{12}, c_4, c_2 c_3 > 0$$

The same condition implies that the adjusted b_2 will be positive but biased toward zero. Other cases can be derived when $r_{12}<0$, $c_4<0$ or $c_2 c_3<0$.

DAG in Figure 4a

$$Y = \beta_0 + c_1 X_1 + c_3 U' + \varepsilon$$
$$\mathrm{cov}[X_1, X_2] = r_{12}, \ \mathrm{cov}[X_1, X_3] = r_{13}, \ \mathrm{cov}[X_2, X_3] = r_{23}$$
$$\mathrm{cov}[X_1, U'] = c_2$$
$$\mathrm{cov}[X_2, U'] = \mathrm{cov}[X_3, U'] = 0$$

Single exposure analyses:

$$b_1 = \mathrm{cov}[X_1, Y] = \mathrm{cov}[X_1, \beta_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 + c_3\,\mathrm{cov}[X_1, U'] = c_1 + c_3 c_2$$
$$b_2 = \mathrm{cov}[X_2, Y] = \mathrm{cov}[X_2, \beta_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 r_{12} + c_3\,\mathrm{cov}[X_2, U'] = c_1 r_{12}$$
$$b_3 = \mathrm{cov}[X_3, Y] = \mathrm{cov}[X_3, \beta_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 r_{13} + c_3\,\mathrm{cov}[X_3, U'] = c_1 r_{13}$$

Mutually adjusted for X_1 and X_i (either X_2 or X_3):

$$b_1 = \frac{\mathrm{cov}[X_1, Y] - r_{1i}\,\mathrm{cov}[X_i, Y]}{1 - r_{1i}^2} = \frac{\left(c_1 + c_3 c_2\right) - r_{1i}\left(c_1 r_{1i}\right)}{1 - r_{1i}^2} = c_1 + \frac{c_3 c_2}{1 - r_{1i}^2}$$
$$b_i = \frac{\mathrm{cov}[X_i, Y] - r_{1i}\,\mathrm{cov}[X_1, Y]}{1 - r_{1i}^2} = \frac{\left(c_1 r_{1i}\right) - r_{1i}\left(c_1 + c_3 c_2\right)}{1 - r_{1i}^2} = -\frac{r_{1i} c_3 c_2}{1 - r_{1i}^2}$$

Mutually adjusted for all three exposures (ignoring the constant coefficient b_0):

$$b=\left(X'X\right)^{-1}X'Y$$

$$A=X'X=\begin{pmatrix}1 & r_{12} & r_{13}\\ r_{12} & 1 & r_{23}\\ r_{13} & r_{23} & 1\end{pmatrix}$$

$$X'Y=\begin{pmatrix}\operatorname{cov}[X_1,Y]\\ \operatorname{cov}[X_2,Y]\\ \operatorname{cov}[X_3,Y]\end{pmatrix}=\begin{pmatrix}c_1+c_3c_2\\ c_1r_{12}\\ c_1r_{13}\end{pmatrix}$$

$$\det[A]=1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}$$

$$b=\frac{1}{\det[A]}\begin{pmatrix}1-r_{23}^2 & r_{13}r_{23}-r_{12} & r_{12}r_{23}-r_{13}\\ r_{13}r_{23}-r_{12} & 1-r_{13}^2 & r_{12}r_{13}-r_{23}\\ r_{12}r_{23}-r_{13} & r_{12}r_{13}-r_{23} & 1-r_{12}^2\end{pmatrix}\begin{pmatrix}c_1+c_3c_2\\ c_1r_{12}\\ c_1r_{13}\end{pmatrix}$$

$$b_1=\frac{\left(1-r_{23}^2\right)\left(c_1+c_3c_2\right)+\left(r_{13}r_{23}-r_{12}\right)c_1r_{12}+\left(r_{12}r_{23}-r_{13}\right)c_1r_{13}}{\det[A]}=c_1+\frac{\left(1-r_{23}^2\right)c_3c_2}{\det[A]}$$

$$b_2=\frac{\left(r_{13}r_{23}-r_{12}\right)\left(c_1+c_3c_2\right)+\left(1-r_{13}^2\right)c_1r_{12}+\left(r_{12}r_{13}-r_{23}\right)c_1r_{13}}{\det[A]}=\frac{\left(r_{13}r_{23}-r_{12}\right)c_3c_2}{\det[A]}$$

$$b_3=\frac{\left(r_{12}r_{23}-r_{13}\right)\left(c_1+c_3c_2\right)+\left(r_{12}r_{13}-r_{23}\right)c_1r_{12}+\left(1-r_{12}^2\right)c_1r_{13}}{\det[A]}=\frac{\left(r_{12}r_{23}-r_{13}\right)c_3c_2}{\det[A]}$$

Adjusting for X_3 amplifies the bias for b_1 beyond adjusting for just X_2 when:

$$\frac{\left(1-r_{23}^2\right)c_2c_3}{1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}}>\frac{c_2c_3}{1-r_{12}^2}\leftrightarrow\left(1-r_{23}^2\right)\left(1-r_{12}^2\right)>1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23},$$

if $\det[A]>0$, $c_2c_3>0$

$$\leftrightarrow 1-r_{12}^2-r_{23}^2+r_{12}^2r_{23}^2>1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}\leftrightarrow r_{12}^2r_{23}^2+r_{13}^2-2r_{12}r_{13}r_{23}=\left(r_{12}r_{23}-r_{13}\right)^2>0$$

Note that the determinant of the exposure correlation matrix is the denominator of the amplification factor (as it also is for 2 exposure variables).
$\det[A]>0$ because exposure variance-covariance matrices are positive semidefinite, positive definite if non-singular (Horn RA and Johnson CR. *Matrix Analysis*. Cambridge University Press 1985; pg 392, 396-399). Therefore, the amplification is always larger for b_1 unless $r_{12}r_{23}=r_{13}$ (and similarly for adjusting for X_2 beyond adjusting for X_3). We hypothesize that this will also be true for higher order systems.

Addition to DAG in Figure 4a of an arrow from X_2 to Y

We'll analyze the effect on b_1 (analyses for b_2 and b_3 not shown). No new computations are required as the results can be derived from earlier DAGs, i.e., adding X_3 does not change the pathways between X_1 and Y or X_2 and Y. The crude estimate of b_1 is equivalent to that in DAG 3b:

$$b_1 = c_1 + r_{12}c_4 + c_2c_3$$

The estimate of b_1 adjusted for X_2 alone is also equivalent to that in DAG 3b:

________ $$b_1 = c_1 + \frac{c_2c_3}{1-r_{12}^2}$$

The estimate of b_1 adjusted for X_2 and X_3 is the same as that given above for DAG 4a:

________ $$b_1 = c_1 + \frac{\left(1-r_{23}^2\right)c_2c_3}{\det[A]}$$

Adding the arrow from X2 to Y does not change this computation because the new pathway from X_3 to Y through X_2 is blocked.

Compared to DAG 3b, the confounding caused by U' may be further amplified by adding X_3 and certainly would be if $r_{23}=0$ (as in DAG 4b). Similar to DAG 3b, the choice of whether to adjust the X_1-Y association for both X_2 and X_3 requires a balancing of confounding through X_2 vs. amplification of confounding through U', but the latter may now be even larger. The bias of the crude b_1 is larger than fully adjusted

when $r_{12}c_4 + c_2c_3 > \frac{c_2c_3\left(1-r_{23}^2\right)}{\det[A]} \leftrightarrow r_{12}c_4 > \frac{c_2c_3\left(1-r_{23}^2-\det[A]\right)}{\det[A]} \leftrightarrow c_4 > \frac{c_2c_3\left(1-r_{23}^2-\det[A]\right)}{r_{12}\det[A]}$ if $r_{12}, c_4, c_2c_3 > 0$

DAG in Figure 4b:

This is a special case of the previous derivation with $r_{23}=0$. The amplification factor for b_1 then depends solely on det(A).

DAG in Figure 5a

$$Y = b_0 + c_1X_1 + c_3U' + \varepsilon$$
$$\text{cov}[X_1, U'] = c_2c_4$$
$$\text{cov}[X_2, U'] = c_2c_5$$
$$\text{cov}[X_1, X_2] = c_4c_5 = r_{12} \text{ since variables are standardized}$$

Single exposure analyses:

$$b_1 = \text{cov}[X_1, Y] = \text{cov}[X_1, b_0 + c_1X_1 + c_3U'] = c_1 + c_3\,\text{cov}[X_1, U'] = c_1 + c_3c_2c_4$$
$$b_2 = \text{cov}[X_2, Y] = \text{cov}[X_2, b_0 + c_1X_1 + c_3U'] = c_1r_{12} + c_3\,\text{cov}[X_2, U'] = c_1r_{12} + c_3c_2c_5$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{cov}[X_1, Y] - r_{12}\,\text{cov}[X_2, Y]}{1-r_{12}^2} = \frac{\left(c_1 + c_3c_2c_4\right) - r_{12}\left(c_1r_{12} + c_3c_2c_5\right)}{1-r_{12}^2}$$
$$= c_1 + c_3c_2\left(\frac{c_4 - r_{12}c_5}{1-r_{12}^2}\right) = c_1 + c_3c_2c_4\left(\frac{1-c_5^2}{1-c_4^2c_5^2}\right)$$
$$b_2 = \frac{\text{cov}[X_2, Y] - r_{12}\,\text{cov}[X_1, Y]}{1-r_{12}^2} = \frac{\left(c_1r_{12} + c_3c_2c_5\right) - r_{12}\left(c_1 + c_3c_2c_4\right)}{1-r_{12}^2}$$
$$= c_3c_2\left(\frac{c_5 - r_{12}c_4}{1-r_{12}^2}\right) = c_3c_2c_5\left(\frac{1-c_4^2}{1-c_4^2c_5^2}\right)$$

b_1: adjusted bias < crude bias when:

$$c_2c_3c_4\left(\frac{1-c_5^2}{1-c_4^2c_5^2}\right)<c_2c_3c_4 \leftrightarrow 1-c_5^2<1-c_4^2c_5^2 \leftrightarrow c_4^2c_5^2<c_5^2 \leftrightarrow c_4^2<1 \quad \text{if } c_2c_3c_4>0$$

b_2: adjusted bias < crude bias when:

$$c_2c_3c_5\left(\frac{1-c_4^2}{1-c_4^2c_5^2}\right)<c_1c_4c_5+c_2c_3c_5 \leftrightarrow c_2c_3\left(\frac{1-c_4^2}{1-c_4^2c_5^2}-1\right)<c_1c_4 \leftrightarrow c_2c_3\left(\frac{-c_4^2+c_4^2c_5^2}{1-c_4^2c_5^2}\right)<c_1c_4$$

$$\leftrightarrow c_2c_3c_4^2\left(\frac{c_5^2-1}{1-c_4^2c_5^2}\right)<0<c_1c_4 \quad \text{since } c_5^2-1<0 \text{ and } 1-c_4^2c_5^2>0, \quad \text{if } c_5,\ c_2c_3,\ c_1c_4>0$$

Other cases can be derived when c_5, c_2c_3, or $c_1c_4<0$.

DAG in Figure 5b

$$Y=b_0+c_1X_1+c_3U'+\varepsilon$$
$$\text{cov}[X_1,U']=c_2$$
$$\text{cov}[X_2,U']=c_4$$
$$\text{cov}[X_1,X_2]=c_2c_4=r_{12} \quad \text{since variables are standardized}$$

Single exposure analyses:

$$b_1=\text{cov}[X_1,Y]=\text{cov}[X_1,b_0+c_1X_1+c_3U']=c_1+c_3\,\text{cov}[X_1,U']=c_1+c_3c_2$$
$$b_2=\text{cov}[X_2,Y]=\text{cov}[X_2,b_0+c_1X_1+c_3U']=c_1r_{12}+c_3\,\text{cov}[X_2,U']=c_1r_{12}+c_3c_4=c_1c_2c_4+c_3c_4$$

Mutually adjusted analysis:

$$b_1=\frac{\text{cov}[X_1,Y]-r_{12}\,\text{cov}[X_2,Y]}{1-r_{12}^2}=\frac{(c_1+c_3c_2)-r_{12}(c_1r_{12}+c_3c_4)}{1-r_{12}^2}$$
$$=c_1+\frac{c_3(c_2-r_{12}c_4)}{1-r_{12}^2}=c_1+c_3c_2\left(\frac{1-c_4^2}{1-c_2^2c_4^2}\right)$$
$$b_2=\frac{\text{cov}[X_2,Y]-r_{12}\,\text{cov}[X_1,Y]}{1-r_{12}^2}=\frac{(c_1r_{12}+c_3c_4)-r_{12}(c_1+c_3c_2)}{1-r_{12}^2}$$
$$=\frac{c_3(c_4-r_{12}c_2)}{1-r_{12}^2}=c_3c_4\left(\frac{1-c_2^2}{1-c_2^2c_4^2}\right)$$

DAG in Figure 5c

$$Y=b_0+c_1X_1+c_3U'+\varepsilon$$
$$\text{cov}[X_1,U']=c_2 \quad \text{because } X_2 \text{ is a collider on } X_1 \leftarrow U \rightarrow X_2 \leftarrow U'$$
$$\text{cov}[X_2,U']=c_4 \quad \text{because } X_1 \text{ is a collider on } X_2 \leftarrow U \rightarrow X_1 \leftarrow U'$$
$$\text{cov}[X_1,X_2]=r_{12} \quad \text{includes components on both } X_1 \leftarrow U \rightarrow X_2 \text{ and } X_1 \leftarrow U' \rightarrow X_2$$

Single exposure analyses:

$$b_1 = \text{cov}[X_1, Y] = \text{cov}[X_1, b_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 + c_3 \text{cov}[X_1, U'] = c_1 + c_3 c_2$$

$$b_2 = \text{cov}[X_2, Y] = \text{cov}[X_2, b_0 + c_1 X_1 + c_3 U' + \varepsilon] = c_1 r_{12} + c_3 \text{cov}[X_2, U'] = r_{12} c_1 + c_3 c_4$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{cov}[X_1, Y] - r_{12}\text{cov}[X_2, Y]}{1 - r_{12}^2} = \frac{(c_1 + c_3 c_2) - r_{12}(c_1 r_{12} + c_3 c_4)}{1 - r_{12}^2}$$

$$= c_1 + \frac{c_3(c_2 - r_{12} c_4)}{1 - r_{12}^2} = c_1 + \frac{c_2 c_3}{1 - r_{12}^2}\left(1 - r_{12}\frac{c_4}{c_2}\right)$$

$$b_2 = \frac{\text{cov}[X_2, Y] - r_{12}\text{cov}[X_1, Y]}{1 - r_{12}^2} = \frac{(c_1 r_{12} + c_3 c_4) - r_{12}(c_1 + c_3 c_2)}{1 - r_{12}^2}$$

$$= \frac{c_3(c_4 - r_{12} c_2)}{1 - r_{12}^2} = \frac{c_3 c_4}{1 - r_{12}^2}\left(1 - r_{12}\frac{c_2}{c_4}\right) = -\frac{c_2 c_3}{1 - r_{12}^2}\left(r_{12} - \frac{c_4}{c_2}\right) = -\frac{r_{12} c_2 c_3}{1 - r_{12}^2}\left(1 - \frac{c_4}{r_{12} c_2}\right)$$

DAG in Figure 5d

$$Y = b_0 + c_1 X_1 + c_5 X_2 + c_3 U' + \varepsilon$$

$\text{cov}[X_1, U'] = c_2$ because X_2 is a collider on $X_1 \leftarrow U \rightarrow X_2 \leftarrow U'$

$\text{cov}[X_2, U'] = c_4$ because X_1 is a collider on $X_2 \leftarrow U \rightarrow X_1 \leftarrow U'$

$\text{cov}[X_1, X_2] = r_{12}$ includes components on both $X_1 \leftarrow U \rightarrow X_2$ and $X_1 \leftarrow U' \rightarrow X_2$

Single exposure analysis:

$$b_1 = \text{cov}[X_1, Y] = \text{cov}[X_1, b_0 + c_1 X_1 + c_5 X_2 + c_3 U' + \varepsilon] = c_1 + c_5 r_{12} + c_3 \text{cov}[X_1, U'] = c_1 + c_5 r_{12} + c_3 c_2$$

$$b_2 = \text{cov}[X_2, Y] = \text{cov}[X_2, b_0 + c_1 X_1 + c_5 X_2 + c_3 U' + \varepsilon] = c_1 r_{12} + c_5 + c_3 \text{cov}[X_2, U'] = r_{12} c_1 + c_5 + c_3 c_4$$

Mutually adjusted analysis:

$$b_1 = \frac{\text{cov}[X_1, Y] - r_{12}\text{cov}[X_2, Y]}{1 - r_{12}^2} = \frac{(c_1 + r_{12} c_5 + c_3 c_2) - r_{12}(c_1 r_{12} + c_5 + c_3 c_4)}{1 - r_{12}^2}$$

$$= c_1 + \frac{c_3(c_2 - r_{12} c_4)}{1 - r_{12}^2} = c_1 + \frac{c_2 c_3}{1 - r_{12}^2}\left(1 - r_{12}\frac{c_4}{c_2}\right)$$

$$b_2 = \frac{\text{cov}[X_2, Y] - r_{12}\text{cov}[X_1, Y]}{1 - r_{12}^2} = \frac{(c_1 r_{12} + c_5 + c_3 c_4) - r_{12}(c_1 + r_{12} c_5 + c_3 c_2)}{1 - r_{12}^2}$$

$$= c_5 + \frac{c_3(c_4 - r_{12} c_2)}{1 - r_{12}^2} = c_5 + \frac{c_3 c_4}{1 - r_{12}^2}\left(1 - r_{12}\frac{c_2}{c_4}\right) = c_5 - \frac{c_2 c_3}{1 - r_{12}^2}\left(r_{12} - \frac{c_4}{c_2}\right) = c_5 - \frac{r_{12} c_2 c_3}{1 - r_{12}^2}\left(1 - \frac{c_4}{r_{12} c_2}\right)$$

Suppose the crude and adjusted b_1 and b_2 are greater than zero.
b_1: The bias of the crude estimate exceeds bias of the mutually adjusted estimate when

$$r_{12}c_5 + c_2c_3 > \frac{c_3\left(c_2 - r_{12}c_4\right)}{1-r_{12}^2} \leftrightarrow r_{12}c_5 > \frac{\left(c_2c_3 - r_{12}c_3c_4 - \left(1-r_{12}^2\right)c_2c_3\right)}{1-r_{12}^2} \leftrightarrow r_{12}c_5 > \frac{\left(-r_{12}c_3c_4 + r_{12}^2c_2c_3\right)}{1-r_{12}^2}$$

$$\leftrightarrow c_5 > \frac{c_3\left(-c_4 + r_{12}c_2\right)}{1-r_{12}^2} \quad \text{if } r_{12} > 0$$

$$\leftrightarrow c_5 + \frac{c_3\left(c_4 - r_{12}c_2\right)}{1-r_{12}^2} > 0$$

$$\leftrightarrow \text{adjusted } b_2 > 0$$

b_2: The bias of the crude estimate exceeds bias of the mutually adjusted estimate when

$$r_{12}c_1 + c_3c_4 > \frac{c_3\left(c_4 - r_{12}c_2\right)}{1-r_{12}^2} \leftrightarrow r_{12}c_1 > \frac{\left(c_3c_4 - r_{12}c_2c_3 - \left(1-r_{12}^2\right)c_3c_4\right)}{1-r_{12}^2} \leftrightarrow r_{12}c_1 > \frac{\left(-r_{12}c_2c_3 + r_{12}^2c_3c_4\right)}{1-r_{12}^2}$$

$$\leftrightarrow c_1 > \frac{c_3\left(-c_2 + r_{12}c_4\right)}{1-r_{12}^2} \quad \text{if } r_{12} > 0$$

$$\leftrightarrow c_1 + \frac{c_3\left(c_2 - r_{12}c_4\right)}{1-r_{12}^2} > 0$$

$$\leftrightarrow \text{adjusted } b_1 > 0$$