All equation references in this eAppendix refer to equations in the main manuscript.

eAppendix A: Asymptotic bias under data generating models (5), (8), (9) and (10)

Preliminaries

Identity link

Suppose that data consist of n sibling-pairs. Define $\bar{X}_i = (X_{i1} + X_{i2})/2$ and $\bar{C}_i = (C_{i1} + C_{i2})/2$. The conditional maximum likelihood estimator $(\hat{\beta}_{ide}, \hat{\delta}_{ide})$ solves the equation system

$$\sum_{i=1}^{n} S_{i,ide}(\hat{\beta}_{ide}, \hat{\delta}_{ide}) = 0,$$

where

$$\begin{aligned} S_{i,ide}(\hat{\beta}_{ide}, \hat{\delta}_{ide}) &= \sum_{j=1}^{2} \begin{pmatrix} X_{ij} - \bar{X}_i \\ C_{ij} - \bar{C}_i \end{pmatrix} (Y_{ij} - \hat{\beta}_{ide} X_{ij} - \hat{\delta}_{ide} C_{ij}) \\ &= \frac{1}{2} \begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} \{(Y_{i1} - \hat{\beta}_{ide} X_{i1} - \hat{\delta}_{ide} C_{i1}) - (Y_{i2} - \hat{\beta}_{ide} X_{i2} - \hat{\delta}_{ide} C_{i2})\}. \end{aligned}$$

It follows from standard theory that $(\hat{\beta}_{ide}, \hat{\delta}_{ide})$ converges to (β^*, δ^*), defined as the solution to

$$E\{S_{i,ide}(\beta^*, \delta^*)\} = 0.$$

Logit link

The conditional maximum likelihood estimator $(\hat{\beta}_{ide}, \hat{\delta}_{ide})$ solves the equation system

$$\sum_{i=1}^{n} S_{i,logit}(\hat{\beta}_{logit}, \hat{\delta}_{logit}) = 0,$$

where

$$S_{i,logit}(\hat{\beta}_{logit}, \hat{\delta}_{logit}) = I(Y_{i1} \neq Y_{i2}) \sum_{j=1}^{2} \begin{pmatrix} X_{ij} \\ C_{ij} \end{pmatrix} \left(Y_{ij} - \frac{e^{\hat{\beta}_{logit} X_{ij} + \hat{\delta}_{logit} C_{ij}}}{e^{\hat{\beta}_{logit} X_{i1} + \hat{\delta}_{logit} C_{i1}} + e^{\hat{\beta}_{logit} X_{i2} + \hat{\delta}_{logit} C_{i2}}} \right)$$

It follows from standard theory that $(\hat{\beta}_{logit}, \hat{\delta}_{logit})$ converges to (β^*, δ^*), defined as the solution to

$$E\{S_{i,logit}(\beta^*, \delta^*)|Y_{i1} \neq Y_{i2}\} = 0.$$

Carry over from X_{i1} to Y_{i2}

Identity link

Define $\mathbf{Q}_i = (X_{i1}, X_{i2}, C_{i1}, C_{i2}, U_i)$. Under the causal diagram in Figure 1C) and model (5) with $g(\cdot) =$ identity we have that

$$\begin{aligned} E(Y_{i1}|\mathbf{Q}_i) &= \alpha_i + \beta X_{i1} + \delta C_{i1} \\ E(Y_{i2}|\mathbf{Q}_i) &= \alpha_i + \beta X_{i2} + \gamma X_{i1} + \delta C_{i2} + \psi. \end{aligned}$$

It follows that

$$\begin{aligned} &E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\ &= \frac{1}{2} \begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} \{(\beta X_{i1} + \delta C_{i1} - \beta^* X_{i1} - \delta^* C_{i1}) \\ &\quad - (\beta X_{i2} + \gamma X_{i1} + \delta C_{i2} + \psi - \beta^* X_{i2} - \delta^* C_{i2})\}. \end{aligned}$$

When $(\beta^*, \delta^*) = (\beta - \gamma/2, \delta)$ this expression simplifies to

$$\frac{1}{2} \begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} \{-\gamma(X_{i1} + X_{i2})/2 - \psi\},$$

which has conditional mean equal to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

given U_i, under the symmetry condition in (6).

Logit link

We give a simulated example where $\hat{\beta}_{logit}$ does not converge to $\beta - \gamma/2$ under the causal diagram in Figure 1C) and model (5). In this example we generated 10000 samples consisting of 10000 pairs each from the model

$$\begin{aligned} U_i &\sim N(0,1) \\ X_{i1}|U_i &\sim Ber\{\text{expit}(U_i)\} \\ Y_{i1}|X_{i1}, U_i &\sim Ber\{\text{expit}(U_i + \beta X_{i1})\} \\ X_{i2}|U_i &\sim Ber\{\text{expit}(U_i)\} \\ Y_{i2}|X_{i2}, X_{i1}, U_i &\sim Ber\{\text{expit}(U_i + \beta X_{i2} + \gamma X_{i1} + \psi)\}, \end{aligned}$$

with $\beta = 1$, $\gamma = 0.5$, $\psi = 2$, and no non-shared confounders. For each sample we calculated $\hat{\beta}_{logit}$. The mean (over the 10000 samples) of $\hat{\beta}_{logit}$ was 0.62, which is not equal to $\beta - \gamma/2 = 0.75$.

Carry over from X_{i1} to Y_{i2} and from X_{i2} to Y_{i1}

Identity link

Under the causal diagram in Figure 1D) and model (8) with $g(\cdot)$ = identity we have that

$$E(Y_{ij}|\mathbf{Q}_i) = \alpha_i + \beta X_{ij} + \gamma X_{ij'} + \delta C_{ij}.$$

It follows that

$$\begin{aligned} & E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\ & \quad = \frac{1}{2}\begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} \{(\beta X_{i1} + \gamma X_{i2} + \delta C_{i1} - \beta^* X_{i1} - \delta^* C_{i1}) \\ & \quad\quad - (\beta X_{i2} + \gamma X_{i1} + \delta C_{i2} - \beta^* X_{i2} - \delta^* C_{i2})\}. \end{aligned}$$

When $(\beta^*, \delta^*) = (\beta - \gamma, \delta)$ this expression equals

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Logit link

Under the causal diagram in Figure 1D) and model (8) with $g(\cdot)$ = logit we have that

$$\begin{aligned}
E(Y_{ij}|\mathbf{Q}_i, Y_{i1} \neq Y_{i2}) &= \frac{\Pr(Y_{ij}=1|\mathbf{Q}_i)\Pr(Y_{ij'}=0|\mathbf{Q}_i)}{\Pr(Y_{i1}=1|\mathbf{Q}_i)\Pr(Y_{i2}=0|\mathbf{Q}_i) + \Pr(Y_{i1}=0|\mathbf{Q}_i)\Pr(Y_{i2}=1|\mathbf{Q}_i)} \\
&= \frac{e^{\beta X_{ij}+\gamma X_{ij'}+\delta C_{ij}}}{e^{\beta X_{i1}+\gamma X_{i2}+\delta C_{i1}} + e^{\beta X_{i2}+\gamma X_{i1}+\delta C_{i2}}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&E\{S_{i,logit}(\beta^*, \delta^*)|\mathbf{Q}_i, Y_{i1} \neq Y_{i2}\} \\
&\quad = \sum_{j=1}^{2} \begin{pmatrix} X_{ij} \\ C_{ij} \end{pmatrix} \left(\frac{e^{\beta X_{ij}+\gamma X_{ij'}+\delta C_{ij}}}{e^{\beta X_{i1}+\gamma X_{i2}+\delta C_{i1}} + e^{\beta X_{i2}+\gamma X_{i1}+\delta C_{i2}}} - \frac{e^{\beta^* X_{ij}+\delta^* C_{ij}}}{e^{\beta^* X_{i1}+\delta^* C_{i1}} + e^{\beta^* X_{i2}+\delta^* C_{i2}}} \right) \\
&\quad = \sum_{j=1}^{2} \begin{pmatrix} X_{ij} \\ C_{ij} \end{pmatrix} \left(\frac{e^{\beta X_{ij}-\gamma X_{ij}+\delta C_{ij}}}{e^{\beta X_{i1}-\gamma X_{i1}+\delta C_{i1}} + e^{\beta X_{i2}-\gamma X_{i2}+\delta C_{i2}}} - \frac{e^{\beta^* X_{ij}+\delta^* C_{ij}}}{e^{\beta^* X_{i1}+\delta^* C_{i1}} + e^{\beta^* X_{i2}+\delta^* C_{i2}}} \right).
\end{aligned}$$

When $(\beta^*, \delta^*) = (\beta - \gamma, \delta)$ this expression equals

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Carry over from Y_{i1} to Y_{i2}

Identity link

Under the causal diagram in Figure 1E) and model (9) with $g(\cdot)$ = identity we have that

$$\begin{aligned}
E(Y_{i1}|\mathbf{Q}_i) &= \alpha_i + \beta X_{i1} + \delta C_{i1} \\
E(Y_{i2}|\mathbf{Q}_i) &= E\{E(Y_{i2}|Y_{i1}, \mathbf{Q}_i)|\mathbf{Q}_i\} \\
&= E(\alpha_i + \beta X_{i2} + \gamma Y_{i1} + \delta C_{i2} + \psi|\mathbf{Q}_i) \\
&= \alpha_i + \beta X_{i2} + \gamma(\alpha_i + \beta X_{i1} + \delta C_{i1}) + \psi.
\end{aligned}$$

It follows that

$$\begin{aligned}
&E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\
&\quad = \frac{1}{2} \begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} [(\beta X_{i1} + \delta C_{i1} - \beta^* X_{i1} - \delta^* C_{i1}) \\
&\qquad - \{\beta X_{i2} + \gamma(\alpha_i + \beta X_{i1} + \delta C_{i1}) + \psi - \beta^* X_{i2} - \delta^* C_{i2}\}].
\end{aligned}$$

When $(\beta^*, \delta^*) = (\beta - \beta\gamma/2, \delta - \delta\gamma/2)$ this expression simplifies to

$$\frac{1}{2}\begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix} \{-\beta\gamma(X_{i1} + X_{i2})/2 - \delta\gamma(C_{i1} + C_{i2})/2 - \gamma\alpha_i - \psi\},$$

which has conditional mean equal to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

given U_i, under the symmetry condition in (6).

Logit link

We first show that $\hat{\beta}_{logit}$ converges to 0 if there are no non-shared confounders (C_{ij} is the empty set) and $\beta = 0$. Under the causal diagram in Figure 1E) and model (9) with $g(\cdot)$ = identity we then have that

$$\begin{aligned}
E(Y_{i1}|\mathbf{Q}_i, Y_{i1} \neq Y_{i2}) &= \frac{\Pr(Y_{i1} = 1|\mathbf{Q}_i)\Pr(Y_{i2} = 0|\mathbf{Q}_i)}{\Pr(Y_{i1} = 1|\mathbf{Q}_i)\Pr(Y_{i2} = 0|\mathbf{Q}_i) + \Pr(Y_{i1} = 0|\mathbf{Q}_i)\Pr(Y_{i2} = 1|\mathbf{Q}_i)} \\
&= \frac{\frac{e^{\alpha_i}}{1+e^{\alpha_i}}\frac{1}{1+e^{\alpha_i+\gamma+\psi}}}{\frac{e^{\alpha_i}}{1+e^{\alpha_i}}\frac{1}{1+e^{\alpha_i+\gamma+\psi}} + \frac{1}{1+e^{\alpha_i}}\frac{e^{\alpha_i+\psi}}{1+e^{\alpha_i+\psi}}} \\
&= V_i \\
E(Y_{i2}|\mathbf{Q}_i, Y_{i1} \neq Y_{i2}) &= \frac{\Pr(Y_{i2} = 1|\mathbf{Q}_i)\Pr(Y_{i1} = 0|\mathbf{Q}_i)}{\Pr(Y_{i1} = 0|\mathbf{Q}_i)\Pr(Y_{i2} = 1|\mathbf{Q}_i) + \Pr(Y_{i1} = 1|\mathbf{Q}_i)\Pr(Y_{i2} = 0|\mathbf{Q}_i)} \\
&= \frac{\frac{1}{1+e^{\alpha_i}}\frac{e^{\alpha_i+\psi}}{1+e^{\alpha_i+\psi}}}{\frac{1}{1+e^{\alpha_i}}\frac{e^{\alpha_i+\psi}}{1+e^{\alpha_i+\psi}} + \frac{e^{\alpha_i}}{1+e^{\alpha_i}}\frac{1}{1+e^{\alpha_i+\gamma+\psi}}} \\
&= 1 - V_i
\end{aligned}$$

.

It follows that

$$E\{S_{i,logit}(\beta^*)|\mathbf{Q}_i, Y_{i1} \neq Y_{i2}\} = X_{i1}\left(V_i - \frac{e^{\beta^* X_{i1}}}{e^{\beta^* X_{i1}} + e^{\beta^* X_{i2}}}\right) + X_{i2}\left(1 - V_i - \frac{e^{\beta^* X_{i2}}}{e^{\beta^* X_{i1}} + e^{\beta^* X_{i2}}}\right).$$

When $\beta^* = 0$ this expression simplifies to

$$(X_{i1} - X_{i2})(V_i - 1/2),$$

which has conditional mean equal to 0, given U_i, under the symmetry condition in (6).

We next give a simulated example where $\hat{\beta}_{logit}$ does not converge to 0 in the presence of non-shared confounders. In this example we generated 10000 samples consisting of 10000 pairs each from the model

$$\begin{aligned} U_i &\sim N(0,1) \\ C_{ij} &\sim N(0,1) \\ X_{ij}|C_{ij},U_i &\sim N\{\text{abs}(U_i+C_{ij}),0.01\} \\ Y_{i1}|X_{i1},C_{i1},U_i &\sim Ber\{\text{expit}(U_i+\beta X_{i1}+\delta C_{i1})\} \\ Y_{i2}|X_{i2},Y_{i1},U_i &\sim Ber\{\text{expit}(U_i+\beta X_{i2}+\gamma Y_{i1}+\delta C_{i2}+\psi)\}, \end{aligned}$$

with $\beta = 0$, $\delta = 1$, $\gamma = 1$, $\psi = 2$, and no non-shared confounders. For each sample we calculated $\hat{\beta}_{logit}$. The mean (over the 10000 samples) of $\hat{\beta}_{logit}$ was 0.05.

Carry over from Y_{i1} to X_{i2}

Identity link

In the absence of non-shared confounders, $\hat{\beta}_{ide}$ is equal to

$$\frac{\sum_i (X_{i1}-X_{i2})(Y_{i1}-Y_{i2})}{\sum_i (X_{i1}-X_{i2})^2},$$

which converges to

$$\frac{E\{(X_{i1}-X_{i2})(Y_{i1}-Y_{i2})\}}{E\{(X_{i1}-X_{i2})^2\}} = \frac{E\{(X_{i1}-X_{i2})E(Y_{i1}-Y_{i2}|X_{i1},X_{i2},U_i)\}}{E\{(X_{i1}-X_{i2})^2\}}.$$

Under model (10) we have that $Y_{i1} = \alpha_i + \beta(\alpha_i + \epsilon_{i1}) + \epsilon_{i2}$, $X_{i2} = \alpha_i + \gamma\{\alpha_i + \beta(\alpha_i + \epsilon_{i1}) + \epsilon_{i2}\} + \epsilon_{i3}$ and $Y_{i2} = \alpha_i + \beta[\alpha_i + \gamma\{\alpha_i + \beta(\alpha_i + \epsilon_{i1}) + \epsilon_{i2}\} + \epsilon_{i3}] + \epsilon_{i4}$. Plugging these into the RHS of the equation above, and assuming that α_i, ϵ_{i1}, ϵ_{i2}, ϵ_{i3} and ϵ_{i4} are independent with mean 0 and variance 1, gives the expression in (11).

Even though it is difficult to avoid bias under this type of carryover effect, it is possible to construct a valid test of the null hypothesis. To see this, consider removing the arrows from from X_{i1} to Y_{i1} and from X_{i2} to Y_{i2} in Figure 1F), so that the null hypothesis holds. Then, consider a relabelling of the variables in Figure 1F), so that X_{i1} and X_{i2} are relabelled as Y_{i1} and Y_{i2}, respectively, and Y_{i1} and Y_{i2} are relabelled as X_{i1} and X_{i2}, respectively. After these manipulations, the causal diagram in Figure 1F) has the same structure as the causal diagram in Figure 1C). In Section 'Carryover from X_{i1} to Y_{i2}' we considered the data generating model in (5), which may be plausible under the causal diagram in Figure 1C), and we showed that this model can be fitted without bias. Thus, to test the null hypothesis in Figure 1F) we may swap the exposure and outcome, fit the model in (5), and test whether the obtained estimate of β is equal to 0. Under the null hypothesis we would expect this estimate to converge to 0, and thus it provides a valid test. We emphasize that when the null hypothesis doesn't hold, the obtained estimate has no simple interpretation as an exposure effect in Figure 1F).

eAppendix B: Robustness of model (1) to misspecification of the intercept

Under models (5) and (9), with $g(\cdot)$ = identity, we have that

$$E(Y_{ij}|X_{ij}, C_{ij}, U_i) = \alpha_{ij} + \beta X_{ij} + \delta C_{ij},$$

where $\alpha_{i1} = \alpha_i$. Under model (5), $\alpha_{i2} = \alpha_i + \gamma E(X_{i1}|X_{i2}, C_{i2}, U_i) = \alpha_i + \gamma E(X_{i1}|U_i)$, whereas under model (9), $\alpha_{i2} = \alpha_i + \gamma E(Y_{i1}|X_{i2}, C_{i2}, U_i) = \alpha_i + \gamma E(Y_{i1}|U_i)$. Thus, the carryover effects in Figures 1C) and 1E) can be viewed as making the intercept in model (1) both family- and sibling-specific. It follows that

$$\begin{aligned} &E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\ &\quad = \frac{1}{2}\begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix}\{(\alpha_{i1} + \beta X_{i1} + \delta C_{i1} - \beta^* X_{i1} - \delta^* C_{i1}) \\ &\quad - (\alpha_{i2} + \beta X_{i2} + \delta C_{i2} - \beta^* X_{i2} - \delta^* C_{i2})\}. \end{aligned}$$

When $(\beta^*, \delta^*) = (\beta, \delta)$ this expression simplifies to

$$\frac{1}{2}\begin{pmatrix} X_{i1} - X_{i2} \\ C_{i1} - C_{i2} \end{pmatrix}(\alpha_{i1} - \alpha_{i2}),$$

which has conditional mean equal to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

given U_i, under the symmetry condition in (6).

eAppendix C: Identifiability problem of model (8)

Under model (8), the conditional maximum likelihood estimator $(\hat{\beta}_{ide}, \hat{\gamma}_{ide}, \hat{\delta}_{ide})$ solves the equation system

$$\sum_{i=1}^{n} S_{i,ide}(\hat{\beta}_{ide}, \hat{\gamma}_{ide}, \hat{\delta}_{ide}) = 0,$$

where

$$\begin{aligned} S_{i,ide}(\hat{\beta}_{ide}, \hat{\gamma}_{ide}, \hat{\delta}_{ide}) &= \sum_{j=1}^{2} \begin{pmatrix} X_{ij} - \bar{X}_i \\ X_{ij'} - \bar{X}_i \\ C_{ij} - \bar{C}_i \end{pmatrix} (Y_{ij} - \hat{\beta}_{ide} X_{ij} - \hat{\gamma}_{ide} X_{ij'} - \hat{\delta}_{ide} C_{ij}) \\ &= \frac{1}{2} \begin{pmatrix} X_{i1} - X_{i2} \\ X_{i2} - X_{i1} \\ C_{i1} - C_{i2} \end{pmatrix} \{(Y_{i1} - \hat{\beta}_{ide} X_{i1} - \hat{\gamma}_{ide} X_{i2} - \hat{\delta}_{ide} C_{i1}) \\ &\quad - (Y_{i2} - \hat{\beta}_{ide} X_{i2} - \hat{\gamma}_{ide} X_{i1} - \hat{\delta}_{ide} C_{i2})\}. \end{aligned}$$

It is easy to see that the first two equations are identical, and thus the equation system has no unique solution.

eAppendix D: The bidirectional design in the presence of carryover from X_{i1} to Y_{i2}.

Identity link

Define $\mathbf{Q}_i = (X_{i1} = 1, X_{i2} = 0, C_{i1}, C_{i2}, U_i)$. Under the causal diagram in Figure 1C) and model (5) with $g(\cdot) =$ identity we have that

$$\begin{aligned} E(Y_{i1}|\mathbf{Q}_i) &= \alpha_i + \beta + \delta C_{i1} \\ E(Y_{i2}|\mathbf{Q}_i) &= \alpha_i + \gamma + \delta C_{i2} + \psi. \end{aligned}$$

It follows that

$$
\begin{aligned}
&E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\
&\quad = \frac{1}{2}\begin{pmatrix} 1 \\ C_{i1} - C_{i2} \end{pmatrix}\{(\beta + \delta C_{i1} - \beta^* - \delta^* C_{i1}) - (\gamma + \delta C_{i2} + \psi - \delta^* C_{i2})\}.
\end{aligned}
$$

When $(\beta^*, \delta^*) = (\beta - \gamma - \psi, \delta)$ this expression equals

$$
\begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$

Define $\mathbf{Q}_i = (X_{i1} = 0, X_{i2} = 1, C_{i1}, C_{i2}, U_i)$. Under the causal diagram in Figure 1C) and model (5) with $g(\cdot) =$ identity we have that

$$
\begin{aligned}
E(Y_{i1}|\mathbf{Q}_i) &= \alpha_i + \delta C_{i1} \\
E(Y_{i2}|\mathbf{Q}_i) &= \alpha_i + \beta + \delta C_{i2} + \psi.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&E\{S_{i,ide}(\beta^*, \delta^*)|\mathbf{Q}_i\} \\
&\quad = \frac{1}{2}\begin{pmatrix} -1 \\ C_{i1} - C_{i2} \end{pmatrix}\{(\delta C_{i1} - \delta^* C_{i1}) - (\beta + \delta C_{i2} + \psi - \beta^* - \delta^* C_{i2})\}.
\end{aligned}
$$

When $(\beta^*, \delta^*) = (\beta + \psi, \delta)$ this expression equals

$$
\begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$