Supplementary material

Section S1 gives the proofs of the theorems.

Section S2 gives parametric examples for the completeness conditions.

Section S3 gives counterexamples and further discussion for the unidentifiable cases.

Section S4 provides details for the parametric estimation.

Section S5 provides details on the simulation studies.

Section S6 provides sensitivity analysis results on the NJCS.

S1 Proofs

S1.1 Proof of Theorem 1

The identification of $\mathbb{P}(Y = y \mid M = m, T = t, X = x)$ follows from

$$\mathbb{P}(Y = y \mid M = m, T = t, X = x) = \mathbb{P}(Y = y \mid R^M = 1, M = m, T = t, X = x).$$

We now focus on identifying $\mathbb{P}(M = m \mid T = t, X = x)$. Define

$$\begin{aligned}
\mathbb{P}_{my1|t,x} &= \mathbb{P}(M = m, Y = y, R^M = 1 \mid T = t, X = x),\\
\mathbb{P}_{+y0|t,x} &= \mathbb{P}(Y = y, R^M = 0 \mid T = t, X = x),\\
\zeta_{t,x}(m) &= \frac{\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}.
\end{aligned}$$

Since

$$\mathbb{P}_{my1|t,x} = \mathbb{P}(M = m, Y = y \mid T = t, X = x)\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x),$$

we have:

$$
\begin{aligned}
\mathbb{P}_{+y0|t,x} &= \int_{m\in\mathcal{M}} \mathbb{P}(M=m, Y=y, R^M=0 \mid T=t, X=x)\mathrm{d}m \\
&= \int_{m\in\mathcal{M}} \mathbb{P}(M=m, Y=y \mid T=t, X=x)\mathbb{P}(R^M=0 \mid M=m, T=t, X=x)\mathrm{d}m \\
&= \int_{m\in\mathcal{M}} \mathbb{P}_{my1|t,x}\frac{\mathbb{P}(R^M=0 \mid M=m, T=t, X=x)}{\mathbb{P}(R^M=1 \mid M=m, T=t, X=x)}\mathrm{d}m \\
&= \int_{m\in\mathcal{M}} \mathbb{P}_{my1|t,x}\zeta_{t,x}(m)\mathrm{d}m
\end{aligned}
$$

for each $y \in \mathcal{Y}$. The uniqueness of solutions $\zeta_{t,x}(m)$ requires that $\mathbb{P}(Y, M, R^M = 1 \mid T = t, X = x)$ is complete in Y for all t and x. For discrete M and discrete Y, the completeness assumption is equivalent to Rank $(\Theta_{tx}) = J$, where Θ_{tx} is a $J \times K$ matrix with $\mathbb{P}_{my1|t,x}$ as the (m, y)th element. For binary M, the rank condition further reduces to $M \not\perp\!\!\!\perp Y \mid (T, X)$, which is equivalent to the testable condition $M \not\perp\!\!\!\perp Y \mid (T, X, R^M = 1)$. For continuous M and continuous Y, the dimension of Y must be no smaller than that of M in general, as required by the completeness assumption.

We can subsequently identify $\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)$ once $\zeta_{t,x}(m)$ is identified. Then, the identification of $\mathbb{P}(M = m \mid T = t, X = x)$ follows from

$$
\mathbb{P}(M=m \mid T=t, X=x) = \frac{\mathbb{P}(M=m, R^M=1 \mid T=t, X=x)}{\mathbb{P}(R^M=1 \mid M=m, T=t, X=x)}.
$$

S1.2 Proof of Theorem 2

Theorem 2 (i)

The identification of $\mathbb{P}(Y = y \mid M = m, T = t, X = x)$ follows from

$$
\mathbb{P}(Y=y \mid M=m, T=t, X=x) = \mathbb{P}(Y=y \mid R^M=1, R^Y=1, M=m, T=t, X=x).
$$

We now focus on identifying $\mathbb{P}(M = m \mid T = t, X = x)$. Define

$$\begin{aligned} \mathbb{P}_{my11|t,x} &= \mathbb{P}(M = m, Y = y, R^M = 1, R^Y = 1 \mid T = t, X = x), \\ \mathbb{P}_{+y01|t,x} &= \mathbb{P}(Y = y, R^M = 0, R^Y = 1 \mid T = t, X = x), \\ \zeta_{t,x}(m) &= \frac{\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}. \end{aligned}$$

Since

$$\begin{aligned} \mathbb{P}_{my11|t,x} &= \mathbb{P}(M = m, Y = y \mid T = t, X = x) \\ &\quad \cdot \mathbb{P}(R^Y = 1 \mid R^M = 1, T = t, X = x)\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x), \end{aligned}$$

we have:

$$\begin{aligned} \mathbb{P}_{+y01|t,x} &= \int_{m \in \mathcal{M}} \mathbb{P}(M = m, Y = y, R^M = 0, R^Y = 1 \mid T = t, X = x)\mathrm{d}m \\ &= \int_{m \in \mathcal{M}} \mathbb{P}_{my11|t,x} \frac{\mathbb{P}(R^Y = 1 \mid R^M = 0, T = t, X = x)\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^Y = 1 \mid R^M = 1, T = t, X = x)\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}\mathrm{d}m \\ &= \frac{\mathbb{P}(R^Y = 1 \mid R^M = 0, T = t, X = x)}{\mathbb{P}(R^Y = 1 \mid R^M = 1, T = t, X = x)} \int_{m \in \mathcal{M}} \mathbb{P}_{my11|t,x}\zeta_{t,x}(m)\mathrm{d}m \end{aligned}$$

for each $y \in \mathcal{Y}$. The uniqueness of solutions $\zeta_{t,x}(m)$ requires that $\mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x)$ is complete in Y for all t and x. For discrete M and discrete Y, the completeness assumption is equivalent to Rank $(\Theta_{tx}) = J$, where Θ_{tx} is a $J \times K$ matrix with $\mathbb{P}_{my11|t,x}$ as the (m, y)th element. For binary M, the rank condition further reduces to $M \not\perp\!\!\!\perp Y \mid (T, X)$, which is equivalent to the testable condition $M \not\perp\!\!\!\perp Y \mid (T, X, R^M = 1, R^Y = 1)$. For continuous M and continuous Y, the dimension of Y must be no smaller than that of M in general, as required by the completeness assumption.

We can subsequently identify $\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)$ once $\zeta_{t,x}(m)$ is identified. Then, the identification of $\mathbb{P}(M = m \mid T = t, X = x)$ follows from

$$\mathbb{P}(M = m \mid T = t, X = x) = \frac{\mathbb{P}(M = m, R^M = 1 \mid T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}.$$

Theorem 2 *(ii)*

The identification of $\mathbb{P}(Y=y, M=m \mid T=t, X=x)$ follows from

$$\mathbb{P}(Y=y, M=m \mid T=t, X=x) = \mathbb{P}(Y=y, M=m \mid R^M=1, R^Y=1, T=t, X=x).$$

S1.3 Proof of Theorem 3

We discuss the identification of $\mathbb{P}(M=m, Y=y \mid T=t, X=x)$. Define

$$\begin{aligned}
\mathbb{P}_{my11|t,x} &= \mathbb{P}(M=m, Y=y, R^M=1, R^Y=1 \mid T=t, X=x), \\
\mathbb{P}_{+y01|t,x} &= \mathbb{P}(Y=y, R^M=0, R^Y=1 \mid T=t, X=x), \\
\mathbb{P}_{m+10|t,x} &= \mathbb{P}(M=m, R^M=1, R^Y=0 \mid T=t, X=x), \\
\zeta_{t,x}(m) &= \frac{\mathbb{P}(R^M=0 \mid M=m, T=t, X=x)}{\mathbb{P}(R^M=1 \mid M=m, T=t, X=x)}, \\
\eta_{t,x}(y) &= \frac{\mathbb{P}(R^Y=0 \mid Y=y, T=t, X=x)}{\mathbb{P}(R^Y=1 \mid Y=y, T=t, X=x)}.
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{P}_{my11|t,x} &= \mathbb{P}(M=m, Y=y \mid T=t, X=x) \\
&\quad \cdot \mathbb{P}(R^Y=1 \mid Y=y, T=t, X=x)\mathbb{P}(R^M=1 \mid M=m, T=t, X=x),
\end{aligned}$$

we have:

$$\begin{aligned}
\mathbb{P}_{+y01|t,x} &= \int_{m\in\mathcal{M}} \mathbb{P}(M=m, Y=y, R^M=0, R^Y=1 \mid T=t, X=x)\mathrm{d}m \\
&= \int_{m\in\mathcal{M}} \mathbb{P}_{my11|t,x}\frac{\mathbb{P}(R^M=0 \mid M=m, T=t, X=x)}{\mathbb{P}(R^M=1 \mid M=m, T=t, X=x)}\mathrm{d}m \\
&= \int_{m\in\mathcal{M}} \mathbb{P}_{my11|t,x}\zeta_{t,x}(m)\mathrm{d}m
\end{aligned}$$

for each $y \in \mathcal{Y}$, and

$$\begin{aligned}
\mathbb{P}_{m+10|t,x} &= \int_{y\in\mathcal{Y}} \mathbb{P}(M = m, Y = y, R^M = 1, R^Y = 0 \mid T = t, X = x)\mathrm{d}y \\
&= \int_{y\in\mathcal{Y}} \mathbb{P}_{my11|t,x} \frac{\mathbb{P}(R^Y = 0 \mid Y = y, T = t, X = x)}{\mathbb{P}(R^Y = 1 \mid Y = y, T = t, X = x)}\mathrm{d}y \\
&= \int_{y\in\mathcal{Y}} \mathbb{P}_{my11|t,x}\eta_{t,x}(y)\mathrm{d}y
\end{aligned}$$

for each $m \in \mathcal{M}$. The uniqueness of solutions $\zeta_{t,x}(m)$ requires that $\mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x)$ is complete in Y for all t and x, and the uniqueness of solutions $\eta_{t,x}(y)$ require that $\mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x)$ is complete in M for all t and x. For discrete M and discrete Y, the above completeness assumptions are equivalent to $J = K$ and Rank $(\Theta_{tx}) = J$, where Θ_{tx} is a $J \times J$ matrix with $\mathbb{P}_{my11|t,x}$ as the (m, y)th element. For binary M and binary Y, the rank condition reduces to $M \not\perp\!\!\!\perp Y \mid (T, X)$. For continuous M and continuous Y, the dimension of Y needs to be the same as the dimension of M in general as required by $\mathbb{P}(Y, M, R^Y = 1, R^M = 1 \mid T, X)$ being complete in M and being complete in Y.

We can subsequently identify $\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)$ and $\mathbb{P}(R^Y = 1 \mid Y = y, T = t, X = x)$ once $\zeta_{t,x}(m)$ and $\eta_{t,x}(y)$ are identified. Then, the identification of $\mathbb{P}(Y = y, M = m \mid T = t, X = x)$ follows from

$$\begin{aligned}
&\mathbb{P}(Y = y, M = m \mid T = t, X = x) \\
= \;&\frac{\mathbb{P}_{my11|t,x}}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)\mathbb{P}(R^Y = 1 \mid Y = y, T = t, X = x)}.
\end{aligned}$$

S1.4 Proof of Theorem 4

The identification of $\mathbb{P}(Y = y \mid M = m, T = t, X = x)$ follows from

$$\mathbb{P}(Y = y \mid M = m, T = t, X = x) = \mathbb{P}(Y = y \mid R^M = 1, R^Y = 1, M = m, T = t, X = x).$$

We now focus on identifying $\mathbb{P}(M = m \mid T = t, X = x)$. Define

$$\begin{aligned}
\mathbb{P}_{my^\dagger 1|t,x} &= \mathbb{P}(M = m, Y^\dagger = y^\dagger, R^M = 1 \mid T = t, X = x), \\
\mathbb{P}_{my11|t,x} &= \mathbb{P}(M = m, Y = y, R^M = 1, R^Y = 1 \mid T = t, X = x), \\
\mathbb{P}_{+y01|t,x} &= \mathbb{P}(Y = y, R^M = 0, R^Y = 1 \mid T = t, X = x), \\
\mathbb{P}_{m+10|t,x} &= \mathbb{P}(M = m, R^M = 1, R^Y = 0 \mid T = t, X = x), \\
\mathbb{P}_{++00|t,x} &= \mathbb{P}(R^M = 0, R^Y = 0 \mid T = t, X = x), \\
\zeta_{t,x}(m) &= \frac{\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}.
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{P}_{my11|t,x} &= \mathbb{P}(M = m, Y = y \mid T = t, X = x) \\
&\quad \cdot \mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)\mathbb{P}(R^Y = 1 \mid M = m, T = t, X = x),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}_{m+10|t,x} &= \mathbb{P}(M = m \mid T = t, X = x) \\
&\quad \cdot \mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)\mathbb{P}(R^Y = 0 \mid M = m, T = t, X = x),
\end{aligned}$$

we have:

$$\begin{aligned}
\mathbb{P}_{+y01|t,x} &= \int_{m \in \mathcal{M}} \mathbb{P}(M = m, Y = y, R^M = 0, R^Y = 1 \mid T = t, X = x)\mathrm{d}m \\
&= \int_{m \in \mathcal{M}} \mathbb{P}_{my11|t,x} \frac{\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}\mathrm{d}m \\
&= \int_{m \in \mathcal{M}} \mathbb{P}_{my11|t,x}\zeta_{t,x}(m)\mathrm{d}m
\end{aligned}$$

for each $y \in \mathcal{Y}$, and

$$\begin{aligned}
\mathbb{P}_{++00|t,x} &= \int_{m \in \mathcal{M}} \mathbb{P}(M = m, R^M = 0, R^Y = 0 \mid T = t, X = x)\mathrm{d}m \\
&= \int_{m \in \mathcal{M}} \mathbb{P}_{m+10|t,x} \frac{\mathbb{P}(R^M = 0 \mid M = m, T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}\mathrm{d}m \\
&= \int_{m \in \mathcal{M}} \mathbb{P}_{m+10|t,x}\zeta_{t,x}(m)\mathrm{d}m.
\end{aligned}$$

The uniqueness of solutions $\zeta_{t,x}(m)$ requires that $\mathbb{P}(M, Y^\dagger, R^M = 1 \mid T = t, X = x)$ is complete in $Y^\dagger$ for all t and x. For discrete M and discrete Y, the completeness assumption is equivalent to Rank $(\Theta_{tx}) = J$, where Θ_{tx} is a $J \times (K+1)$ matrix with $\mathbb{P}_{my11|t,x}$ as the (m, y)th element and $\mathbb{P}_{m+10|t,x}$ as the $(m, K+1)$th element. If it exists, the effect of M on R^Y provides one additional constraint to identify $\zeta_{t,x}(m)$. For binary M, the rank condition further reduces to $M \not\perp\!\!\!\perp Y^\dagger \mid (T, X)$, that is $M \not\perp\!\!\!\perp Y \mid (T, X)$ or $M \not\perp\!\!\!\perp R^Y \mid (T, X)$, which is equivalent to the testable condition $M \not\perp\!\!\!\perp Y \mid (T, X, R^M = 1, R^Y = 1)$ or $M \not\perp\!\!\!\perp R^Y \mid (T, X, R^M = 1)$. For continuous M and continuous Y, the dimension of $Y^\dagger$ needs to be no smaller than the dimension of M in general, as required by the completeness assumption.

We can subsequently identify $\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)$ once $\zeta_{t,x}(m)$ is identified. Then, the identification of $\mathbb{P}(M = m \mid T = t, X = x)$ follows from

$$\mathbb{P}(M = m \mid T = t, X = x) = \frac{\mathbb{P}(M = m, R^M = 1 \mid T = t, X = x)}{\mathbb{P}(R^M = 1 \mid M = m, T = t, X = x)}.$$

S2 Parametric examples

Theorem 2.2 in Newey and Powell (2003) presents the following result on the completeness of data distributions from an exponential family.

Result 1 *The distribution $\mathbb{P}(Y, M) = \psi(M)h(Y)\exp\{\lambda(Y)^{\mathrm{T}}\eta(M)\}$ is complete in Y if (i) $\psi(M) > 0$, (ii) the support of $\lambda(Y)$ is an open set, and (iii) the mapping $M \to \eta(M)$ is one to one.*

For illustration, we present examples of parametric models below that satisfy the corresponding completeness assumption for each of Theorems 1 to 4.

S2.1 An example for Theorem 1

Proposition 1 *For continuous Y, under a linear model*

$$Y = \beta_0 + \beta_m M + \beta_t T + \beta_{mt} M \cdot T + \beta_x X + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, if $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$, the distribution

$$\begin{aligned} & \mathbb{P}(Y, M, R^M = 1 \mid T = t, X = x) \\ = \;& \mathbb{P}(Y \mid M, T = t, X = x)\mathbb{P}(M, R^M = 1 \mid T = t, X = x) \\ = \;& \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(Y - \beta_0 - \beta_m M - \beta_t t - \beta_{mt} M \cdot t - \beta_x x)^2}{2\sigma^2}\right\} \mathbb{P}(M, R^M = 1 \mid T = t, X = x) \end{aligned}$$

is complete in Y for all t and x.

Proposition 1 follows from Result 1 with $\lambda(Y) = \sigma^{-2}(\beta_m + \beta_{mt} t)Y$ and $\eta(M) = M$.

S2.2 An example for Theorem 2

Proposition 2 *For continuous Y, under a linear model*

$$Y = \beta_0 + \beta_m M + \beta_t T + \beta_{mt} M \cdot T + \beta_x X + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, if $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$, the distribution

$$\begin{aligned} & \mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x) \\ = \;& \mathbb{P}(Y \mid M, T = t, X = x)\mathbb{P}(M, R^M = 1, R^Y = 1 \mid T = t, X = x) \\ = \;& \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(Y - \beta_0 - \beta_m M - \beta_t t - \beta_{mt} M \cdot t - \beta_x x)^2}{2\sigma^2}\right\} \\ & \cdot \mathbb{P}(M, R^M = 1, R^Y = 1 \mid T = t, X = x) \end{aligned}$$

is complete in Y for all t and x.

Proposition 2 follows from Result 1 with $\lambda(Y) = \sigma^{-2}(\beta_m + \beta_{mt} t)Y$ and $\eta(M) = M$.

S2.3 An example for Theorem 3

Proposition 3 *For continuous Y and continuous M, under a linear model*

$$Y = \beta_0 + \beta_m M + \beta_t T + \beta_{mt} M \cdot T + \beta_x X + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, if $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$, the distribution

$$\begin{aligned}
& \mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x) \\
= \; & \mathbb{P}(Y \mid M, T = t, X = x)\mathbb{P}(M, R^M = 1 \mid T = t, X = x)\mathbb{P}(R^Y = 1 \mid Y, T = t, X = x) \\
= \; & \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(Y - \beta_0 - \beta_m M - \beta_t t - \beta_{mt} M \cdot t - \beta_x x)^2}{2\sigma^2}\right\} \\
& \cdot\mathbb{P}(M, R^M = 1 \mid T = t, X = x)\mathbb{P}(R^Y = 1 \mid Y, T = t, X = x)
\end{aligned}$$

is complete in Y and is complete in M for all t and x.

Proposition 3 follows from Result 1 with $\lambda(Y) = \sigma^{-2}(\beta_m + \beta_{mt} t)Y$ and $\eta(M) = M$ and with $\lambda(M) = \sigma^{-2}(\beta_m + \beta_{mt} t)M$ and $\eta(Y) = Y$.

S2.4 Examples for Theorem 4

Proposition 4 *For continuous Y, under a linear model*

$$Y = \beta_0 + \beta_m M + \beta_t T + \beta_{mt} M \cdot T + \beta_x X + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, if $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$, the distribution

$$\begin{aligned}
& \mathbb{P}(Y, M, R^M = 1, R^Y = 1 \mid T = t, X = x) \\
= \; & \mathbb{P}(Y \mid M, T = t, X = x)\mathbb{P}(M, R^M = 1, R^Y = 1 \mid T = t, X = x) \\
= \; & \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(Y - \beta_0 - \beta_m M - \beta_t t - \beta_{mt} M \cdot t - \beta_x x)^2}{2\sigma^2}\right\} \\
& \cdot\mathbb{P}(M, R^M = 1, R^Y = 1 \mid T = t, X = x)
\end{aligned}$$

is complete in Y for all t and x.

Proposition 4 follows from Result 1 with $\lambda(Y) = \sigma^{-2}(\beta_m + \beta_{mt} t)Y$ and $\eta(M) = M$.

Proposition 5 *For binary M and binary R^Y, under a logistic regression model*

$$\text{logit } \mathbb{P}(R^Y = 1 \mid M, T, X) = \beta_0 + \beta_m M + \beta_t T + \beta_{mt} M \cdot T + \beta_x X,$$

if $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$, the distribution

$$\begin{aligned} & \mathbb{P}(M, R^M = 1, R^Y \mid T = t, X = x) \\ = & \mathbb{P}(R^Y \mid M, T = t, X = x)\mathbb{P}(M, R^M = 1 \mid T = t, X = x) \\ = & \frac{\exp\{R^Y(\beta_0 + \beta_m M + \beta_t t + \beta_{mt} M \cdot t + \beta_x x)\}}{1 + \exp(\beta_0 + \beta_m M + \beta_t t + \beta_{mt} M \cdot t + \beta_x x)} \mathbb{P}(M, R^M = 1 \mid T = t, X = x) \end{aligned}$$

is complete in R^Y for all t and x.

For binary M and binary R^Y, the completeness condition reduces to $M \not\perp\!\!\!\perp R^Y \mid (T = t, X = x)$ for all t and x, and therefore, Proposition 5 follows when $\beta_m \neq 0$ and $\beta_m + \beta_{mt} \neq 0$.

S3 The unidentifiable cases: counterexamples and conditions for identification

As discussed in section 2, identifying the NIE and NDE relies on identifying the joint distribution $\mathbb{P}(Y, M \mid T, X)$. Below, we first show that the observable data probabilities cannot uniquely determine $\mathbb{P}(Y, M \mid T, X)$ without further assumptions if R^Y depends on more than one of (R^M, Y, M) as the missingness mechanisms in Figure S1 or the completeness assumption is violated. We explain the reasons and provide concrete examples in subsection S3.1. We then show that the identification is plausible for these complex MNAR mechanisms by exploiting the information on a future outcome in subsection S3.2. To simplify the notation, all DAGs and probabilities below condition on T and X.

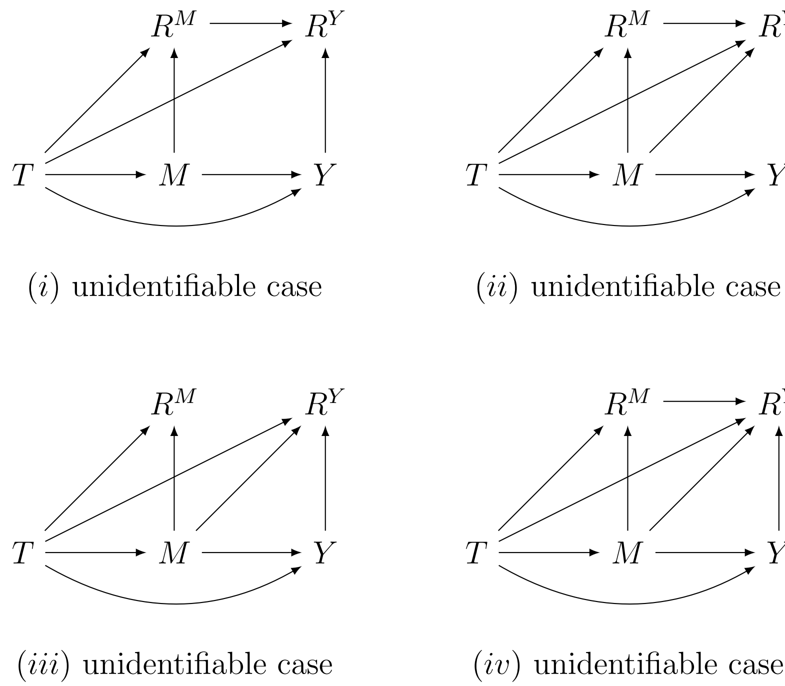


Figure S1: The DAGs in (i) to (iv) describe the unidentifiable missingness mechanisms when missingness exists in both the mediator and outcome. All DAGs condition on X and allow X to have directed arrows to all variables in the DAGs.

S3.1 Counterexamples

Define

$$\begin{aligned}
\mathbb{P}_{my11} &= \mathbb{P}(M = m, Y = y, R^M = 1, R^Y = 1), \\
\mathbb{P}_{+y01} &= \mathbb{P}(Y = y, R^M = 0, R^Y = 1), \\
\mathbb{P}_{m+10} &= \mathbb{P}(M = m, R^M = 1, R^Y = 0), \\
\mathbb{P}_{++00} &= \mathbb{P}(R^M = 0, R^Y = 0).
\end{aligned}$$

In (i) to (iv) below, we present examples where the identification cannot be achieved without further assumptions if R^Y depends on more than one of (R^M, Y, M) in a simple setup of a binary mediator M and a binary outcome Y. Based on the observable data probabilities, we can identify $\mathbb{P}_{my11}$, $\mathbb{P}_{+y01}$, $\mathbb{P}_{m+10}$ and $\mathbb{P}_{++00}$, and $\sum_{m=0}^{1}\sum_{y=0}^{1}\mathbb{P}_{my11} +$

$\sum_{y=0}^{1}\mathbb{P}_{+y01}+\sum_{m=0}^{1}\mathbb{P}_{m+10}+\mathbb{P}_{++00}=1$. In (v), we present an unidentifiable case when M has more categories than Y under Assumptions 1 to 4. In (vi), we present an unidentifiable case when Y has more categories than M under Assumption 3.

(i) **We present below an unidentified case when R^Y depends on both Y and R^M as described by Figure S2**

$$\begin{tikzcd}
R^M \arrow[r] & R^Y \\
M \arrow[u] \arrow[r] & Y \arrow[u]
\end{tikzcd}$$

Figure S2: R^Y depends on both Y and R^M

Consider the following observable data probabilities:

$$(\mathbb{P}_{1111},\mathbb{P}_{0111},\mathbb{P}_{1011},\mathbb{P}_{0011},\mathbb{P}_{+101},\mathbb{P}_{+001},\mathbb{P}_{1+10},\mathbb{P}_{0+10},\mathbb{P}_{++00})=\left(\frac{6}{20},\frac{2}{20},\frac{1}{20},\frac{1}{20},\frac{2}{20},\frac{1}{20},\frac{2}{20},\frac{1}{20},\frac{4}{20}\right).$$

The key to identify $\mathbb{P}(Y=y,M=m)$ is to identify both $\mathbb{P}(R^M=1\mid M=m)$ and $\mathbb{P}(R^Y=1\mid Y=y,R^M=1)$ in the following formula

$$\mathbb{P}(Y=y,M=m)=\frac{\mathbb{P}_{my11}}{\mathbb{P}(R^M=1\mid M=m)\mathbb{P}(R^Y=1\mid Y=y,R^M=1)}.$$

We now show that the identification of $\mathbb{P}(R^Y=1\mid Y=y,R^M=1)$ can be achieved. We have

$$\begin{aligned}\mathbb{P}_{m+10}&=\sum_{y\in\mathcal{Y}}\mathbb{P}(M=m,Y=y,R^M=1,R^Y=0)\\&=\sum_{y\in\mathcal{Y}}\mathbb{P}_{my11}\frac{\mathbb{P}(R^Y=0\mid Y=y,R^M=1)}{\mathbb{P}(R^Y=1\mid Y=y,R^M=1)}.\end{aligned}$$

By plugging in the two possible values for m and y in the above formula, we have

$$\mathbb{P}_{1+10}=\mathbb{P}_{1111}\frac{\mathbb{P}(R^Y=0\mid Y=1,R^M=1)}{\mathbb{P}(R^Y=1\mid Y=1,R^M=1)}+\mathbb{P}_{1011}\frac{\mathbb{P}(R^Y=0\mid Y=0,R^M=1)}{\mathbb{P}(R^Y=1\mid Y=0,R^M=1)},\tag{S1}$$

$$\mathbb{P}_{0+10} = \mathbb{P}_{0111}\frac{\mathbb{P}(R^Y = 0 \mid Y = 1, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 1)} + \mathbb{P}_{0011}\frac{\mathbb{P}(R^Y = 0 \mid Y = 0, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 1)}. \qquad \text{(S2)}$$

Therefore, $\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 1)$ can be identified by solving the linear equations (S1) and (S2). Based on the observable data probabilities, $\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 1) = \frac{4}{5}$ and $\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 1) = \frac{2}{3}$.

We now focus on the identifiability of $\mathbb{P}(R^M = 1 \mid M = m)$ and show that $\mathbb{P}(R^M = 1 \mid M = m)$ cannot be identified without further assumptions. We have

$$\begin{aligned}
\mathbb{P}_{+y01} &= \sum_{m\in\mathcal{M}} \mathbb{P}(M = m, Y = y, R^M = 0, R^Y = 1) \\
&= \sum_{m\in\mathcal{M}} \mathbb{P}_{my11}\frac{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 0)\mathbb{P}(R^M = 0 \mid M = m)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 1)\mathbb{P}(R^M = 1 \mid M = m)} \\
&= \frac{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 0)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 1)} \sum_{m\in\mathcal{M}} \mathbb{P}_{my11}\frac{\mathbb{P}(R^M = 0 \mid M = m)}{\mathbb{P}(R^M = 1 \mid M = m)},
\end{aligned}$$

and as a result,

$$\mathbb{P}_{+y01}\frac{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 0)} = \sum_{m\in\mathcal{M}} \mathbb{P}_{my11}\frac{\mathbb{P}(R^M = 0 \mid M = m)}{\mathbb{P}(R^M = 1 \mid M = m)}.$$

By plugging in the two possible values for m and y in the above formula, we have

$$\mathbb{P}_{+101}\frac{\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 0)} = \mathbb{P}_{1111}\frac{\mathbb{P}(R^M = 0 \mid M = 1)}{\mathbb{P}(R^M = 1 \mid M = 1)} + \mathbb{P}_{0111}\frac{\mathbb{P}(R^M = 0 \mid M = 0)}{\mathbb{P}(R^M = 1 \mid M = 0)},$$

$$\mathbb{P}_{+001}\frac{\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 0)} = \mathbb{P}_{1011}\frac{\mathbb{P}(R^M = 0 \mid M = 1)}{\mathbb{P}(R^M = 1 \mid M = 1)} + \mathbb{P}_{0011}\frac{\mathbb{P}(R^M = 0 \mid M = 0)}{\mathbb{P}(R^M = 1 \mid M = 0)}.$$

Since $\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 1)$ are identified from the previous step, the identifiability of $\mathbb{P}(R^M = 1 \mid M = m)$ depends on the identifiability of $\mathbb{P}(R^Y = 1 \mid Y = y, R^M = 0)$.

We have

$$\begin{aligned}
& \mathbb{P}(R^Y = 1 \mid Y = y, R^M = 0) \\
= \; & \frac{\mathbb{P}(Y = y, R^M = 0, R^Y = 1)}{\mathbb{P}(Y = y, R^M = 0)} \\
= \; & \frac{\mathbb{P}(Y = y, R^M = 0, R^Y = 1)}{\mathbb{P}(Y = y, R^M = 0, R^Y = 0) + \mathbb{P}(Y = y, R^M = 0, R^Y = 1)} \\
= \; & \frac{\mathbb{P}(Y = y, R^M = 0, R^Y = 1)}{\mathbb{P}(Y = y \mid R^M = 0, R^Y = 0)\mathbb{P}(R^M = 0, R^Y = 0) + \mathbb{P}(Y = y, R^M = 0, R^Y = 1)} \\
= \; & \frac{\mathbb{P}_{+y01}}{\mathbb{P}(Y = y \mid R^M = 0, R^Y = 0)\mathbb{P}_{++00} + \mathbb{P}_{+y01}}.
\end{aligned}$$

In the above expression, $\mathbb{P}_{+y01}$ and $\mathbb{P}_{++00}$ are known, but $\mathbb{P}(Y = y \mid R^Y = 0, R^M = 0)$ is not observable or identifiable. Different values of $\mathbb{P}(Y = y \mid R^Y = 0, R^M = 0)$ will result in different values of $\mathbb{P}(R^M = 1 \mid M = m)$, which in turn will give different values of $\mathbb{P}(Y = y, M = m)$. For example, let $\mathbb{P}(Y = 1 \mid R^Y = 0, R^M = 0) = \frac{5}{6}$, and the corresponding $\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 0)$ and $\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 0)$ equal $\frac{3}{8}$ and $\frac{3}{5}$, respectively. As a result, we have $\mathbb{P}(R^M = 1 \mid M = 1) = \frac{45}{68}$ and $\mathbb{P}(R^M = 1 \mid M = 0) = \frac{5}{8}$. Subsequently, we have $\mathbb{P}(Y = 1, M = 1) = \frac{17}{30}$, $\mathbb{P}(Y = 0, M = 1) = \frac{17}{150}$, $\mathbb{P}(Y = 1, M = 0) = \frac{1}{5}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{3}{25}$. Alternatively, let $\mathbb{P}(Y = 1 \mid R^Y = 0, R^M = 0) = \frac{7}{8}$, and the corresponding $\mathbb{P}(R^Y = 1 \mid Y = 1, R^M = 0)$ and $\mathbb{P}(R^Y = 1 \mid Y = 0, R^M = 0)$ equal $\frac{4}{11}$ and $\frac{2}{3}$, respectively. As a result, we have $\mathbb{P}(R^M = 1 \mid M = 1) = \frac{5}{8}$ and $\mathbb{P}(R^M = 1 \mid M = 0) = \frac{5}{7}$. Subsequently, we have $\mathbb{P}(Y = 1, M = 1) = \frac{3}{5}$, $\mathbb{P}(Y = 0, M = 1) = \frac{3}{25}$, $\mathbb{P}(Y = 1, M = 0) = \frac{7}{40}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{21}{200}$.

The two sets of values of $\mathbb{P}(Y = y, M = m)$ correspond to the same observable data probabilities: $(\mathbb{P}_{1111}, \mathbb{P}_{0111}, \mathbb{P}_{1011}, \mathbb{P}_{0011}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{++00})$, and therefore, $\mathbb{P}(Y = y, M = m)$ can not be uniquely identified without further assumptions.

This unidentifiable result does not contradict the conclusion in Li et al. (2023) discussing the identifiability of the self-censoring model under assumptions imposed by chain graphs instead of DAGs.

(ii) **We present below an unidentified case when R^Y depends on both M and R^M as described by Figure S3**

$$
\begin{tikzcd}
R^M \arrow[r] & R^Y \\
M \arrow[u] \arrow[ur] \arrow[r] & Y
\end{tikzcd}
$$

Figure S3: R^Y depends on both M and R^M

Consider the following observable data probabilities:

$$(\mathbb{P}_{1111}, \mathbb{P}_{0111}, \mathbb{P}_{1011}, \mathbb{P}_{0011}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{++00}) = \left(\frac{12}{40}, \frac{4}{40}, \frac{2}{40}, \frac{4}{40}, \frac{4}{40}, \frac{1}{40}, \frac{4}{40}, \frac{4}{40}, \frac{5}{40}\right).$$

Define

$$
\begin{aligned}
\mathbb{P}_M &= \mathbb{P}(M = 1), \\
\mathbb{P}_{Y1} &= \mathbb{P}(Y = 1 \mid M = 1), \\
\mathbb{P}_{Y0} &= \mathbb{P}(Y = 1 \mid M = 0), \\
\mathbb{P}_{R^M 1} &= \mathbb{P}(R^M = 1 \mid M = 1), \\
\mathbb{P}_{R^M 0} &= \mathbb{P}(R^M = 1 \mid M = 0), \\
\mathbb{P}_{R^Y 00} &= \mathbb{P}(R^Y = 1 \mid M = 0, R^M = 0), \\
\mathbb{P}_{R^Y 01} &= \mathbb{P}(R^Y = 1 \mid M = 0, R^M = 1), \\
\mathbb{P}_{R^Y 10} &= \mathbb{P}(R^Y = 1 \mid M = 1, R^M = 0), \\
\mathbb{P}_{R^Y 11} &= \mathbb{P}(R^Y = 1 \mid M = 1, R^M = 1).
\end{aligned}
$$

Below, we study the identifiability of the above 9 parameters describing the graphical model in Figure S3 based on the observable data probabilities. Although there are 9 observable data probabilities, the degree of freedom in the probabilities is only 8 given that they sum up to 1.

The following relationships between the observable data probabilities and the parameters hold,

$$\mathbb{P}_{1111} = \mathbb{P}_M \mathbb{P}_{Y1} \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 11}, \tag{S3}$$

$$\mathbb{P}_{1011} = \mathbb{P}_M (1 - \mathbb{P}_{Y1}) \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 11}, \tag{S4}$$

$$\mathbb{P}_{0111} = (1 - \mathbb{P}_M) \mathbb{P}_{Y0} \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 01}, \tag{S5}$$

$$\mathbb{P}_{0011} = (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y0}) \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 01}, \tag{S6}$$

$$\mathbb{P}_{1+10} = \mathbb{P}_M \mathbb{P}_{R^M 1} (1 - \mathbb{P}_{R^Y 11}), \tag{S7}$$

$$\mathbb{P}_{0+10} = (1 - \mathbb{P}_M) \mathbb{P}_{R^M 0} (1 - \mathbb{P}_{R^Y 01}), \tag{S8}$$

$$\mathbb{P}_{+101} = \mathbb{P}_M \mathbb{P}_{Y1} (1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 10} + (1 - \mathbb{P}_M) \mathbb{P}_{Y0} (1 - \mathbb{P}_{R^M 0}) \mathbb{P}_{R^Y 00}, \tag{S9}$$

$$\mathbb{P}_{+001} = \mathbb{P}_M (1 - \mathbb{P}_{Y1})(1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 10} + (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y0})(1 - \mathbb{P}_{R^M 0}) \mathbb{P}_{R^Y 00}. \tag{S10}$$

By solving the equations (S3) to (S8), we can identify the parameters $\mathbb{P}_{Y1}$, $\mathbb{P}_{Y0}$, $\mathbb{P}_{R^Y 11}$ and $\mathbb{P}_{R^Y 01}$:

$$\mathbb{P}_{Y1} = \frac{\mathbb{P}_{1111}}{\mathbb{P}_{1111} + \mathbb{P}_{1011}},$$
$$\mathbb{P}_{Y0} = \frac{\mathbb{P}_{0111}}{\mathbb{P}_{0111} + \mathbb{P}_{0011}},$$
$$\mathbb{P}_{R^Y 11} = \frac{\mathbb{P}_{1111} + \mathbb{P}_{1011}}{\mathbb{P}_{1111} + \mathbb{P}_{1011} + \mathbb{P}_{1+10}},$$
$$\mathbb{P}_{R^Y 01} = \frac{\mathbb{P}_{0111} + \mathbb{P}_{0011}}{\mathbb{P}_{0111} + \mathbb{P}_{0011} + \mathbb{P}_{0+10}}.$$

Based on the observable data probabilities, $\mathbb{P}_{Y1} = \frac{6}{7}$, $\mathbb{P}_{Y0} = \frac{1}{2}$, $\mathbb{P}_{R^Y 11} = \frac{7}{9}$ and $\mathbb{P}_{R^Y 01} = \frac{2}{3}$. In addition, we can identify the following products of parameters based on equations (S7) to (S10): $\mathbb{P}_M \mathbb{P}_{R^M 1}$, $(1 - \mathbb{P}_M) \mathbb{P}_{R^M 0}$, $\mathbb{P}_M (1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 10}$ and $(1 - \mathbb{P}_M)(1 - \mathbb{P}_{R^M 0}) \mathbb{P}_{R^Y 00}$. As a result, when $\mathbb{P}_M$ is known, one can solve for $\mathbb{P}_{R^M 1}$, $\mathbb{P}_{R^M 0}$, $\mathbb{P}_{R^Y 10}$ and $\mathbb{P}_{R^Y 00}$.

For example, let $\mathbb{P}_M = \frac{3}{5}$, we have $\mathbb{P}_{R^M 1} = \frac{3}{4}$, $\mathbb{P}_{R^M 0} = \frac{3}{4}$, $\mathbb{P}_{R^Y 10} = \frac{7}{10}$ and $\mathbb{P}_{R^Y 00} = \frac{1}{5}$. This set of parameter values gives us the following joint probabilities of M and Y as $\mathbb{P}(Y = 1, M = 1) = \frac{18}{35}$, $\mathbb{P}(Y = 0, M = 1) = \frac{3}{35}$, $\mathbb{P}(Y = 1, M = 0) = \frac{1}{5}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{1}{5}$. Alternatively, let $\mathbb{P}_M = \frac{13}{20}$, we have $\mathbb{P}_{R^M 1} = \frac{9}{13}$, $\mathbb{P}_{R^M 0} = \frac{6}{7}$, $\mathbb{P}_{R^Y 10} = \frac{21}{40}$ and $\mathbb{P}_{R^Y 00} = \frac{2}{5}$. This alternative set of parameter values gives us the following joint probabilities of M and Y as $\mathbb{P}(Y = 1, M = 1) = \frac{39}{70}$, $\mathbb{P}(Y = 0, M = 1) = \frac{13}{140}$, $\mathbb{P}(Y = 1, M = 0) = \frac{7}{40}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{7}{40}$.

The two sets of values of $\mathbb{P}(Y = y, M = m)$ correspond to the same observable data probabilities: $(\mathbb{P}_{1111}, \mathbb{P}_{0111}, \mathbb{P}_{1011}, \mathbb{P}_{0011}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{++00})$, and therefore, $\mathbb{P}(Y = y, M = m)$ can not be uniquely identified without further assumptions.

(iii) **We present below an unidentified case when R^Y depends on both Y and M as described by Figure S4**

$$\begin{tikzcd}
R^M & R^Y \\
M \arrow[u] \arrow[ur] \arrow[r] & Y \arrow[u]
\end{tikzcd}$$

Figure S4: R^Y depends on both Y and M

Consider the following probabilities from the observable data:

$$(\mathbb{P}_{1111}, \mathbb{P}_{0111}, \mathbb{P}_{1011}, \mathbb{P}_{0011}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{++00}) = \left(\frac{18}{96}, \frac{6}{96}, \frac{3}{96}, \frac{2}{96}, \frac{12}{96}, \frac{3}{96}, \frac{15}{96}, \frac{16}{96}, \frac{21}{96}\right).$$

Define

$$
\begin{aligned}
\mathbb{P}_M &= \mathbb{P}(M = 1),\\
\mathbb{P}_{Y1} &= \mathbb{P}(Y = 1 \mid M = 1),\\
\mathbb{P}_{Y0} &= \mathbb{P}(Y = 1 \mid M = 0),\\
\mathbb{P}_{R^M 1} &= \mathbb{P}(R^M = 1 \mid M = 1),\\
\mathbb{P}_{R^M 0} &= \mathbb{P}(R^M = 1 \mid M = 0),\\
\mathbb{P}_{R^Y 00} &= \mathbb{P}(R^Y = 1 \mid M = 0, Y = 0),\\
\mathbb{P}_{R^Y 01} &= \mathbb{P}(R^Y = 1 \mid M = 0, Y = 1),\\
\mathbb{P}_{R^Y 10} &= \mathbb{P}(R^Y = 1 \mid M = 1, Y = 0),\\
\mathbb{P}_{R^Y 11} &= \mathbb{P}(R^Y = 1 \mid M = 1, Y = 1).
\end{aligned}
$$

Below, we study the identifiability of the above 9 parameters describing the graphical model in Figure S4 based on the observable data probabilities. Although there are 9 observable data probabilities, the degree of freedom in the probabilities is only 8 given that they sum up to 1.

The following relationships between the observable data probabilities and the parameters hold,

$$\mathbb{P}_{1111} = \mathbb{P}_M \mathbb{P}_{Y1} \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 11}, \tag{S11}$$

$$\mathbb{P}_{1011} = \mathbb{P}_M (1 - \mathbb{P}_{Y1}) \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 10}, \tag{S12}$$

$$\mathbb{P}_{0111} = (1 - \mathbb{P}_M) \mathbb{P}_{Y0} \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 01}, \tag{S13}$$

$$\mathbb{P}_{0011} = (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y0}) \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 00}, \tag{S14}$$

$$\mathbb{P}_{1+10} = \mathbb{P}_M \mathbb{P}_{Y1} \mathbb{P}_{R^M 1}(1 - \mathbb{P}_{R^Y 11}) + \mathbb{P}_M (1 - \mathbb{P}_{Y1}) \mathbb{P}_{R^M 1}(1 - \mathbb{P}_{R^Y 10}), \tag{S15}$$

$$\mathbb{P}_{0+10} = (1 - \mathbb{P}_M) \mathbb{P}_{Y0} \mathbb{P}_{R^M 0}(1 - \mathbb{P}_{R^Y 01}) + (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y0}) \mathbb{P}_{R^M 0}(1 - \mathbb{P}_{R^Y 00}), \tag{S16}$$

$$\mathbb{P}_{+101} = \mathbb{P}_M \mathbb{P}_{Y1}(1 - \mathbb{P}_{R^M 1})\mathbb{P}_{R^Y 11} + (1 - \mathbb{P}_M)\mathbb{P}_{Y0}(1 - \mathbb{P}_{R^M 0})\mathbb{P}_{R^Y 01}, \tag{S17}$$

$$\mathbb{P}_{+001} = \mathbb{P}_M (1 - \mathbb{P}_{Y1})(1 - \mathbb{P}_{R^M 1})\mathbb{P}_{R^Y 10} + (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y0})(1 - \mathbb{P}_{R^M 0})\mathbb{P}_{R^Y 00}. \tag{S18}$$

By solving the equations (S11) to (S18), we can identify the parameters $\mathbb{P}_M$, $\mathbb{P}_{R^M 1}$ and $\mathbb{P}_{R^M 0}$. Given the observable data probabilities, $\mathbb{P}_M = \frac{1}{2}$, $\mathbb{P}_{R^M 1} = \frac{3}{4}$ and $\mathbb{P}_{R^M 0} = \frac{1}{2}$. However, $\mathbb{P}_{Y1}$, $\mathbb{P}_{Y0}$, $\mathbb{P}_{R^Y 11}$, $\mathbb{P}_{R^Y 10}$, $\mathbb{P}_{R^Y 01}$ and $\mathbb{P}_{R^Y 00}$ are not identifiable. For example, we can have $\mathbb{P}_{Y1} = \frac{3}{4}$, $\mathbb{P}_{Y0} = \frac{1}{2}$, $\mathbb{P}_{R^Y 11} = \frac{2}{3}$, $\mathbb{P}_{R^Y 10} = \frac{1}{3}$, $\mathbb{P}_{R^Y 01} = \frac{1}{2}$ and $\mathbb{P}_{R^Y 00} = \frac{1}{6}$, which in turn give us $\mathbb{P}(Y = 1, M = 1) = \frac{3}{8}$, $\mathbb{P}(Y = 1, M = 0) = \frac{1}{4}$, $\mathbb{P}(Y = 0, M = 1) = \frac{1}{8}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{1}{4}$. Alternatively, we can have $\mathbb{P}_{Y1} = \frac{2}{3}$, $\mathbb{P}_{Y0} = \frac{3}{4}$, $\mathbb{P}_{R^Y 11} = \frac{3}{4}$, $\mathbb{P}_{R^Y 10} = \frac{1}{4}$, $\mathbb{P}_{R^Y 01} = \frac{1}{3}$ and $\mathbb{P}_{R^Y 00} = \frac{1}{3}$, which in turn give us $\mathbb{P}(Y = 1, M = 1) = \frac{1}{3}$, $\mathbb{P}(Y = 1, M = 0) = \frac{3}{8}$, $\mathbb{P}(Y = 0, M = 1) = \frac{1}{6}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{1}{8}$.

The two sets of values of $\mathbb{P}(Y = y, M = m)$ correspond to the same observable data probabilities: $(\mathbb{P}_{1111}, \mathbb{P}_{0111}, \mathbb{P}_{1011}, \mathbb{P}_{0011}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{++00})$, and therefore, $\mathbb{P}(Y = y, M = m)$ can not be uniquely identified without further assumptions.

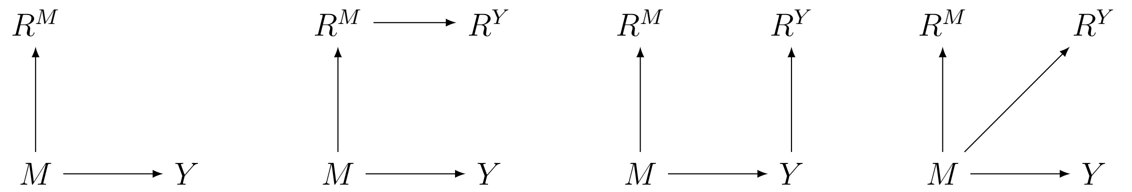
(iv) **We present below an unidentified case when R^Y depends on Y, M and R^M as described by Figure S5**

$$\begin{tikzcd}
R^M \arrow[r] & R^Y \\
M \arrow[u] \arrow[ur] \arrow[r] & Y \arrow[u]
\end{tikzcd}$$

Figure S5: R^Y depends on Y, M and R^M

The counterexamples presented in *(i)* to *(iii)* can all be viewed as special cases of the missingness mechanism described by Figure S5.

(v) **We present below an unidentified case when M has more categories than Y under Assumptions 1 to 4 as described by Figure S6**



(a) Assumption 1 (b) Assumption 2 (c) Assumption 3 (d) Assumption 4

Figure S6: M has more categories than Y under Assumptions 1 to 4

As an illustration, we present a counterexample for the missingness mechanism under Assumption 1 (a). The counterexample for (a) can be viewed as a special case for the missingness mechanism under Assumptions 2, 3, and 4 with $\mathbb{P}(R^Y = 1) = 1$.

For (d), when M has more categories than Y, we can still achieve identification if the rank condition holds as illustrated in Theorem 4. We provide a simulation result in section S5 showing that the identifiability of model parameters is improved under Assumption 4 compared to Assumption 1 when M has more categories than Y.

Define

$$\begin{aligned}
\mathbb{P}_{my1} &= \mathbb{P}(M = m, Y = y, R^M = 1), \\
\mathbb{P}_{+y0} &= \mathbb{P}(Y = y, R^M = 0).
\end{aligned}$$

Consider a binary outcome Y and a discrete M with three categories, denoted as 0, 1, and 2, respectively. Consider the following probabilities from the observable data:

$$(\mathbb{P}_{211}, \mathbb{P}_{111}, \mathbb{P}_{011}, \mathbb{P}_{201}, \mathbb{P}_{101}, \mathbb{P}_{001}, \mathbb{P}_{+10}, \mathbb{P}_{+00}) = \left(\frac{4}{96}, \frac{4}{96}, \frac{6}{96}, \frac{8}{96}, \frac{8}{96}, \frac{6}{96}, \frac{22}{96}, \frac{38}{96}\right).$$

Define

$$\mathbb{P}_{M^2} = \mathbb{P}(M = 2),$$

$$\mathbb{P}_{M^1} = \mathbb{P}(M = 1),$$

$$\mathbb{P}_{Y2} = \mathbb{P}(Y = 1 \mid M = 2),$$

$$\mathbb{P}_{Y1} = \mathbb{P}(Y = 1 \mid M = 1),$$

$$\mathbb{P}_{Y0} = \mathbb{P}(Y = 1 \mid M = 0),$$

$$\mathbb{P}_{R^M 2} = \mathbb{P}(R^M = 1 \mid M = 2),$$

$$\mathbb{P}_{R^M 1} = \mathbb{P}(R^M = 1 \mid M = 1),$$

$$\mathbb{P}_{R^M 0} = \mathbb{P}(R^M = 1 \mid M = 0).$$

The following relationships between the observable data probabilities and the parameters hold,

$$\mathbb{P}_{211} = \mathbb{P}_{M^2}\mathbb{P}_{Y2}\mathbb{P}_{R^M 2}, \tag{S19}$$

$$\mathbb{P}_{111} = \mathbb{P}_{M^1}\mathbb{P}_{Y1}\mathbb{P}_{R^M 1}, \tag{S20}$$

$$\mathbb{P}_{011} = (1 - \mathbb{P}_{M^2} - \mathbb{P}_{M^1})\mathbb{P}_{Y0}\mathbb{P}_{R^M 0}, \tag{S21}$$

$$\mathbb{P}_{201} = \mathbb{P}_{M^2}(1 - \mathbb{P}_{Y2})\mathbb{P}_{R^M 2}, \tag{S22}$$

$$\mathbb{P}_{101} = \mathbb{P}_{M^1}(1 - \mathbb{P}_{Y1})\mathbb{P}_{R^M 1}, \tag{S23}$$

$$\mathbb{P}_{001} = (1 - \mathbb{P}_{M^2} - \mathbb{P}_{M^1})(1 - \mathbb{P}_{Y0})\mathbb{P}_{R^M 0}, \tag{S24}$$

$$\begin{aligned}\mathbb{P}_{+10} = {} & \mathbb{P}_{M^2}\mathbb{P}_{Y2}(1 - \mathbb{P}_{R^M 2}) + \mathbb{P}_{M^1}\mathbb{P}_{Y1}(1 - \mathbb{P}_{R^M 1}) \\ & + (1 - \mathbb{P}_{M^2} - \mathbb{P}_{M^1})\mathbb{P}_{Y0}(1 - \mathbb{P}_{R^M 0}),\end{aligned} \tag{S25}$$

$$\begin{aligned}\mathbb{P}_{+00} = {} & \mathbb{P}_{M^2}(1 - \mathbb{P}_{Y2})(1 - \mathbb{P}_{R^M 2}) + \mathbb{P}_{M^1}(1 - \mathbb{P}_{Y1})(1 - \mathbb{P}_{R^M 1}) \\ & + (1 - \mathbb{P}_{M^2} - \mathbb{P}_{M^1})(1 - \mathbb{P}_{Y0})(1 - \mathbb{P}_{R^M 0}).\end{aligned} \tag{S26}$$

By solving the equations (S19) to (S26), we can identify the parameters $\mathbb{P}_{Y2}$, $\mathbb{P}_{Y1}$ and

$\mathbb{P}_{Y0}$. Given the observable data probabilities, $\mathbb{P}_{Y2} = \frac{1}{3}$, $\mathbb{P}_{Y1} = \frac{1}{3}$ and $\mathbb{P}_{Y0} = \frac{1}{2}$. However, $\mathbb{P}_{M^2}$, $\mathbb{P}_{M^1}$, $\mathbb{P}_{R^M 2}$, $\mathbb{P}_{R^M 1}$ and $\mathbb{P}_{R^M 0}$ are not identifiable. For example, we can have $\mathbb{P}_{M^2} = \frac{1}{4}$, $\mathbb{P}_{M^1} = \frac{1}{2}$, $\mathbb{P}_{R^M 2} = \frac{1}{2}$, $\mathbb{P}_{R^M 1} = \frac{1}{4}$ and $\mathbb{P}_{R^M 0} = \frac{1}{2}$, which in turn give us $\mathbb{P}(Y = 1, M = 2) = \frac{1}{12}$, $\mathbb{P}(Y = 1, M = 1) = \frac{1}{6}$, $\mathbb{P}(Y = 1, M = 0) = \frac{1}{8}$, $\mathbb{P}(Y = 0, M = 2) = \frac{1}{6}$, $\mathbb{P}(Y = 0, M = 1) = \frac{1}{3}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{1}{8}$. Alternatively, we can have $\mathbb{P}_{M^2} = \frac{3}{8}$, $\mathbb{P}_{M^1} = \frac{3}{8}$, $\mathbb{P}_{R^M 2} = \frac{1}{3}$, $\mathbb{P}_{R^M 1} = \frac{1}{3}$ and $\mathbb{P}_{R^M 0} = \frac{1}{2}$, which in turn give us $\mathbb{P}(Y = 1, M = 2) = \frac{1}{8}$, $\mathbb{P}(Y = 1, M = 1) = \frac{1}{8}$, $\mathbb{P}(Y = 1, M = 0) = \frac{1}{8}$, $\mathbb{P}(Y = 0, M = 2) = \frac{1}{4}$, $\mathbb{P}(Y = 0, M = 1) = \frac{1}{4}$ and $\mathbb{P}(Y = 0, M = 0) = \frac{1}{8}$.

The two sets of values of $\mathbb{P}(Y = y, M = m)$ correspond to the same observable data probabilities: $(\mathbb{P}_{211}, \mathbb{P}_{111}, \mathbb{P}_{011}, \mathbb{P}_{201}, \mathbb{P}_{101}, \mathbb{P}_{001}, \mathbb{P}_{+10}, \mathbb{P}_{+00})$, and therefore, $\mathbb{P}(Y = y, M = m)$ can not be uniquely identified without further assumptions.

(vi) **We present below an unidentified case when Y has more categories than M under Assumption 3 as described by Figure S7**

$$\begin{tikzcd}
R^M & R^Y \\
M \arrow[u] \arrow[r] & Y \arrow[u]
\end{tikzcd}$$

Figure S7: Y has more categories than M under Assumption 3

Consider a binary M and an outcome Y with three categories, denoted as 0, 1, and 2, respectively. Consider the following probabilities from the observable data:

$$\begin{aligned}
& (\mathbb{P}_{1211}, \mathbb{P}_{1111}, \mathbb{P}_{1011}, \mathbb{P}_{0211}, \mathbb{P}_{0111}, \mathbb{P}_{0011}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{+201}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{++00}) \\
= & \left(\frac{180}{1440}, \frac{60}{1440}, \frac{15}{1440}, \frac{120}{1440}, \frac{30}{1440}, \frac{15}{1440}, \frac{285}{1440}, \frac{195}{1440}, \frac{180}{1440}, \frac{50}{1440}, \frac{20}{1440}, \frac{290}{1440}\right).
\end{aligned}$$

Define

$$
\begin{aligned}
&\mathbb{P}_M = \mathbb{P}(M = 1),\\
&\mathbb{P}_{Y^2 1} = \mathbb{P}(Y = 2 \mid M = 1),\\
&\mathbb{P}_{Y^1 1} = \mathbb{P}(Y = 1 \mid M = 1),\\
&\mathbb{P}_{Y^2 0} = \mathbb{P}(Y = 2 \mid M = 0),\\
&\mathbb{P}_{Y^1 0} = \mathbb{P}(Y = 1 \mid M = 0),\\
&\mathbb{P}_{R^M 1} = \mathbb{P}(R^M = 1 \mid M = 1),\\
&\mathbb{P}_{R^M 0} = \mathbb{P}(R^M = 1 \mid M = 0),\\
&\mathbb{P}_{R^Y 2} = \mathbb{P}(R^Y = 1 \mid Y = 2),\\
&\mathbb{P}_{R^Y 1} = \mathbb{P}(R^Y = 1 \mid Y = 1),\\
&\mathbb{P}_{R^Y 0} = \mathbb{P}(R^Y = 1 \mid Y = 0).
\end{aligned}
$$

The following relationships between the observable data probabilities and the parameters hold,

$$\mathbb{P}_{1211} = \mathbb{P}_M \mathbb{P}_{Y^2 1} \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 2}, \tag{S27}$$

$$\mathbb{P}_{1111} = \mathbb{P}_M \mathbb{P}_{Y^1 1} \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 1}, \tag{S28}$$

$$\mathbb{P}_{1011} = \mathbb{P}_M (1 - \mathbb{P}_{Y^2 1} - \mathbb{P}_{Y^1 1}) \mathbb{P}_{R^M 1} \mathbb{P}_{R^Y 0}, \tag{S29}$$

$$\mathbb{P}_{0211} = (1 - \mathbb{P}_M) \mathbb{P}_{Y^2 0} \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 2}, \tag{S30}$$

$$\mathbb{P}_{0111} = (1 - \mathbb{P}_M) \mathbb{P}_{Y^1 0} \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 1}, \tag{S31}$$

$$\mathbb{P}_{0011} = (1 - \mathbb{P}_M)(1 - \mathbb{P}_{Y^2 0} - \mathbb{P}_{Y^1 0}) \mathbb{P}_{R^M 0} \mathbb{P}_{R^Y 0}, \tag{S32}$$

$$\mathbb{P}_{+201} = \mathbb{P}_M \mathbb{P}_{Y^2 1} (1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 2} + (1 - \mathbb{P}_M) \mathbb{P}_{Y^2 0} (1 - \mathbb{P}_{R^M 0}) \mathbb{P}_{R^Y 2}, \tag{S33}$$

$$\mathbb{P}_{+101} = \mathbb{P}_M \mathbb{P}_{Y^1 1} (1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 1} + (1 - \mathbb{P}_M) \mathbb{P}_{Y^1 0} (1 - \mathbb{P}_{R^M 0}) \mathbb{P}_{R^Y 1}, \tag{S34}$$

$$\mathbb{P}_{+001} = \mathbb{P}_M (1 - \mathbb{P}_{Y^2 1} - \mathbb{P}_{Y^1 1})(1 - \mathbb{P}_{R^M 1}) \mathbb{P}_{R^Y 0}$$

$$+ (1-\mathbb{P}_M)(1-\mathbb{P}_{Y^2 0}-\mathbb{P}_{Y^1 0})(1-\mathbb{P}_{R^M 0})\mathbb{P}_{R^Y 0}, \tag{S35}$$

$$\mathbb{P}_{1+10} = \mathbb{P}_M\mathbb{P}_{Y^2 1}\mathbb{P}_{R^M 1}(1-\mathbb{P}_{R^Y 2}) + \mathbb{P}_M\mathbb{P}_{Y^1 1}\mathbb{P}_{R^M 1}(1-\mathbb{P}_{R^Y 1})$$

$$+ \mathbb{P}_M(1-\mathbb{P}_{Y^2 1}-\mathbb{P}_{Y^1 1})\mathbb{P}_{R^M 1}(1-\mathbb{P}_{R^Y 0}), \tag{S36}$$

$$\mathbb{P}_{0+10} = (1-\mathbb{P}_M)\mathbb{P}_{Y^2 0}\mathbb{P}_{R^M 0}(1-\mathbb{P}_{R^Y 2}) + (1-\mathbb{P}_M)\mathbb{P}_{Y^1 0}\mathbb{P}_{R^M 0}(1-\mathbb{P}_{R^Y 1})$$

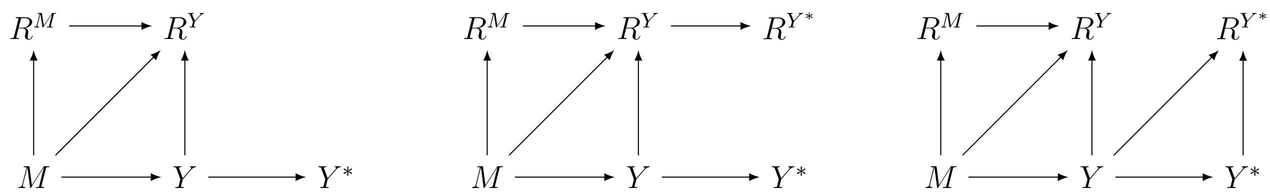
$$+ (1-\mathbb{P}_M)(1-\mathbb{P}_{Y^2 0}-\mathbb{P}_{Y^1 0})\mathbb{P}_{R^M 0}(1-\mathbb{P}_{R^Y 0}). \tag{S37}$$

By solving the equations (S27) to (S37), we can identify the parameters $\mathbb{P}_M$, $\mathbb{P}_{R^M 1}$ and $\mathbb{P}_{R^M 0}$. Given the observable data probabilities, $\mathbb{P}_M = \frac{1}{2}$, $\mathbb{P}_{R^M 1} = \frac{3}{4}$ and $\mathbb{P}_{R^M 0} = \frac{1}{2}$. However, $\mathbb{P}_{Y^2 1}$, $\mathbb{P}_{Y^1 1}$, $\mathbb{P}_{Y^2 0}$, $\mathbb{P}_{Y^1 0}$, $\mathbb{P}_{R^Y 2}$, $\mathbb{P}_{R^Y 1}$ and $\mathbb{P}_{R^Y 0}$ are not identifiable. For example, we can have $\mathbb{P}_{Y^2 1} = \frac{1}{2}$, $\mathbb{P}_{Y^1 1} = \frac{1}{3}$, $\mathbb{P}_{Y^2 0} = \frac{1}{2}$, $\mathbb{P}_{Y^1 0} = \frac{1}{4}$, $\mathbb{P}_{R^Y 2} = \frac{2}{3}$, $\mathbb{P}_{R^Y 1} = \frac{1}{3}$ and $\mathbb{P}_{R^Y 0} = \frac{1}{6}$, which in turn give us $\mathbb{P}(Y=2, M=1) = \frac{1}{4}$, $\mathbb{P}(Y=1, M=1) = \frac{1}{6}$, $\mathbb{P}(Y=0, M=1) = \frac{1}{12}$, $\mathbb{P}(Y=2, M=0) = \frac{1}{4}$, $\mathbb{P}(Y=1, M=0) = \frac{1}{8}$ and $\mathbb{P}(Y=0, M=0) = \frac{1}{8}$. Alternatively, we can have $\mathbb{P}_{Y^2 1} = \frac{5}{8}$, $\mathbb{P}_{Y^1 1} = \frac{1}{4}$, $\mathbb{P}_{Y^2 0} = \frac{5}{8}$, $\mathbb{P}_{Y^1 0} = \frac{3}{16}$, $\mathbb{P}_{R^Y 2} = \frac{8}{15}$, $\mathbb{P}_{R^Y 1} = \frac{4}{9}$ and $\mathbb{P}_{R^Y 0} = \frac{2}{9}$, which in turn give us $\mathbb{P}(Y=2, M=1) = \frac{5}{16}$, $\mathbb{P}(Y=1, M=1) = \frac{1}{8}$, $\mathbb{P}(Y=0, M=1) = \frac{1}{16}$, $\mathbb{P}(Y=2, M=0) = \frac{5}{16}$, $\mathbb{P}(Y=1, M=0) = \frac{3}{32}$ and $\mathbb{P}(Y=0, M=0) = \frac{3}{32}$.

The two sets of values of $\mathbb{P}(Y=y, M=m)$ correspond to the same observable data probabilities: $(\mathbb{P}_{1211}, \mathbb{P}_{1111}, \mathbb{P}_{1011}, \mathbb{P}_{0211}, \mathbb{P}_{0111}, \mathbb{P}_{0011}, \mathbb{P}_{1+10}, \mathbb{P}_{0+10}, \mathbb{P}_{+201}, \mathbb{P}_{+101}, \mathbb{P}_{+001}, \mathbb{P}_{++00})$, and therefore, $\mathbb{P}(Y=y, M=m)$ can not be uniquely identified without further assumptions.

S3.2 Improve identifiability with a future outcome

Ma et al. (2003) proposed to enhance the identifiability of an unidentifiable model by incorporating a future outcome. We show similar ideas apply to our setting. We use Y^* to denote the future outcome with $\mathcal{Y}^*$ denoting its support, and let R^{Y^*} be the response indicator for Y^* such that $R^{Y^*} = 1$ if Y^* is observed and $R^{Y^*} = 0$ otherwise. We provide some scenarios where the identification of $\mathbb{P}(Y = y, M = m)$ is plausible under the unidentifiable case (iv) by exploiting the information on a future outcome as described in Figure S8 (a) to (c). The same results can apply to the reduced unidentifiable cases (i) to (iii). To simplify notation, all DAGs and discussions in this subsection condition on T and X and allow T and X to have directed arrows to all variables in the DAGs.



(a) (b) (c)

Figure S8: The DAGs in (a) to (c) describe the unidentifiable case (iv) that can become identifiable with a fully observed Y^* or Y^* subject to missingness.

According to the structures of the DAGs in Figure S8, the identification of $\mathbb{P}(Y = y)$ in (a) to (c) can be established based on the theoretical results presented in the main paper under some completeness assumptions. Specifically, let Y, Y^*, R^Y, R^{Y^*} play the roles as M, Y, R^M, R^Y, respectively, the identification of $\mathbb{P}(Y = y)$ in (a) to (c) can be achieved following the identification of $\mathbb{P}(M = m)$ in the proofs of Theorems 1 to 4.

In all of the DAGs in Figure S8, $R^M \perp\!\!\!\perp Y \mid M$, we can identify $\mathbb{P}(Y = y \mid M = m)$ if $\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)$ is identifiable. We have

$$\begin{aligned}
\mathbb{P}(Y = y \mid M = m) &= \mathbb{P}(Y = y \mid M = m, R^M = 1) \\
&= \frac{\mathbb{P}(Y = y, R^Y = 1 \mid M = m, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)}. \qquad \text{(S38)}
\end{aligned}$$

In the expression (S38), $\mathbb{P}(Y = y, R^Y = 1 \mid M = m, R^M = 1)$ is observable. Below, we show that the identification of $\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)$ can be achieved with a fully observed Y^* or Y^* subject to missingness according to (a) to (c). Define

$$\begin{aligned}
\mathbb{P}_{y^*y1|m1} &= \mathbb{P}(Y^* = y^*, Y = y, R^Y = 1 \mid M = m, R^M = 1), \\
\mathbb{P}_{y^*+0|m1} &= \mathbb{P}(Y^* = y^*, R^Y = 0 \mid M = m, R^M = 1), \\
\mathbb{P}_{y^*y11|m1} &= \mathbb{P}(Y^* = y^*, Y = y, R^{Y^*} = 1, R^Y = 1 \mid M = m, R^M = 1), \\
\mathbb{P}_{y^*+10|m1} &= \mathbb{P}(Y^* = y^*, R^{Y^*} = 1, R^Y = 0 \mid M = m, R^M = 1), \\
\mathbb{P}_{+y01|m1} &= \mathbb{P}(Y = y, R^{Y^*} = 0, R^Y = 1 \mid M = m, R^M = 1), \\
\mathbb{P}_{++00|m1} &= \mathbb{P}(R^{Y^*} = 0, R^Y = 0 \mid M = m, R^M = 1).
\end{aligned}$$

In (a), we have

$$\mathbb{P}_{y^*y1|m1} = \mathbb{P}(Y^* = y^* \mid Y = y)\mathbb{P}(Y = y \mid M = m)\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1),$$

and therefore, for each $y^* \in \mathcal{Y}^*$,

$$\begin{aligned}
\mathbb{P}_{y^*+0|m1} &= \int_{y \in \mathcal{Y}} \mathbb{P}(Y^* = y^*, Y = y, R^Y = 0 \mid M = m, R^M = 1)\mathrm{d}y \\
&= \int_{y \in \mathcal{Y}} \mathbb{P}_{y^*y1|m1} \frac{\mathbb{P}(R^Y = 0 \mid Y = y, M = m, R^M = 1)}{\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)}\mathrm{d}y.
\end{aligned}$$

The uniqueness of solutions $\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)$ in (a) requires that $\mathbb{P}(Y^*, Y, R^Y = 1 \mid M = m, R^M = 1)$ is complete in Y^* for all m.

In (b), we have

$$\begin{aligned}
\mathbb{P}_{y^*y11|m1} &= \mathbb{P}(Y^* = y^* \mid Y = y)\mathbb{P}(Y = y \mid M = m) \\
&\quad \cdot \mathbb{P}(R^{Y^*} = 1 \mid R^Y = 1)\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1),
\end{aligned}$$

and therefore, for each $y^* \in \mathcal{Y}^*$,

$$\begin{aligned}
\mathbb{P}_{y^*+10|m1} &= \int_{y \in \mathcal{Y}} \mathbb{P}(Y^* = y^*, Y = y, R^{Y^*} = 1, R^Y = 0 \mid M = m, R^M = 1)\mathrm{d}y \\
&= \int_{y \in \mathcal{Y}} \mathbb{P}_{y^*y11|m1} \frac{\mathbb{P}(R^Y = 0 \mid Y = y, M = m, R^M = 1)\mathbb{P}(R^{Y^*} = 1 \mid R^Y = 0)}{\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)\mathbb{P}(R^{Y^*} = 1 \mid R^Y = 1)}\mathrm{d}y.
\end{aligned}$$

The uniqueness of solutions $\mathbb{P}(R^Y = 1 \mid Y = y, M = m, R^M = 1)$ in (b) requires that

$\mathbb{P}(Y^*, Y, R^{Y^*}=1, R^Y=1 \mid M=m, R^M=1)$ is complete in Y^* for all m.

In (c), we have

$$\begin{aligned}\mathbb{P}_{y^*y11|m1} &= \mathbb{P}(Y^*=y^* \mid Y=y)\mathbb{P}(Y=y \mid M=m) \\ &\quad \cdot \mathbb{P}(R^{Y^*}=1 \mid Y=y, Y^*=y^*)\mathbb{P}(R^Y=1 \mid Y=y, M=m, R^M=1),\end{aligned}$$

and

$$\mathbb{P}_{+y01|m1} = \mathbb{P}(Y=y \mid M=m)\mathbb{P}(R^{Y^*}=0 \mid Y=y)\mathbb{P}(R^Y=1 \mid Y=y, M=m, R^M=1).$$

Therefore, we have

$$\begin{aligned}\mathbb{P}_{y^*+10|m1} &= \int_{y\in\mathcal{Y}} \mathbb{P}(Y^*=y^*, Y=y, R^{Y^*}=1, R^Y=0 \mid M=m, R^M=1)\mathrm{d}y \\ &= \int_{y\in\mathcal{Y}} \mathbb{P}_{y^*y11|m1}\frac{\mathbb{P}(R^Y=0 \mid Y=y, M=m, R^M=1)}{\mathbb{P}(R^Y=1 \mid Y=y, M=m, R^M=1)}\mathrm{d}y\end{aligned}$$

for each $y^* \in \mathcal{Y}^*$, and

$$\begin{aligned}\mathbb{P}_{++00|m1} &= \int_{y\in\mathcal{Y}} \mathbb{P}(Y=y, R^{Y^*}=0, R^Y=0 \mid M=m, R^M=1)\mathrm{d}y \\ &= \int_{y\in\mathcal{Y}} \mathbb{P}_{+y01|m1}\frac{\mathbb{P}(R^Y=0 \mid Y=y, M=m, R^M=1)}{\mathbb{P}(R^Y=1 \mid Y=y, M=m, R^M=1)}\mathrm{d}y.\end{aligned}$$

Further define a random vector $Y^{*\dagger} = (Y^* \cdot R^{Y^*}, R^{Y^*})$. The uniqueness of solutions $\mathbb{P}(R^Y = 1 \mid Y=y, M=m, R^M=1)$ in (c) requires that $\mathbb{P}(Y^{*\dagger}, Y, R^Y=1 \mid M=m, R^M=1)$ is complete in $Y^{*\dagger}$ for all m.

So far, we have shown that the identification of $\mathbb{P}(Y=y)$ and $\mathbb{P}(Y=y \mid M=m)$ can be established in (a) to (c) under some completeness assumptions. Subsequently, if the joint distribution $\mathbb{P}(Y, M)$ is complete in Y, we can identify $\mathbb{P}(M=m)$ by solving the following linear equations:

$$\mathbb{P}(Y=y) = \int_{m\in\mathcal{M}} \mathbb{P}(Y=y \mid M=m)\mathbb{P}(M=m)\mathrm{d}m$$

for each $y \in \mathcal{Y}$.

Therefore, the identification of $\mathbb{P}(Y=y, M=m)$ can be achieved in (a) to (c) by exploiting the information on a future outcome. Since all the probabilities and statements involved conditions on T and X, the corresponding completeness conditions need to hold for all t and x.

S4 Details for the parametric estimation

For illustration, we describe the parametric methods in the scenarios considered in Theorem 1 when missingness exists only in the mediator. The likelihoods below are implicitly conditional on T and X to simplify the notation.

Under Assumption 1, the log of the complete-data likelihood is

$$\begin{aligned}\ell_c(\theta) = \sum_{i=1}^{n} \log\ \mathbb{P}(Y_i = y_i \mid M_i = m_i, T_i = t_i, X_i = x_i) + \log\ \mathbb{P}(M_i = m_i \mid T_i = t_i, X_i = x_i) \\ + \log\ \mathbb{P}(R_i^M = r_i^M \mid M_i = m_i, T_i = t_i, X_i = x_i).\end{aligned}$$

Under Assumption 1, the observed-data likelihood is

$$\begin{aligned}L_{obs}(\theta) = &\prod_{\{i:R_i^M=1\}} \mathbb{P}(Y_i = y_i \mid M_i = m_i, T_i = t_i, X_i = x_i)\mathbb{P}(M_i = m_i \mid T_i = t_i, X_i = x_i) \\ &\cdot \mathbb{P}(R_i^M = 1 \mid M_i = m_i, T_i = t_i, X_i = x_i) \\ &\cdot \prod_{\{i:R_i^M=0\}} \int_{\mathcal{M}} \mathbb{P}(Y_i = y_i \mid M_i = m, T_i = t_i, X_i = x_i)\mathbb{P}(M_i = m \mid T_i = t_i, X_i = x_i) \\ &\cdot \mathbb{P}(R_i^M = 0 \mid M_i = m, T_i = t_i, X_i = x_i)\ \mathrm{d}m.\end{aligned}$$

When M is discrete, the integral involved in the above expression is reduced to summation. Since the value of M is missing for some subjects, we implement the expectation-maximization algorithm (Dempster et al., 1977) to obtain the maximum likelihood estimates by treating the missing M as a latent variable. Specifically, in the E-step, we find the conditional expectation of complete-data log-likelihood by calculating the conditional expectation of M for subjects with missing M. For example, if M is binary,

$$E\{I(M_i = m) \mid Y_i, R_i^M = 0, T_i, X_i; \theta^{(t)}\} = \frac{\mathbb{P}(Y_i, M_i = m, R_i^M = 0 \mid T_i, X_i)}{\sum_{m=0,1} \mathbb{P}(Y_i, M_i = m, R_i^M = 0 \mid T_i, X_i)}.$$

When M is continuous, the conditional expectation of complete-data log-likelihood may be complicated to calculate. Therefore, we apply fractional imputation (Kim, 2011) using the idea of importance sampling and weighting method to approximate the conditional expectation. Specifically, we generate the fractionally imputed data $m_i^{(1)}, \ldots, m_i^{(S)}$ from a proposed distribution $h(M_i \mid T_i, X_i)$ for subjects with missing M. Then, we compute the fractional weight for each imputed observation. The Monte Carlo approximation of the conditional expectation becomes more accurate when S is large:

$$E\{\ell_{ci}(M_i = m; \theta) | Y_i, R_i^M = 0, T_i, X_i; \theta^{(t)}\} \approx \sum_{j=1}^{S} \ell_{ci}(M_i = m_i^{(j)}; \theta)\hat{w}(m_i^{(j)}),$$

where

$$\hat{w}(m_i^{(j)}) \propto \frac{\mathbb{P}(Y_i, M_i = m_i^{(j)}, R_i^M = 0 \mid T_i, X_i)}{h(M_i = m_i^{(j)} \mid T_i, X_i)}$$

is the fractional weight for $m_i^{(j)}$ that satisfies $\hat{w}(m_i^{(j)}) \geq 0$ and $\sum_{j=1}^{S} \hat{w}(m_i^{(j)}) = 1$. We iterate between the E-step and M-step until convergence.

The same estimation methods can be applied to the situation where missingness exists in both the mediator and outcome. We generate the imputed data sequentially for subjects with both M_i and Y_i missing. For binary M and binary Y, we generate the possible value of (m_i, y_i). For binary M and continuous Y, we generate the possible value of m_i and then the fractionally imputed data $y_i^{(1)}, \ldots, y_i^{(S)}$ for each possible value of m_i. For continuous M and continuous Y, we generate the fractionally imputed data $(m_i^{(1)}, y_i^{(1)})$,...,$(m_i^{(S)}, y_i^{(S)})$. For continuous M and binary Y, we generate the fractionally imputed data $m_i^{(1)}, \ldots, m_i^{(S)}$ and then the possible value of y_i for each fractionally imputed m_i.

The outcome model is identifiable using complete cases under Assumptions 1, 2, and

4. Therefore, an alternative approach for those scenarios is to estimate the outcome model first using complete cases, then estimate the parameters in other models through the expectation-maximization algorithm by plugging in the estimated outcome model. We tried those two slightly different approaches to our simulation settings, and both provided consistent results, with the alternative approach enjoying higher computation efficiency as expected. However, under Assumption 3, the alternative approach does not work because $\mathbb{P}(Y \mid M, T, X)$ is not identifiable using complete cases.

S5 Details on the simulation studies

In this section, we show that when $M \perp\!\!\!\perp Y \mid (T, X)$, our methods recover the underlying true values of the NIE and NDE under Assumptions 1, 2, and 4 as expected. However, we observe biases under Assumption 3 when $M \perp\!\!\!\perp Y \mid (T, X)$. In addition, we demonstrate that when M has more categories than Y, the identifiability of the model parameters is improved under Assumption 4 compared to Assumption 1 due to the additional constraint provided by the effect of M on R^Y. Furthermore, our results suggest that certain parametric assumptions outperform others in recovering the underlying model parameter values when the completeness assumption is violated.

Continuing the simulation studies in the main paper, Figure S9 presents the boxplots of percentages of bias with respect to the true values for each of the simulation scenarios when $M \perp\!\!\!\perp Y \mid (T, X)$ across 500 replications. Under Assumption 1, as shown in Figure S9 A.I (0) to D.I (0) with (0) indicating that NIE $= 0$, the percentages of bias for both the NIE and NDE estimated using all three methods are close to zero. This is because $\mathbb{P}(Y \mid M, T, X)$ is identifiable using complete cases under Assumption 1. Under Assumption 2 (A.II(0) to D.II(0)) and Assumption 4 (A.IV(0) to D.IV(0)), we reach the same conclusions as those under Assumption 1 except the fact that the estimated NIE and NDE from the multiple imputation under MAR have biases in some cases (e.g. C.II(0) and C.IV(0)) where the NIE and NDE are identifiable using complete cases. Under Assumption 3, when $M \perp\!\!\!\perp Y \mid (T, X)$, both $\mathbb{P}(Y \mid M, T, X)$ and $\mathbb{P}(M \mid T, X)$ are not identifiable, and we observe biases using all three methods.

We also check the performance of the proposed estimator for a discrete M with three categories and a binary Y under Assumptions 1 and 4, respectively, where M is generated according to a multinomial logistic regression model and Y is generated according to a logistic regression model. We consider a single covariate $X \sim \mathcal{N}(0, 1)$ and a randomized $T \sim \text{Bernoulli}(0.5)$. We generate the mediator M from

$$\log \frac{\mathbb{P}(M = 1 \mid T, X)}{\mathbb{P}(M = 0 \mid T, X)} = \alpha_{10} + \alpha_{1t}T + \alpha_{1x}X,$$
$$\log \frac{\mathbb{P}(M = 2 \mid T, X)}{\mathbb{P}(M = 0 \mid T, X)} = \alpha_{20} + \alpha_{2t}T + \alpha_{2x}X.$$

We generate the outcome Y from

$$\begin{aligned}\text{logit } \mathbb{P}(Y = 1 \mid M, T, X) = {} & \beta_0 + \beta_{m1}I(M = 1) + \beta_{m2}I(M = 2) + \beta_t T \\ & + \beta_{mt1}I(M = 1) \cdot T + \beta_{mt2}I(M = 2) \cdot T + \beta_x X.\end{aligned}$$

The binary variable R^M is generated from

$$\text{logit } \mathbb{P}(R^M = 1 \mid M, T, X) = \lambda_0 + \lambda_{m1}I(M = 1) + \lambda_{m2}I(M = 2) + \lambda_t T + \lambda_x X.$$

Under (IV) Assumption 4, the binary variable R^Y is generated from

$$\text{logit } \mathbb{P}(R^Y = 1 \mid M, T, X) = \gamma_0 + \gamma_{m1}I(M = 1) + \gamma_{m2}I(M = 2) + \gamma_t T + \gamma_x X.$$

Table S1 (Setting E) presents the specifications of parameter values. The missing rates, sample size, and number of replications are consistent with the simulation studies in the main paper.

Under Assumption 1, when $M \not\perp\!\!\!\perp Y \mid (T, X)$ but the completeness assumption in Theorem 1 does not hold, the Y model is identifiable using complete cases, but the identification of both the M and R^M models requires the completeness assumption in Theorem 1 according to our nonparametric identification results. In Figure S10, we observe that although the proposed estimators of the NIE and NDE are approximately unbiased, the parameter estimates in both the M and R^M models exhibit more complex characteristics compared to the parameter estimates in the Y model. Specifically, the estimates of α_{10} and α_{20} are concentrated around two distinct modes rather than a single point, which indicates that the parameters cannot be uniquely identified based on the observable data. Also, the estimates of α_{1x} and α_{2x} display an imbalance or non-symmetry in the distribution shape, with a long tail on one side while being relatively concentrated on the other. In addition, the irregular distribution patterns of the estimates of λ_0, λ_{m1} and λ_{m2} suggest that the parameter estimates may fail to converge to a reasonable region, which raises issues about identifiability of the model parameters even with parametric assumptions. Furthermore, the proposed estimators of the parameters in the M model are biased. On the other hand, when data are under Assumption 4 (E.IV) or when M is under a linear regression model (D.I), the parameter estimates have an approximately normal distribution shape, and the means of the parameter estimates are close to the true values as shown in Figures S11 and S12.

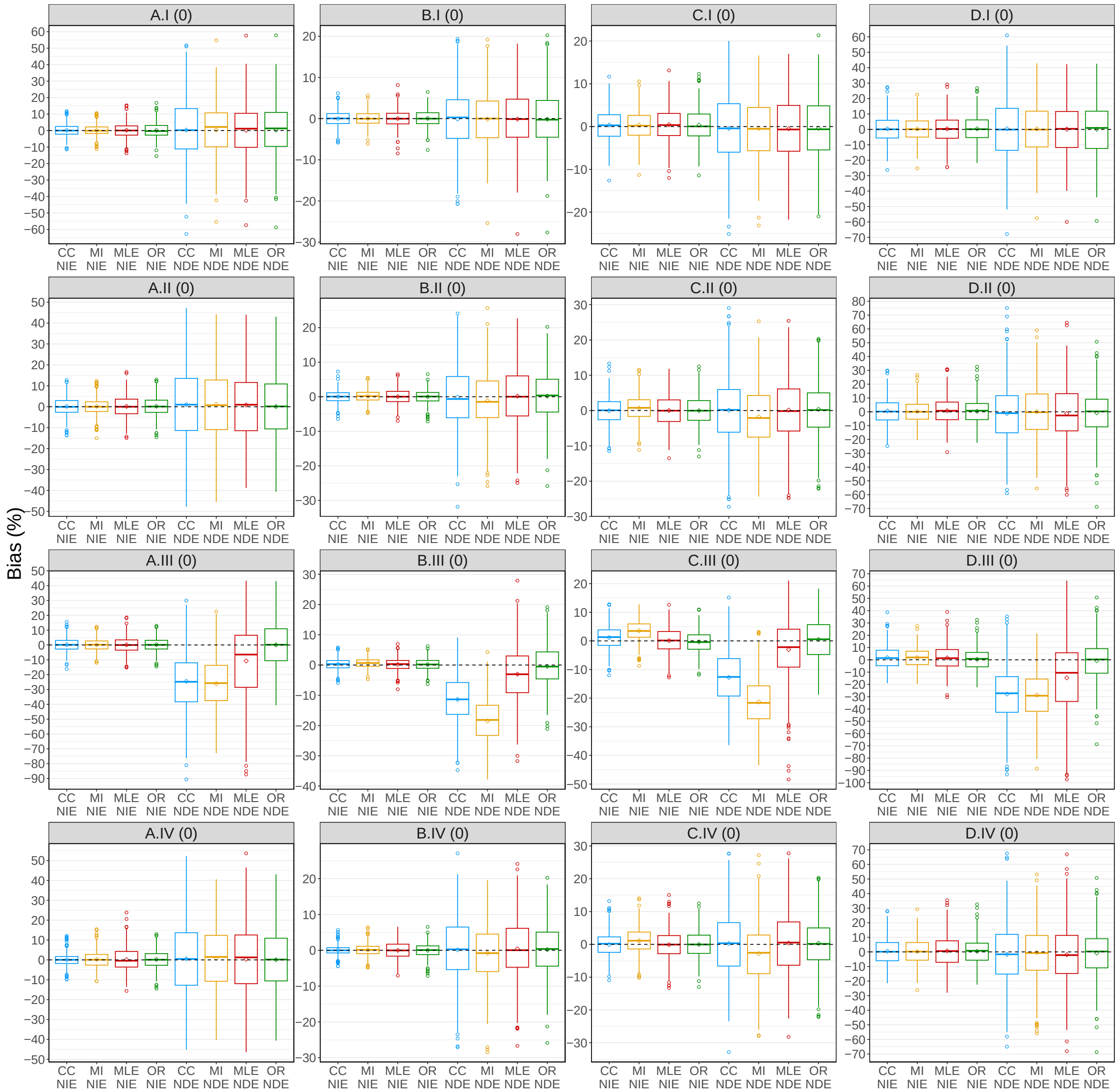


Figure S9: Simulation results when $M \perp\!\!\!\perp Y \mid (T, X)$. A, Binary M and Binary Y; B, Binary M and Continuous Y; C, Continuous M and Continuous Y; D, Continuous M and Binary Y; I, Assumption 1; II, Assumption 2; III, Assumption 3; IV, Assumption 4; CC, complete-case analysis; MI, multiple imputation estimators; MLE, our proposed methods; OR, oracle estimators; (0), $M \perp\!\!\!\perp Y \mid (T, X)$; Bias (%), {(estimate-truth)/truth}*100; The true values of the effects are nonzero except for NIE, and Bias (%) for NIE is calculated as (estimate/NDE)*100.

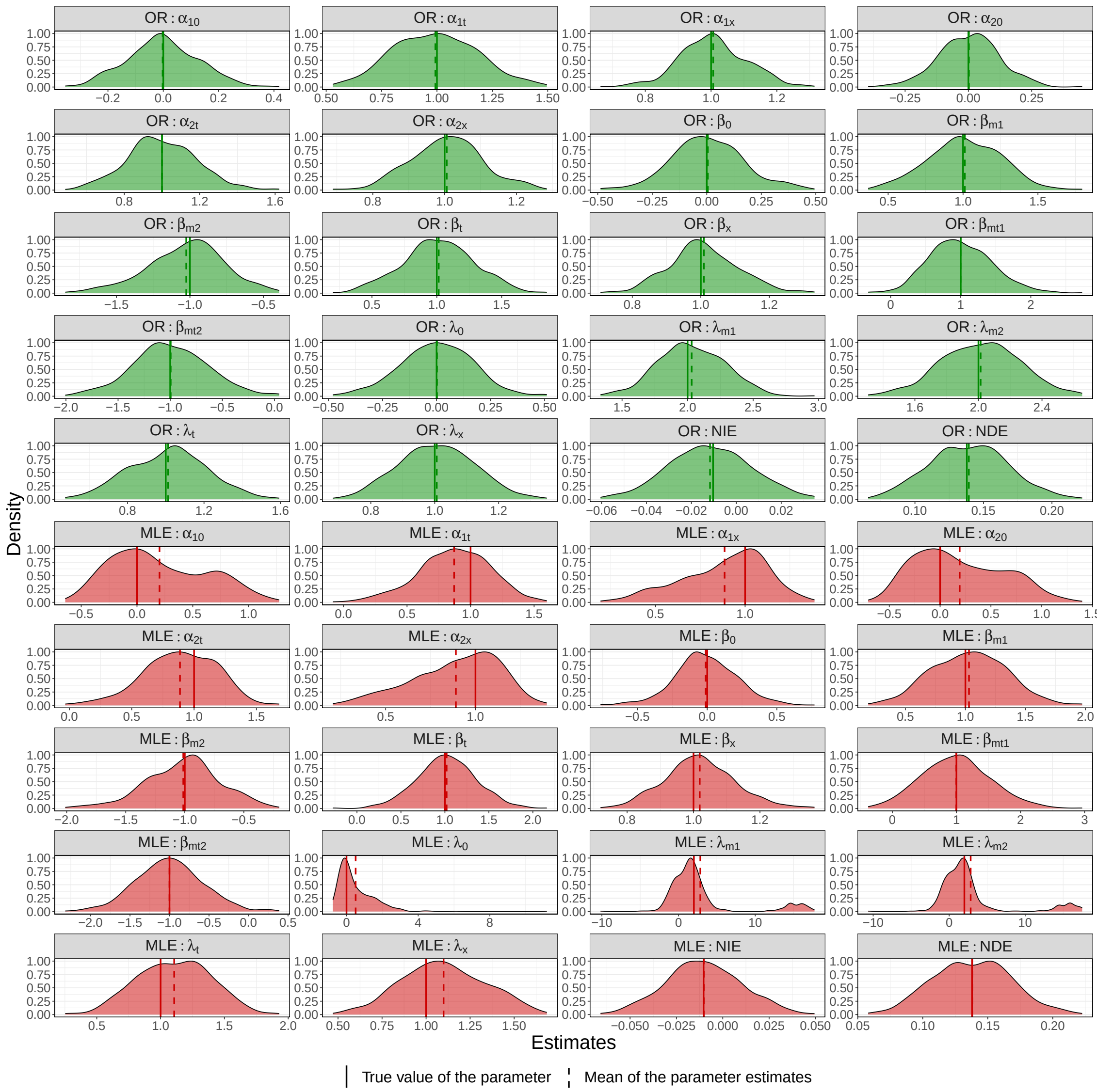


Figure S10: Simulation results under Assumption 1 when M is under a multinomial logistic regression model and Y is under a logistic regression model. $\alpha_{10}, \alpha_{1t}, \alpha_{1x}, \alpha_{20}, \alpha_{2t}, \alpha_{2x}$, parameters in the M model; $\beta_0, \beta_{m1}, \beta_{m2}, \beta_t, \beta_x, \beta_{mt1}, \beta_{mt2}$, parameters in the Y model; $\lambda_0, \lambda_{m1}, \lambda_{m2}, \lambda_t, \lambda_x$, parameters in the R^M model; MLE, our proposed methods; OR, oracle estimators.

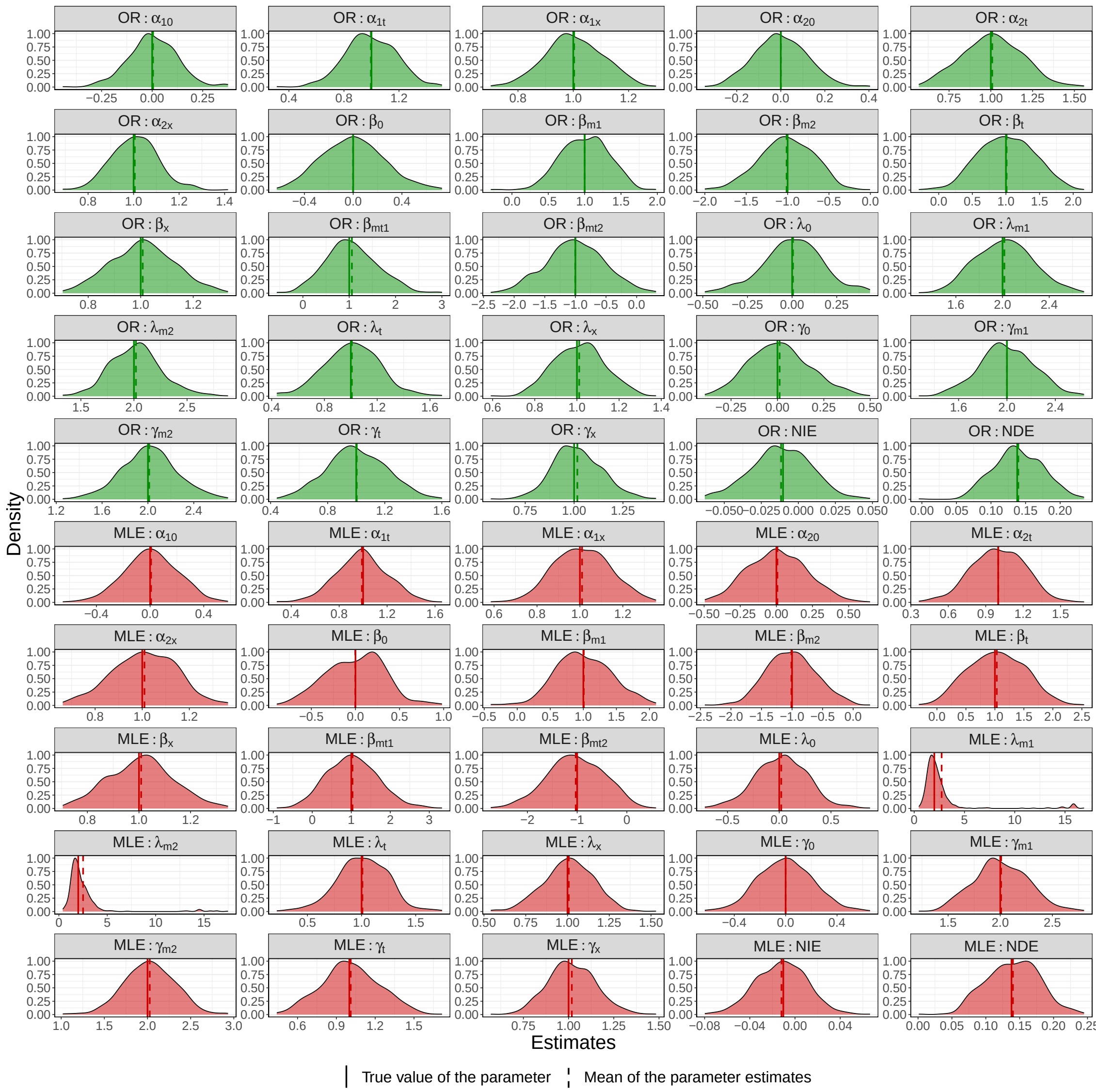


Figure S11: Simulation results under Assumption 4 when M is under a multinomial logistic regression model and Y is under a logistic regression model. $\alpha_{10}, \alpha_{1t}, \alpha_{1x}, \alpha_{20}, \alpha_{2t}, \alpha_{2x}$, parameters in the M model; $\beta_0, \beta_{m1}, \beta_{m2}, \beta_t, \beta_x, \beta_{mt1}, \beta_{mt2}$, parameters in the Y model; $\lambda_0, \lambda_{m1}, \lambda_{m2}, \lambda_t, \lambda_x$, parameters in the R^M model; $\gamma_0, \gamma_{m1}, \gamma_{m2}, \gamma_t, \gamma_x$, parameters in the R^Y model; MLE, our proposed methods; OR, oracle estimators.

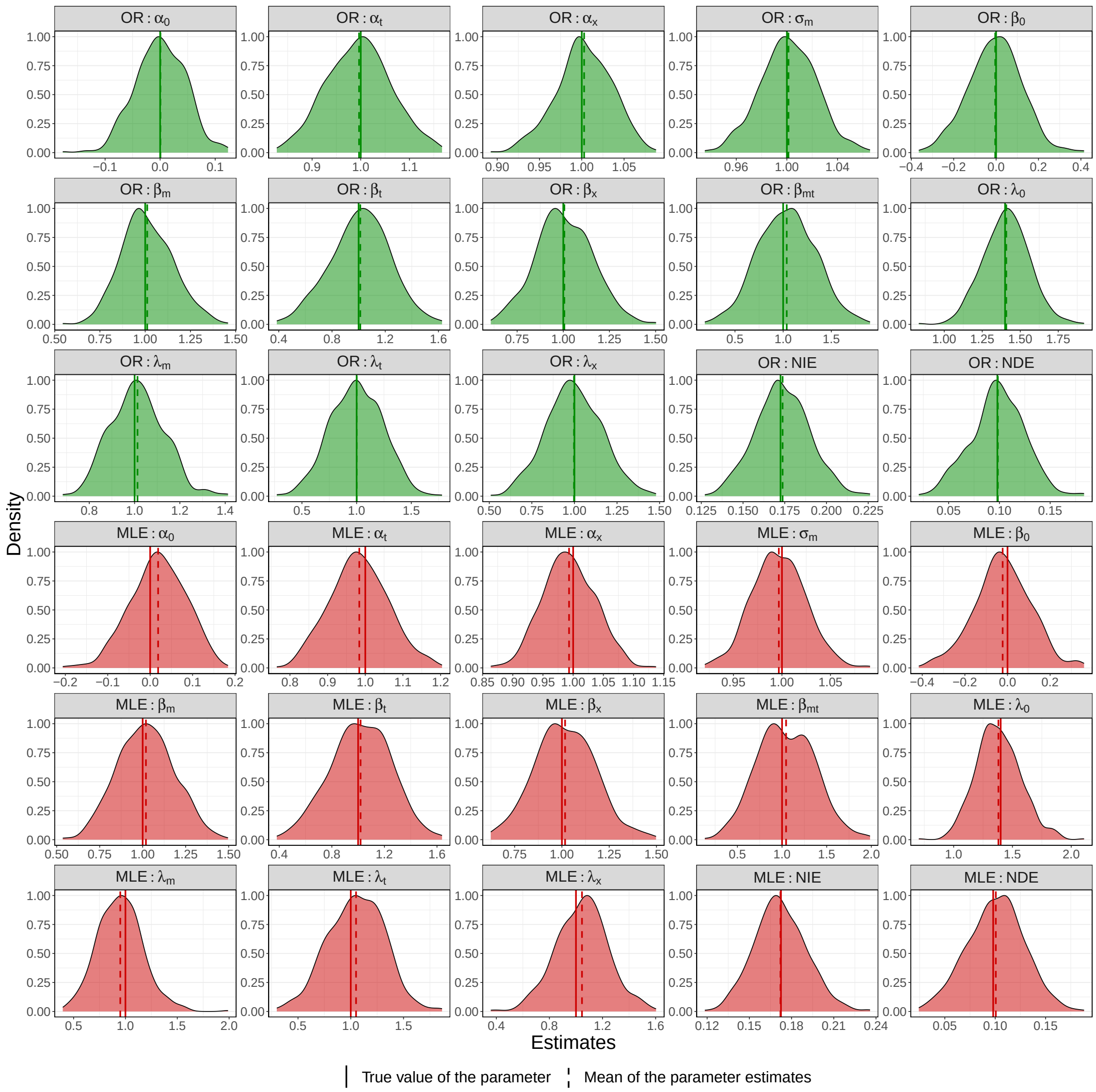


Figure S12: Simulation results under Assumption 1 when M is under a linear regression model and Y is under a logistic regression model. $\alpha_0, \alpha_t, \alpha_x, \sigma_m$ (residual standard error), parameters in the M model; $\beta_0, \beta_m, \beta_t, \beta_x, \beta_{mt}$, parameters in the Y model; $\lambda_0, \lambda_m, \lambda_t, \lambda_x$, parameters in the R^M model; MLE, our proposed methods; OR, oracle estimators.

Table S1: Specifications of the parameter values.

Setting	Model	Parameters	$M \not\perp\!\!\!\perp Y \mid (T, X)$	$M \perp\!\!\!\perp Y \mid (T, X)$
A	binary M	$(\alpha_0, \alpha_t, \alpha_x)$	$(0, 1, 1)$	$(0, 1, 1)$
	binary Y	$(\beta_0, \beta_m, \beta_t, \beta_{mt}, \beta_x)$	$(0, -1, 1, -1, 1)$	$(0, 0, 1, 0, 1)$
	R^M	$(\lambda_0, \lambda_m, \lambda_t, \lambda_x)$	$(0.3, 2, 1, 1)$	$(0.3, 2, 1, 1)$
	R^Y(II)	$(\gamma_0, \gamma_{r^M}, \gamma_t, \gamma_x)$	$(0.4, 1, 1, 1)$	$(0.4, 1, 1, 1)$
	R^Y(III)	$(\gamma_0, \gamma_y, \gamma_t, \gamma_x)$	$(0.6, 2, 1, 1)$	$(0.3, 2, 1, 1)$
	R^Y(IV)	$(\gamma_0, \gamma_m, \gamma_t, \gamma_x)$	$(0.3, 2, 1, 1)$	$(0.3, 2, 1, 1)$
B	binary M	$(\alpha_0, \alpha_t, \alpha_x)$	$(0, 1, 1)$	$(0, 1, 1)$
	continuous Y	$(\beta_0, \beta_m, \beta_t, \beta_{mt}, \beta_x)$	$(0, -1, 1, -1, 1)$	$(0, 0, 1, 0, 1)$
	R^M	$(\lambda_0, \lambda_m, \lambda_t, \lambda_x)$	$(0.3, 2, 1, 1)$	$(0.3, 2, 1, 1)$
	R^Y(II)	$(\gamma_0, \gamma_{r^M}, \gamma_t, \gamma_x)$	$(0.4, 1, 1, 1)$	$(0.4, 1, 1, 1)$
	R^Y(III)	$(\gamma_0, \gamma_y, \gamma_t, \gamma_x)$	$(0.8, -1, 1, 1)$	$(1.4, 1, 1, 1)$
	R^Y(IV)	$(\gamma_0, \gamma_m, \gamma_t, \gamma_x)$	$(0.3, 2, 1, 1)$	$(0.3, 2, 1, 1)$
C	continuous M	$(\alpha_0, \alpha_t, \alpha_x)$	$(0, 1, 1)$	$(0, 1, 1)$
	continuous Y	$(\beta_0, \beta_m, \beta_t, \beta_{mt}, \beta_x)$	$(0, 1, 1, 1, 1)$	$(0, 0, 1, 0, 1)$
	R^M	$(\lambda_0, \lambda_m, \lambda_t, \lambda_x)$	$(1.4, 1, 1, 1)$	$(1.4, 1, 1, 1)$
	R^Y(II)	$(\gamma_0, \gamma_{r^M}, \gamma_t, \gamma_x)$	$(0.4, 1, 1, 1)$	$(0.4, 1, 1, 1)$
	R^Y(III)	$(\gamma_0, \gamma_y, \gamma_t, \gamma_x)$	$(1.8, 1, 1, 1)$	$(1.4, 1, 1, 1)$
	R^Y(IV)	$(\gamma_0, \gamma_m, \gamma_t, \gamma_x)$	$(1.4, 1, 1, 1)$	$(1.4, 1, 1, 1)$
D	continuous M	$(\alpha_0, \alpha_t, \alpha_x)$	$(0, 1, 1)$	$(0, 1, 1)$
	binary Y	$(\beta_0, \beta_m, \beta_t, \beta_{mt}, \beta_x)$	$(0, 1, 1, 1, 1)$	$(0, 0, 1, 0, 1)$
	R^M	$(\lambda_0, \lambda_m, \lambda_t, \lambda_x)$	$(1.4, 1, 1, 1)$	$(1.4, 1, 1, 1)$
	R^Y(II)	$(\gamma_0, \gamma_{r^M}, \gamma_t, \gamma_x)$	$(0.4, 1, 1, 1)$	$(0.4, 1, 1, 1)$
	R^Y(III)	$(\gamma_0, \gamma_y, \gamma_t, \gamma_x)$	$(0.4, 2, 1, 1)$	$(0.3, 2, 1, 1)$
	R^Y(IV)	$(\gamma_0, \gamma_m, \gamma_t, \gamma_x)$	$(1.4, 1, 1, 1)$	$(1.4, 1, 1, 1)$
E	discrete M	$(\alpha_{10}, \alpha_{1t}, \alpha_{1x}, \alpha_{20}, \alpha_{2t}, \alpha_{2x})$	$(0, 1, 1, 0, 1, 1)$	
	binary Y	$(\beta_0, \beta_{m1}, \beta_{m2}, \beta_t, \beta_{mt1}, \beta_{mt2}, \beta_x)$	$(0, 1, -1, 1, 1, -1, 1)$	
	R^M	$(\lambda_0, \lambda_{m1}, \lambda_{m2}, \lambda_t, \lambda_x)$	$(0, 2, 2, 1, 1)$	
	R^Y(IV)	$(\gamma_0, \gamma_{m1}, \gamma_{m2}, \gamma_t, \gamma_x)$	$(0, 2, 2, 1, 1)$	

S6 Sensitivity analysis

We consider the two-part Gamma model under Assumption 2 from the data analysis as a starting model for building the sensitivity analysis. It is possible that the missingness of earnings also depends on the earnings itself and the educational and vocational attainment, in addition to the missingness of the educational and vocational attainment, as described in Figure S13. Now we assess the sensitivity of our conclusions to the additional impacts of H on R^Y and M on R^Y. We revise the model for R^Y as follows:

$$\begin{aligned} & \text{logit}\ \mathbb{P}(R^Y = 1 \mid R^M = r^M, H = h, M = m, T = t, X = x) \\ = \quad & \gamma_0 + \gamma_{r^M} r^M + \gamma_h h + \gamma_m m + \gamma_t t + \gamma_x^{\mathrm{T}} x, \end{aligned}$$

where γ_h and γ_m are the sensitivity parameters. We consider a large log odds ratio (Chen et al., 2010) and let both sensitivity parameters vary among -2, 0 and 2. When $\gamma_h = 0$ and $\gamma_m = 0$, it is the same as the MNAR mechanism under Assumption 2 that stands out in the data analysis.

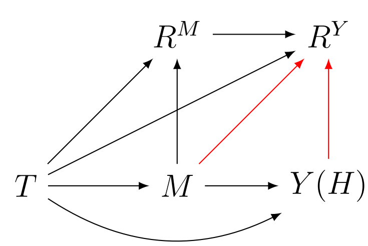


Figure S13: The DAG describes the missing mechanism for the sensitivity analysis. The DAG conditions on X and allows X to have directed arrows to all variables in the DAG.

Table S2 presents the sensitivity analysis results. The estimated NIE increases more than 10% in the case where $\gamma_m = -2$ and $\gamma_h = 2$, and where $\gamma_m = 0$ and $\gamma_h = 2$. The estimated NDE decreases more than 10% in the case where $\gamma_m = 0$ and $\gamma_h = 2$ and increases more than 10% in the case where $\gamma_m = 2$ and $\gamma_h = 2$. However, the estimated NIEs are positive and significant at the 0.05 significance level, and the estimated NDEs are positive but not significant at the 0.05 significance level for all pairs of values (γ_m, γ_h) considered. In summary, the conclusions on the NIE and NDE in the NJCS are not sensitive to some strong impacts of H on R^Y and M on R^Y in addition to the impact of R^M on R^Y.

Table S2: Sensitivity analysis results from the two-part Gamma model under Assumption 2. Est, estimate; CI, confidence interval based on 500 bootstrap samples; γ_h (sensitivity parameter), coefficient of H in the R^Y model; γ_m (sensitivity parameter), coefficient of M in the R^Y model.

		$\gamma_h = -2$		$\gamma_h = 0$		$\gamma_h = 2$	
Parameters	γ_m	Est	95% CI	Est	95% CI	Est	95% CI
NIE	-2	11.15	(7.97, 14.49)	11.49	(8.24, 14.83)	14.33	(11.02, 17.89)
	0	11.30	(8.12, 14.58)	10.94	(7.94, 14.29)	13.40	(10.22, 16.78)
	2	11.39	(8.19, 14.63)	10.83	(7.98, 14.25)	10.48	(7.15, 14.81)
NDE	-2	13.18	(−1.53, 27.88)	13.90	(−1.00, 28.57)	11.72	(−2.33, 26.34)
	0	12.82	(−1.90, 27.52)	12.93	(−1.95, 27.64)	11.27	(−3.12, 25.57)
	2	12.50	(−2.24, 27.16)	12.25	(−2.38, 27.33)	15.43	(−0.18, 29.47)

References

Chen, H., Cohen, P., and Chen, S. (2010). How big is a big odds ratio? interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation*, 39(4):860–864.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from in-complete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika*, 98(1):119–132.

Li, Y., Miao, W., Shpitser, I., and Tchetgen, E. J. T. (2023). A self-censoring model for multivariate nonignorable nonmonotone missing data. *Biometrics*, 79(4):3203–3214.

Ma, W.-Q., Geng, Z., and Hu, Y.-H. (2003). Identification of graphical models for nonignorable nonresponse of binary outcomes in longitudinal studies. *Journal of Multivariate Analysis*, 87(1):24–45.

Newey, W. and Powell, J. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.