Longitudinal mediation analysis with missing not at random data
Supplementary Material

ZHU Yuxuan · ZHANG Hong · ZHAO Saijun

ZHU Yuxuan
Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026*, China.*
Email: zyx1998@mail.ustc.edu.com.
ZHANG Hong (Corresponding author)
Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026*, China.*
Email: zhangh@ustc.edu.cn.
ZHAO Saijun
Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026*, China.*
Email: zsj123@mail.ustc.edu.cn.

Appendix A Proof of Proposition 3.1

It should be noted that the identifiability of the model at subsequent moments depends on the consistency of the estimators of the parameters at previous moments, and the consistency of the estimators of the parameters at each moment also depends on the identifiability of the model at that moment.

We first give a proof of the identifability of the outcome model at time t with the assumption that the parameter estimates of the outcome models at the first $t-1$ time points are consistent. In other words, we only need to prove the identifiability of $P(Y_t = 1 \mid A = a, M_t = m, X = x, Y_{t-1} = y_{t-1})$ for any given (a, m, x, y_{t-1}) without missing data before time t. Following the idea of Li and Zhou[1], we use a proof by contradiction to show this result.

Suppose there are two parameters underlying the observed data distribution:

$$\begin{aligned}&P_1(R_t = 1 \mid A, M_t, Y_t = y_t, U)P_1(Y_t = y_t \mid A, M_t, Y_{t-1}, U, V)\\=&P_2(R_t = 1 \mid A, M_t, Y_t = y_t, U)P_2(Y_t = y_t \mid A, M_t, Y_{t-1}, U, V).\end{aligned}$$

For simplicity, we omit a, m, and u in the corresponding expressions. For $j = 1, 2$, denote

$$p_{ij}(y_t) = P_i(Y_t = y_t \mid V = v_j, Y_{t-1} = y_{t-1})$$

and

$$Q(y_t) = P_1(R_t = 1 \mid Y_t = y_t)/P_2(R_t = 1 \mid Y_t = y_t) - 1,$$

we have

$$P_1(R_t = 1 \mid Y_t = y_t)p_{11}(y_t) = P_2(R_t = 1 \mid Y_t = y_t)p_{21}(y_t) \tag{S.1}$$

and

$$P_1(R_t = 1 \mid Y_t = y_t)p_{12}(y_t) = P_2(R_t = 1 \mid Y_t = y_t)p_{22}(y_t). \tag{S.2}$$

Let $y_t = 0, 1$ in equations (S.1) and (S.2), we have

$$\begin{cases} P_1(R_t = 1 \mid Y_t = 0)p_{11}(0) = P_2(R_t = 1 \mid Y_t = 0)p_{21}(0), \\ P_1(R_t = 1 \mid Y_t = 1)p_{11}(1) = P_2(R_t = 1 \mid Y_t = 1)p_{21}(1), \\ P_1(R_t = 1 \mid Y_t = 0)p_{12}(0) = P_2(R_t = 1 \mid Y_t = 0)p_{22}(0), \\ P_1(R_t = 1 \mid Y_t = 1)p_{12}(1) = P_2(R_t = 1 \mid Y_t = 1)p_{22}(1), \end{cases} \tag{S.3}$$

i.e.,

$$\begin{cases} (Q_1(0) + 1)p_{11}(0) = p_{21}(0), \\ (Q_1(1) + 1)p_{11}(1) = p_{21}(1), \\ (Q_1(0) + 1)p_{12}(0) = p_{22}(0), \\ (Q_1(1) + 1)p_{12}(1) = p_{22}(1). \end{cases} \tag{S.4}$$

Since $p_{ij}(0) + p_{ij}(1) = 1$ for $i = 1, 2$ and $j = 1, 2$, we have

$$Q(0)p_{11}(0) + Q(1)p_{11}(1) = 0, \tag{S.5}$$

$$Q(0)p_{12}(0) + Q(1)p_{12}(1) = 0. \tag{S.6}$$

Consequently,

$$(Q(1) - Q(0))(p_{12}(1) - p_{11}(1)) = 0.$$

From the definition of the shadow variable, we know that there exists v_1 and v_2 such that $p_{12}(1) - p_{11}(1) \neq 0$. So $Q(1) - Q(0) = 0$, which implies $Q(1) = Q(0) = 0$ because they couldn't have the same signs according to equations (S.5) and (S.6). In other words,

$$P_1(R_t = 1 \mid Y_t = y_t) = P_2(R_t = 1 \mid Y_t = y_t),$$

so

$$\begin{aligned} &P_1(Y_t = y_t \mid A = a, M_t = m, Y_{t-1} = y_{t-1}, U = u, V = v) \\ =&P_2(Y_t = y_t \mid A = a, M_t = m, Y_{t-1} = y_{t-1}, U = u, V = v) \end{aligned}$$

for any (a, m, y_{t-1}, u, v), this creates a contradiction. As a result, $P(Y_t = 1 \mid A = a, M_t = m, X = x, Y_{t-1} = y_{t-1})$ is identifiable.

We then establish the consistency of the parameter estimators under the identifiability assumptions. By the condition that the matrices

$$\frac{\partial \omega_{R_t}(\beta_t)}{\partial \beta_t} R_t h_1^\tau \text{ and } -\frac{\partial E(Y_t \mid A, M_t, U, V, Y_{t-1}; \gamma_t)}{\partial \gamma_t} h_2^\tau$$

are invertible, we have that

$$E\begin{pmatrix} \frac{\partial \omega_{R_t}(\beta_t)}{\partial \beta_t} R_t h_1^\tau & 0 \\ \frac{\partial \omega_{R_t}(\beta_t)}{\partial \beta_t} R_t h_2^\tau & -\frac{\partial E(Y_t \mid A, M_t, U, V, Y_{t-1}; \gamma_t)}{\partial \gamma_t} h_2^\tau \end{pmatrix}$$

is invertable. Consequently,

$$\text{equations (7) and (9) have (locally) unique solutions.} \tag{S.7}$$

Let $L_t = (A, M_t, U, V)$ and omits variables in $h_1(\cdot)$ and $h_2(\cdot)$. First,

$$\begin{aligned} &E[\{\omega_{R_t}(A, M_t, Y_t, U; \beta_t)R_t - 1\}h_1] \\ =&E[E[\{\omega_{R_t}(A, M_t, Y_t, U; \beta_t)R_t - 1\}h_1 \mid A, M_t, Y_t, U, V]], \\ =&E[\{1/E(R_t \mid A, M_t, Y_t, U, V)E(R_t \mid A, M_t, Y_t, U, V) - 1\}h_1] = 0. \end{aligned} \tag{S.8}$$

Second,

$$
\begin{aligned}
&E[\{\omega_{R_t}(A, M_t, Y_t, U; \beta_t) R_t Y_t - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2] \\
=&E[E[\{\omega_{R_t}(A, M_t, Y_t, U; \beta_t) R_t Y_t - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2 \mid L_t, Y_t, Y_{t-1}]] \\
=&E[\{1/E(R_t \mid L_t, Y_t, Y_{t-1}; \beta_t) E(R_t \mid L_t, Y_t, Y_{t-1}; \beta_t) Y_t - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2] \\
=&E[\{Y_t - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2] \\
=&E[E[\{Y_t - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2 \mid L_t, Y_{t-1}]] \\
=&E[\{E(Y_t \mid L_t, Y_{t-1}; \gamma_t) - E(Y_t \mid L_t, Y_{t-1}; \gamma_t)\} h_2] \\
=&0. \qquad \text{(S.9)}
\end{aligned}
$$

The second part of the proposition follows from (S.7)-(S.9) by the standard limit theorem for estimation equation.

Appendix B Other simulation results

Appendix B.1 Verification of natural effect model (14)

The distributions of A, C_1, and C_2 in this section are the same as those in Section 4.1. We considered time-varying binary mediators and binary outcomes for $T = 5$ time points. The mediator and outcome were generated from models

$$\text{logit}(P(M_1 = 1)) = \delta_0 + \delta_1 A + \delta_2 C_1 + \delta_3 C_2 \qquad \text{(S.10)}$$

and

$$\text{logit}(P(Y_1 = 1)) = \gamma_0 + \gamma_1 A + \gamma_M M_1 + \gamma_2 C_1 + \gamma_3 C_2 \qquad \text{(S.11)}$$

for $t = 1$ and models

$$\text{logit}(P(M_t = 1)) = \delta_0 + \delta_1 A + \delta_M M_{t-1} + \delta_Y Y_{t-1} + \delta_2 C_1 + \delta_3 C_2 \qquad \text{(S.12)}$$

and

$$\text{logit}(P(Y_t = 1)) = \gamma_0 + \gamma_1 A + \gamma_M M_t + \gamma_Y Y_{t-1} + \gamma_2 C_1 + \gamma_3 C_2 \qquad \text{(S.13)}$$

for $t = 2, \ldots, 5$. We considered two parameter settings for mediator and outcome models. For setting 1,

$$(\delta_0, \delta_1, \delta_2, \delta_3, \delta_M, \delta_Y) = (0.3, -0.8, 0.5, 1, 0.1, 0.1),$$

and

$$(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_M, \gamma_Y) = (-0.4, 1.5, -0.4, 1.2, -1, 0.3).$$

For setting 2,

$$(\delta_0, \delta_1, \delta_2, \delta_3, \delta_M, \delta_Y) = (0.7, -0.5, 0.5, 0.5, 0.15, 0.15),$$

and

$$(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_M, \gamma_Y) = (1, -1.5, 0.4, 0.8, 1.2, 0.3).$$

We generated $B = 500$ samples with sample size $N = 10^6$ and the natural effects (see Table S.1) were calculated by definition ((4)-(6)). It's obvious that the natural effects at time $t = 2, \ldots, 5$ are very close to each other.

Table S.1: True values of the natural effects at five time points under two parameter settings 1 and 2 ($T = 5$).

Natural effect	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
Setting 1					
DE_t	1.304	1.368	1.370	1.370	1.370
IDE_t	0.154	0.162	0.162	0.162	0.162
TE_t	1.458	1.530	1.532	1.532	1.532
Setting 2					
DE_t	-1.303	-1.384	-1.384	-1.383	-1.383
IDE_t	-0.132	-0.141	-0.140	-0.140	-0.140
TE_t	-1.435	-1.525	-1.524	-1.524	-1.524

Appendix B.2 Additional simulations for comparison of models (14) and (13)

First, we conducted some additional simulations as supplements of Section 4.1. The distribution of A, C_1, C_2 in this section is the same as Section 4.1. We consider two parameter settings in models (15)-(18) for $T = 3$. For setting 1,

$$(\delta_0, \delta_1, \delta_2, \delta_3, \delta_M, \delta_Y) = (0.5, 0.5, 1.5, -1, 0.25, 0.25),$$

and

$$(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_M, \gamma_Y) = (0.4, 1.2, -0.8, 1.2, 1.2, 0.3).$$

For setting 2,

$$(\delta_0, \delta_1, \delta_2, \delta_3, \delta_M, \delta_Y) = (0.7, -0.5, 0.5, 0.5, 0.15, 0.15),$$

and

$$(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_M, \gamma_Y) = (1, -1.5, 0.4, 0.8, 1.2, 0.3).$$

We generated $B = 500$ samples with sample size $n = 2000$ or 5000 and estimated natural effects through fitting the natural effect models (13) and (14) using complete data. The results at time $t = 2, \ldots, T = 3$ under settings 1 and 2 are shown in Tables S.2 and S.3, respectively. The estimation biases based on model (13) are much larger than (14) for most natural effects. The RMSEs and SEs of the effects estimated using models (13) and (14) are similar to the results in Section 4.1.

Shown in Tables S.4-S.6 are the results of $T = 5$. Evidently, the biases of model (13) were much larger than model (14) for most effects. For $t = 2, 3$, some RMSEs of model (13) were smaller than model (14) due to smaller SEs of model (13). As t increased to 5, the RMSEs of model (13) were much larger than (14).

Overall, model (14) outperformed model (13) in terms of the estimation of natural effects with complete data.

Next, we conducted some simulations with $T = 5$ as supplements of Section 4.2. The results are presented in Tables S.7 and S.8. The biases of all three methods got larger. As a result, M-EE showed poor CPs for some effects. The biases of O-EE were again much larger than M-EE for most effects, and M-Heck performed poorly in all aspects. RMSE showed a pattern similar to the complete data case.

Table S.2: Comparison of models (13) and (14) in terms of natural effects with complete data under parameter setting 1 ($T = 3$).

Natural effect	True value	Model (13) Bias	Model (13) RMSE	Model (13) SE	Model (14) Bias	Model (14) RMSE	Model (14) SE
$n = 2000$							
DE_2	1.167	-0.014	0.091	0.090	0.004	0.115	0.115
DE_3	1.161	0.021	0.149	0.148	0.009	0.115	0.115
IDE_2	0.093	-0.000	0.016	0.016	0.000	0.022	0.022
IDE_3	0.092	0.001	0.029	0.029	0.001	0.022	0.022
TE_2	1.260	-0.014	0.091	0.091	0.004	0.115	0.115
TE_3	1.253	0.022	0.149	0.147	0.010	0.116	0.115
$n = 5000$							
DE_2	1.167	-0.023	0.063	0.059	-0.002	0.074	0.074
DE_3	1.161	0.016	0.098	0.097	0.003	0.074	0.074
IDE_2	0.093	-0.001	0.010	0.010	-0.001	0.013	0.013
IDE_3	0.092	0.000	0.017	0.017	0.000	0.013	0.013
TE_2	1.260	-0.024	0.064	0.059	-0.003	0.074	0.074
TE_3	1.253	0.016	0.099	0.098	0.003	0.074	0.074

Bias, estimated natural effect minus the true value; RMSE, root mean square error; SE, empirical standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t.

Table S.3: Comparison of models (13) and (14) in terms of natural effects with complete data under parameter setting 2 ($T = 3$).

Natural effect	True value	Model (13) Bias	Model (13) RMSE	Model (13) SE	Model (14) Bias	Model (14) RMSE	Model (14) SE
$n = 2000$							
DE_2	-1.384	0.025	0.073	0.068	-0.005	0.088	0.088
DE_3	-1.384	-0.023	0.112	0.110	-0.005	0.088	0.088
IDE_2	-0.141	0.010	0.022	0.019	0.007	0.026	0.025
IDE_3	-0.140	0.005	0.032	0.032	0.006	0.025	0.025
TE_2	-1.525	0.035	0.076	0.067	0.002	0.086	0.086
TE_3	-1.524	-0.018	0.109	0.108	0.001	0.086	0.086
$n = 5000$							
DE_2	1.167	0.022	0.048	0.043	-0.005	0.050	0.050
DE_3	1.161	-0.023	0.069	0.065	-0.005	0.050	0.050
IDE_2	0.093	0.011	0.016	0.012	0.008	0.016	0.015
IDE_3	0.092	0.006	0.020	0.019	0.007	0.016	0.015
TE_2	1.260	0.032	0.054	0.043	0.003	0.051	0.051
TE_3	1.253	-0.017	0.068	0.066	0.002	0.051	0.051

Bias, estimated natural effect minus the true value; RMSE, root mean square error; SE, empirical standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t.

Table S.4: Comparison of models (13) and (14) in terms of natural effects with complete data under parameter setting in Section 4.1 ($T = 5$).

Natural effect	True value	Model (13)			Model (14)		
		Bias	RMSE	SE	Bias	RMSE	SE
$n = 2000$							
DE_2	1.368	-0.025	0.059	0.053	0.006	0.051	0.050
DE_3	1.370	-0.010	0.046	0.045	0.004	0.050	0.050
DE_4	1.370	0.006	0.057	0.056	0.004	0.050	0.050
DE_5	1.370	0.023	0.082	0.079	0.004	0.050	0.050
IDE_2	0.162	-0.002	0.018	0.018	0.000	0.022	0.022
IDE_3	0.162	-0.002	0.019	0.019	-0.000	0.022	0.022
IDE_4	0.162	-0.001	0.027	0.027	-0.000	0.022	0.022
IDE_5	0.162	-0.001	0.038	0.038	-0.000	0.022	0.022
TE_2	1.530	-0.027	0.058	0.052	0.006	0.048	0.047
TE_3	1.532	-0.012	0.044	0.042	0.003	0.047	0.047
TE_4	1.532	0.005	0.052	0.052	0.003	0.047	0.047
TE_5	1.532	0.022	0.077	0.074	0.003	0.047	0.047
$n = 5000$							
DE_2	1.368	-0.030	0.046	0.035	-0.001	0.034	0.034
DE_3	1.370	-0.017	0.034	0.030	-0.003	0.034	0.034
DE_4	1.370	-0.002	0.037	0.037	-0.003	0.034	0.034
DE_5	1.370	0.013	0.053	0.051	-0.003	0.034	0.034
IDE_2	0.162	-0.002	0.011	0.010	0.003	0.015	0.014
IDE_3	0.162	-0.000	0.012	0.012	0.002	0.015	0.014
IDE_4	0.162	0.003	0.018	0.017	0.002	0.015	0.014
IDE_5	0.162	0.005	0.025	0.024	0.002	0.015	0.014
TE_2	1.530	-0.032	0.048	0.035	0.002	0.032	0.032
TE_3	1.532	-0.017	0.033	0.029	-0.001	0.032	0.032
TE_4	1.532	0.001	0.034	0.035	-0.001	0.032	0.032
TE_5	1.532	0.018	0.052	0.049	-0.001	0.032	0.032

Bias, estimated natural effect minus the true value; RMSE, root mean square error; SE, empirical standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t.

Table S.5: Comparison of models (13) and (14) in terms of natural effects with complete data under parameter setting 1 ($T = 5$).

Natural effect	True value	Model (13) Bias	Model (13) RMSE	Model (13) SE	Model (14) Bias	Model (14) RMSE	Model (14) SE
$n = 2000$							
DE_2	1.166	-0.025	0.087	0.084	0.001	0.084	0.084
DE_3	1.161	-0.006	0.072	0.071	0.006	0.084	0.084
DE_4	1.161	0.010	0.095	0.094	0.007	0.084	0.084
DE_5	1.161	0.025	0.138	0.136	0.007	0.084	0.084
IDE_2	0.093	-0.001	0.016	0.016	-0.000	0.017	0.017
IDE_3	0.092	0.000	0.015	0.015	0.000	0.017	0.017
IDE_4	0.092	0.001	0.020	0.020	0.000	0.017	0.017
IDE_5	0.092	0.001	0.028	0.028	0.000	0.017	0.017
TE_2	1.260	-0.026	0.088	0.084	0.001	0.082	0.082
TE_3	1.253	-0.005	0.071	0.071	0.006	0.082	0.082
TE_4	1.253	0.011	0.092	0.092	0.007	0.082	0.082
TE_5	1.252	0.026	0.134	0.132	0.007	0.082	0.082
$n = 5000$							
DE_2	1.166	-0.025	0.059	0.054	-0.001	0.053	0.053
DE_3	1.161	-0.009	0.047	0.047	0.004	0.053	0.053
DE_4	1.161	0.004	0.061	0.061	0.005	0.053	0.053
DE_5	1.161	0.015	0.088	0.087	0.005	0.053	0.053
IDE_2	0.093	-0.001	0.009	0.009	-0.001	0.011	0.011
IDE_3	0.092	-0.000	0.009	0.009	-0.001	0.011	0.011
IDE_4	0.092	-0.001	0.013	0.013	-0.000	0.011	0.011
IDE_5	0.092	-0.001	0.019	0.019	-0.000	0.011	0.011
TE_2	1.260	-0.026	0.059	0.053	-0.002	0.053	0.053
TE_3	1.253	-0.009	0.047	0.046	0.004	0.053	0.053
TE_4	1.253	0.003	0.061	0.061	0.005	0.053	0.053
TE_5	1.252	0.014	0.087	0.086	0.005	0.053	0.053

Bias, estimated natural effect minus the true value; RMSE, root mean square error; SE, empirical standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t.

Table S.6: Comparison of models (13) and (14) in terms of natural effects with complete data under parameter setting 2 ($T = 5$).

Natural effect	True value	Model (13) Bias	Model (13) RMSE	Model (13) SE	Model (14) Bias	Model (14) RMSE	Model (14) SE
$n = 2000$							
DE_2	-1.384	0.029	0.069	0.063	-0.005	0.065	0.065
DE_3	-1.384	0.011	0.058	0.057	-0.006	0.065	0.065
DE_4	-1.383	-0.006	0.075	0.075	-0.006	0.065	0.065
DE_5	-1.383	-0.023	0.108	0.105	-0.006	0.065	0.065
IDE_2	-0.141	0.010	0.020	0.017	0.007	0.020	0.019
IDE_3	-0.140	0.008	0.018	0.016	0.007	0.020	0.019
IDE_4	-0.140	0.007	0.022	0.021	0.007	0.020	0.019
IDE_5	-0.140	0.005	0.030	0.029	0.007	0.020	0.019
TE_2	-1.525	0.039	0.074	0.063	0.002	0.064	0.064
TE_3	-1.524	0.019	0.059	0.056	0.001	0.064	0.064
TE_4	-1.524	0.001	0.073	0.073	0.000	0.064	0.064
TE_5	-1.524	-0.018	0.105	0.103	0.000	0.064	0.064
$n = 5000$							
DE_2	-1.384	0.035	0.055	0.042	-0.005	0.043	0.043
DE_3	-1.384	0.014	0.041	0.038	-0.005	0.043	0.043
DE_4	-1.383	-0.007	0.048	0.048	-0.005	0.043	0.043
DE_5	-1.383	-0.028	0.071	0.065	-0.005	0.043	0.043
IDE_2	-0.141	0.012	0.016	0.011	0.008	0.014	0.011
IDE_3	-0.140	0.009	0.013	0.010	0.008	0.014	0.011
IDE_4	-0.140	0.007	0.015	0.013	0.008	0.014	0.011
IDE_5	-0.140	0.005	0.019	0.018	0.008	0.014	0.011
TE_2	-1.525	0.047	0.063	0.042	0.004	0.043	0.043
TE_3	-1.524	0.023	0.044	0.038	0.003	0.043	0.043
TE_4	-1.524	0.000	0.047	0.047	0.002	0.043	0.043
TE_5	-1.524	-0.023	0.069	0.065	0.002	0.043	0.043

Bias, estimated natural effect minus the true value; RMSE, root mean square error; SE, empirical standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t.

Table S.7: Comparison of three methods in terms of natural effects ($T = 5$ and $n = 2000$).

Natural	True	M-EE					O-EE					M-Heck				
effect	value	Bias	RMSE	CP	SE	SEE	Bias	RMSE	CP	SE	SEE	Bias	RMSE	CP	SE	SEE
$n = 2000$																
DE_2	1.368	-0.015	0.062	0.976	0.060	0.064	-0.042	0.078	0.906	0.066	0.066	-0.074	0.098	0.910	0.064	0.077
DE_3	1.370	-0.017	0.063	0.974	0.060	0.064	-0.029	0.062	0.934	0.055	0.058	-0.077	0.100	0.902	0.064	0.077
DE_4	1.370	-0.017	0.063	0.974	0.060	0.064	-0.014	0.069	0.974	0.067	0.070	-0.077	0.100	0.902	0.064	0.077
DE_5	1.370	-0.017	0.063	0.974	0.060	0.064	0.001	0.095	0.960	0.095	0.095	-0.077	0.100	0.902	0.064	0.077
IDE_2	0.162	-0.003	0.022	0.966	0.022	0.023	-0.004	0.019	0.904	0.019	0.016	-0.022	0.036	0.878	0.028	0.026
IDE_3	0.162	-0.004	0.022	0.964	0.022	0.023	-0.005	0.019	0.952	0.018	0.019	-0.023	0.036	0.874	0.028	0.026
IDE_4	0.162	-0.004	0.022	0.964	0.022	0.023	-0.004	0.026	0.970	0.026	0.028	-0.023	0.036	0.874	0.028	0.026
IDE_5	0.162	-0.004	0.022	0.964	0.022	0.023	-0.004	0.036	0.980	0.036	0.039	-0.023	0.036	0.874	0.028	0.026
TE_2	1.530	-0.018	0.058	0.978	0.056	0.066	-0.046	0.078	0.920	0.063	0.068	-0.097	0.115	0.856	0.063	0.080
TE_3	1.532	-0.021	0.059	0.974	0.056	0.066	-0.033	0.062	0.956	0.052	0.061	-0.099	0.117	0.844	0.063	0.080
TE_4	1.532	-0.021	0.059	0.974	0.056	0.066	-0.018	0.064	0.972	0.061	0.072	-0.099	0.117	0.844	0.063	0.080
TE_5	1.532	-0.021	0.059	0.974	0.056	0.066	-0.003	0.086	0.970	0.086	0.095	-0.099	0.117	0.844	0.063	0.080

Bias, estimated natural effect minus the true value; RMSE, root mean square error; CP, the coverage probability of 95% confidence intervals; SE, empirical standard error; SEE, average estimated standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t; M-EE, our proposed method; O-EE, using our proposed multiple imputation but using natural effect model proposed by Mittinty and Vansteelandt[2] when estimating natural effects; M-Heck, using the theory based on Heckman model[17] to do multiple imputation and using our natural effect model when estimating natural effects.

Table S.8: Comparison of three methods in terms of natural effects ($T = 5$, $n = 5000$).

Natural	True	M-EE					O-EE					M-Heck				
effect	value	Bias	RMSE	CP	SE	SEE	Bias	RMSE	CP	SE	SEE	Bias	RMSE	CP	SE	SEE
$n = 5000$																
DE_2	1.368	-0.003	0.042	0.960	0.042	0.041	-0.031	0.051	0.900	0.041	0.042	-0.056	0.071	0.916	0.045	0.055
DE_3	1.370	-0.005	0.042	0.958	0.042	0.041	-0.019	0.040	0.924	0.035	0.037	-0.058	0.073	0.912	0.045	0.055
DE_4	1.370	-0.005	0.042	0.958	0.042	0.041	-0.005	0.046	0.952	0.046	0.045	-0.058	0.073	0.910	0.045	0.055
DE_5	1.370	-0.005	0.042	0.958	0.042	0.041	0.010	0.065	0.938	0.065	0.060	-0.058	0.073	0.912	0.045	0.055
IDE_2	0.162	-0.002	0.014	0.966	0.013	0.014	-0.004	0.012	0.916	0.011	0.010	-0.015	0.025	0.876	0.020	0.019
IDE_3	0.162	-0.0043	0.014	0.966	0.013	0.014	-0.004	0.012	0.950	0.011	0.012	-0.015	0.025	0.874	0.020	0.019
IDE_4	0.162	-0.002	0.014	0.966	0.013	0.018	-0.003	0.016	0.972	0.016	0.016	-0.015	0.025	0.876	0.020	0.019
IDE_5	0.162	-0.003	0.014	0.966	0.013	0.014	-0.002	0.023	0.972	0.023	0.025	-0.016	0.026	0.874	0.020	0.019
TE_2	1.530	-0.005	0.038	0.986	0.038	0.042	-0.035	0.052	0.918	0.039	0.043	-0.071	0.083	0.870	0.044	0.056
TE_3	1.532	-0.008	0.039	0.982	0.038	0.042	-0.023	0.040	0.952	0.033	0.039	-0.073	0.086	0.860	0.044	0.056
TE_4	1.532	-0.008	0.039	0.982	0.038	0.042	-0.008	0.042	0.966	0.041	0.046	-0.073	0.086	0.860	0.044	0.056
TE_5	1.532	-0.008	0.039	0.982	0.038	0.042	-0.007	0.059	0.960	0.059	0.061	-0.073	0.086	0.860	0.044	0.056

Bias, estimated natural effect minus the true value; RMSE, root mean square error; CP, the coverage probability of 95% confidence intervals; SE, empirical standard error; SEE, average estimated standard error; DE_t, direct effect at time t; IDE_t, indirect effect at time t; TE_t, total effect at time t; M-EE, our proposed method; O-EE, using our proposed multiple imputation but using natural effect model proposed by Mittinty and Vansteelandt[2] when estimating natural effects; M-Heck, using the theory based on Heckman model[17] to do multiple imputation and using our natural effect model when estimating natural effects.

Table S.9: Point estimates and 95% CIs of causal effects with 50% artificially generated MNAR data at each age (MCS data set).

	M-V	M-EE		O-EE	
	Estimate	Estimate	95 %CI, length	Estimate	95% CI, length
DE_2	0.463	0.319	(0.087,0.551), 0.464	0.444	(0.290,0.597), 0.307
DE_3	0.240	0.319	(0.087,0.551), 0.464	0.363	(0.050,0.676), 0.626
IDE_2	0.136	0.093	(0.071,0.115), 0.044	0.077	(0.061,0.093), 0.032
IDE_3	0.089	0.093	(0.071,0.115), 0.044	0.098	(0.069,0.127), 0.058
TE_2	0.599	0.412	(0.183,0.640), 0.457	0.521	(0.369,0.672), 0.303
TE_3	0.329	0.412	(0.183,0.640), 0.457	0.461	(0.152,0.770), 0.618

M-V, the method of Mittinty and Vansteelandt[2]; M-EE, our proposed method; O-EE, using our proposed multiple imputation but using natural effect model proposed by Mittinty and Vansteelandt[2] when estimating natural effects; 95%CI, 95% confidence intervals; DE, direct effect; IDE, indirect effect; TE, total effect; Subscripts 2 and 3 correspond to children ages of five and seven years.