Bayesian Inference for Sequential Treatments under Latent Sequential Ignorability
Web Supplementary Material

Federico Ricciardi¹, Alessandra Mattei², and Fabrizia Mealli²
¹ Department of Statistical Science, University College London, UK
² Department of Statistics, Computer Science, Applications, University of Florence, Italy
March 25, 2019

1 MCMC & DA algorithms

The posterior distributions of the parameters are obtained from Markov chain Monte Carlo (MCMC) methods.

Three different algorithms are adopted, depending on which assumption (either LSI, SI/SIL with the specification SI-1 or SI/SIL with the specification SI-2) is used in the inferential analysis.

1.1 Bayesian Inference under LSI

The Markov chain algorithm that we adopt is based on Gibbs sampler and uses the data augmentation method to impute at each step the missing principal stratum indicators and to exploit the complete principal strata-data posterior distribution to update the parameters.

1.1.1 Prior Distributions for Parameters

- \( \alpha^g \sim N(\mu_{\alpha^g}, \sigma_{\alpha^g}^2) \) with \( \mu_{\alpha^g} = 0 \) and \( \sigma_{\alpha^g}^2 = 100 \) for \( g \in \{00, 10, 11\} \);

- \( \gamma_{w_1} \sim N(\mu_{\gamma_{w_1}}, \Sigma_{\gamma_{w_1}}) \) with \( \mu_{\gamma_{w_1}} = \left( \mu_{\gamma_{w_1}}, \mu_{\gamma_{w_1}Y_{1(0)}}, \mu_{\gamma_{w_1}Y_{1(1)}}, \mu_{\gamma_{w_1}Y_{1(0)}Y_{1(1)}} \right) = (0, 0, 0, 0) \) and

\[ \Sigma_{\gamma_{w_1}} = \sigma_{\gamma_{w_1}}^2 I = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix} \]

- \( \beta_{w_1,w_2} \sim N(\mu_{\beta_{w_1,w_2}}, \Sigma_{\beta_{w_1,w_2}}) \) with \( \mu_{\beta_{w_1,w_2}} = \left( \mu_{\beta_{w_1,w_2}}, \mu_{\beta_{w_1,w_2}Y_{1(0)}}, \mu_{\beta_{w_1,w_2}Y_{1(1)}}, \mu_{\beta_{w_1,w_2}Y_{1(0)}Y_{1(1)}} \right) = (0, 0, 0, 0) \)

\[ \Sigma_{\beta_{w_1,w_2}} = \sigma_{\beta_{w_1,w_2}}^2 I = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix} \]
- $\sigma_2^{2}_{w_{1},w_{2}} \sim \text{Scale-inv-\chi}^2(\mathcal{L}_w, \mathcal{L}_w^2) \text{ with } \mathcal{L}_w = 0.002 \text{ and } \mathcal{L}_w^2 = 1$, with $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$.

Let $(G_{(it)}^{(it)}, \theta_{(it)})$ denote the state of the chain at iteration $it$, where $\theta_{(it)} = (\alpha_{(it)}^{\theta}, \beta_{(it)}^{\theta}, \sigma_{(it)}^{\theta}, \gamma_{w, v}^{\theta})$. The state of the chain at iteration $it + 1$ is given by the following steps:

1.1.2 sample $G_{(i+1)}^{(it)}$ according to $Pr \left( G_{i} = 01 | \theta, W_{i1} = 0, Y_{i1}(w_1) = 0, W_{i2}, Y_{i2}(0, W_{i2}) = Y_{i2}^{obs} \right)$

(a) For observations $i \in O(W_{i1}, Y_{i1}^{obs}) = O(0, 0)$ we have

\[
Pr \left( G_{i} = 01 | \theta, W_{i1} = 0, Y_{i1}(w_1) = 0, W_{i2}, Y_{i2}(0, W_{i2}) = Y_{i2}^{obs} \right) \propto \frac{\pi_{i01} \times f_{i01}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{001}^{0})^{W_{i2}} (1 - h_{001}^{0})^{(1-W_{i2})} \right]}{\pi_{i01} \times f_{i01}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{011}^{0})^{W_{i2}} (1 - h_{011}^{0})^{(1-W_{i2})} \right] + \pi_{i00} \times f_{i00}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{000}^{0})^{W_{i2}} (1 - h_{000}^{0})^{(1-W_{i2})} \right]}
\]

(b) For observations $i \in O(W_{i1}, Y_{i1}^{obs}) = O(0, 1)$ we have

\[
Pr \left( G_{i} = 11 | \theta, W_{i1} = 0, Y_{i1}(w_1) = 1, W_{i2}, Y_{i2}(0, W_{i2}) = Y_{i2}^{obs} \right) \propto \frac{\pi_{i11} \times f_{i11}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{011}^{0})^{W_{i2}} (1 - h_{011}^{0})^{(1-W_{i2})} \right]}{\pi_{i01} \times f_{i01}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{011}^{0})^{W_{i2}} (1 - h_{011}^{0})^{(1-W_{i2})} \right] + \pi_{i10} \times f_{i10}^{0, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{100}^{0})^{W_{i2}} (1 - h_{100}^{0})^{(1-W_{i2})} \right]}
\]

(c) For observations $i \in O(W_{i1}, Y_{i1}^{obs}) = O(1, 0)$ we have

\[
Pr \left( G_{i} = 00 | \theta, W_{i1} = 1, Y_{i1}(w_1) = 0, W_{i2}, Y_{i2}(1, W_{i2}) = Y_{i2}^{obs} \right) \propto \frac{\pi_{i00} \times f_{i00}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{100}^{1})^{W_{i2}} (1 - h_{100}^{1})^{(1-W_{i2})} \right]}{\pi_{i00} \times f_{i00}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{100}^{1})^{W_{i2}} (1 - h_{100}^{1})^{(1-W_{i2})} \right] + \pi_{i10} \times f_{i10}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{110}^{1})^{W_{i2}} (1 - h_{110}^{1})^{(1-W_{i2})} \right]}
\]

(d) For observations $i \in O(W_{i1}, Y_{i1}^{obs}) = O(1, 1)$ we have

\[
Pr \left( G_{i} = 01 | \theta, W_{i1} = 1, Y_{i1}(w_1) = 1, W_{i2}, Y_{i2}(1, W_{i2}) = Y_{i2}^{obs} \right) \propto \frac{\pi_{i01} \times f_{i01}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{101}^{1})^{W_{i2}} (1 - h_{101}^{1})^{(1-W_{i2})} \right]}{\pi_{i01} \times f_{i01}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{101}^{1})^{W_{i2}} (1 - h_{101}^{1})^{(1-W_{i2})} \right] + \pi_{i11} \times f_{i11}^{1, w_{2}}(Y_{i2}^{obs}) \times \left[ (h_{111}^{1})^{W_{i2}} (1 - h_{111}^{1})^{(1-W_{i2})} \right]}
\]

1.1.3 Sample the latent variables $G_{i}^{*}$

(a) Sample $G_{i}^{*}(11)$ from $N(\alpha^{11}, 1)$ truncated to $(-\infty, 0)$ if $G_{i} = 11$ and to $(0, \infty)$ if $G_{i} \neq 11$.

(b) Sample $G_{i}^{*}(10)$ from $N(\alpha^{10}, 1)$ truncated to $(-\infty, 0)$ if $G_{i} = 10$ and to $(0, \infty)$ if $G_{i} \neq 10$.

(c) Sample $G_{i}^{*}(00)$ from $N(\alpha^{00}, 1)$ truncated to $(-\infty, 0)$ if $G_{i} = 00$ and to $(0, \infty)$ if $G_{i} \neq 00$. 

2
1.1.4 Sample coefficients $\alpha^g$

(a) Let $N$ be the number of units, sample $\alpha^{11}$ from $N(\mu_{\alpha^{11}}, \sigma_{\alpha^{11}}^2)$ where

$$
\sigma_{\alpha^{11}}^2 = \left( \frac{1}{\sigma_{\alpha^{11}}} + N \right)^{-1}, \quad \mu_{\alpha^{11}} = \sigma_{\alpha^{11}}^2 \left( \frac{\mu_{\alpha^{11}}}{\sigma_{\alpha^{11}}} + \sum_{i=1}^{N} G_i^{(11)} \right).
$$

(b) Let $N^{00,01,10}$ be the number of units with $G_i^{(it+1)} \in \{00, 01, 10\}$. Sample $\alpha^{10}$ from $N(\mu_{\alpha^{10}}, \sigma_{\alpha^{10}})$ where

$$
\sigma_{\alpha^{10}}^2 = \left( \frac{1}{\sigma_{\alpha^{10}}} + N^{00,01,10} \right)^{-1}, \quad \mu_{\alpha^{10}} = \sigma_{\alpha^{10}}^2 \left( \frac{\mu_{\alpha^{10}}}{\sigma_{\alpha^{10}}} + \sum_{i: G_i^{(it+1)} \in \{00,01,10\}} G_i^*(10) \right).
$$

(c) Finally let $N^{00,01}$ be the number of units with $G_i^{(it+1)} \in \{00, 01\}$. Sample $\alpha^{00}$ from $N(\mu_{\alpha^{00}}, \sigma_{\alpha^{00}})$ where

$$
\sigma_{\alpha^{00}}^2 = \left( \frac{1}{\sigma_{\alpha^{00}}} + N^{00,01} \right)^{-1}, \quad \mu_{\alpha^{00}} = \sigma_{\alpha^{00}}^2 \left( \frac{\mu_{\alpha^{00}}}{\sigma_{\alpha^{00}}} + \sum_{i: G_i^{(it+1)} \in \{00,01\}} G_i^*(00) \right).
$$

1.1.5 Sample the latent treatment $W_{i2}^*$

For each $i$ with $G_i = g$ and $W_{i1} = w_1$ sample the latent treatment $W_{i2}^*$ from

$$
N \left( \gamma_{w_1} + \gamma_{w_1} Y_{i1}(0), \gamma_{w_1} Y_{i1}(1) + \gamma_{w_1} Y_{i1}(0) Y_{i1}(1), 1 \right)
$$

truncated to $[0, +\infty)$ if $W_{i2} = 1$ and to $(-\infty, 0)$ if $W_{i2} = 0$.

1.1.6 Sample the coefficient $\gamma_{w_1}$

Let $1 = (1, \ldots, 1)'$ and let $YY$ denote the matrix with $i$-th row equal to $(Y_{i1}(0), Y_{i1}(1), Y_{i1}(0) \times Y_{i1}(1))$. Finally let $YY_{w_1}$ and $W_{2;w_1}^*$ be the sub-matrix of $YY$ and the sub-vector of $W_2^*$ for units with $W_{i1} = w_1$. Sample the coefficients $\gamma_{w_1}$ from

$$
N \left( \mu_{\gamma_{w_1}}, \Sigma_{\gamma_{w_1}} \right)
$$

for $w_1 \in \{0, 1\}$, where

$$
\Sigma_{\gamma_{w_1}} = \left( \Sigma_{\gamma_{w_1}^{-1}} + [1|YY_{w_1}]'[1|YY_{w_1}] \right)^{-1} \quad \text{and} \quad \mu_{\gamma_{w_1}} = \Sigma_{\gamma_{w_1}} \left( \Sigma_{\gamma_{w_1}^{-1}} \mu_{\gamma_{w_1}} + [1|YY_{w_1}]'W_{2;w_1}^* \right)
$$

1.1.7 Sample the coefficient $\beta_{w_1,w_2}$

Let $YY_{w_1,w_2}$ and $Y_{w_1,w_2}^{obs}$ be the sub-matrix of $YY$ and sub-vector of $Y_{w_2}^{obs}$ for units with $W_{i1} = w_1$ and $W_{i2} = w_2$. Sample the coefficients $\beta_{w_1,w_2}$ from

$$
N \left( \mu_{\beta_{w_1,w_2}}, \Sigma_{\beta_{w_1,w_2}} \right)
$$
for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$\Sigma_{\beta_{w_1,w_2}} = \left( \Sigma_{\beta_{w_1,w_2}}^{-1} + \frac{1 | YY_{w_1,w_2}|}{\sigma_{w_1,w_2}^2} \right)^{-1}$$

and

$$\mu_{\beta_{w_1,w_2}} = \Sigma_{\beta_{w_1,w_2}} \left( \Sigma_{\beta_{w_1,w_2}}^{-1} \mu_{\beta_{w_1,w_2}} + \frac{1 | YY_{w_1,w_2}| Y_{w_1,w_2}^{obs}}{\sigma_{w_1,w_2}^2} \right)$$

1.1.8 Sample the parameters $\sigma_{w_1,w_2}^2$

Let $RY_2$ be the vector with the $i$-th element equal to

$$Y_{i2}^{obs} - \left( \beta_{w_1,w_2} \gamma_{w_1} + \beta_{w_1,w_2}^Y Y_{i,1}(0) + \beta_{w_1,w_2}^Y Y_{i,1}(1) + \beta_{w_1,w_2} Y_{i,1}^Y Y_{i,1}(0) Y_{i,1}(1) \right).$$

Let $RY_{2;w_1,w_2}$ be the sub vector of $RY_2$ for units with $W_{i1} = w_1$ and $W_{i2} = w_2$. Finally let $N_{w_1,w_2}$ be the number of units in the sample with $W_{i1} = w_1$ and $W_{i2} = w_2$.

Sample the parameters $\sigma_{w_1,w_2}^2$ from

$$\text{Scale-inv-}\chi^2(\nu_{w_1,w_2}, \sigma_{w_1,w_2}^2)$$

for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$\nu_{w_1,w_2} = \nu_{w_1,w_2} + N_{w_1,w_2}$$

and

$$\tau_{w_1,w_2}^2 = \frac{\nu_{w_1,w_2} \sigma_{w_1,w_2}^2}{\nu_{w_1,w_2} + \bar{RY}_{2;w_1,w_2}^2}.$$

1.2 Bayesian Inference under SI/SIL: SI-1 Specification

1.2.1 Prior Distributions for Parameters

- $\alpha^g \sim N\left( \mu_{\alpha^g}, \sigma_{\alpha^g}^2 \right)$ with $\mu_{\alpha^g} = 0$ and $\sigma_{\alpha^g}^2 = 100$ for $g \in \{00, 10, 11\}$;

- $\gamma_{w_1} \sim N\left( \mu_{\gamma_{w_1}}, \Sigma_{\gamma_{w_1}} \right)$ with $\mu_{\gamma_{w_1}} = (\mu_{\gamma_{w_1,1}}, \mu_{\gamma_{w_1,1}}(1)) = (0, 0)$ and

$$\Sigma_{\gamma_{w_1}} = \sigma_{\gamma_{w_1}}^2 I = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

- $\beta_{w_1,w_2} \sim N\left( \mu_{\beta_{w_1,w_2}}, \Sigma_{\beta_{w_1,w_2}} \right)$ with $\mu_{\beta_{w_1,w_2}} = (\mu_{\beta_{w_1,w_2,0}}, \mu_{\beta_{w_1,w_2,1}}, \mu_{\beta_{w_1,w_2,1}}, \mu_{\beta_{w_1,w_2,1}}(1)) = (0, 0, 0, 0)$ and

$$\Sigma_{\beta_{w_1,w_2}} = \sigma_{\beta_{w_1,w_2}}^2 I = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

- $\sigma_{w_1,w_2}^2 \sim \text{Scale-inv-}\chi^2(\nu_{w_1,w_2}, \sigma_{w_1,w_2}^2)$ with $\nu_{w_1,w_2} = 0.002$ and $\sigma_{w_1,w_2}^2 = 1$, with $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$.

Let $(G^{(it)}, \theta^{(it)})$ denote the state of the chain at iteration $it$, where $\theta^{(it)} = (\alpha^{g^{(it)}}, \gamma_{w_1}^{(it)}, \beta_{w_1,w_2}^{(it)}, \sigma_{w_1,w_2}^{2,(it)})$. The state of the chain at iteration $it + 1$ is given by the following steps:
1.2.2 DA: sample $G^{(i+1)}$ according to $P_r\left(G_i \mid \theta, W_{i1} = w_1, Y_{i1}(w_1) = y_{i1}, W_{i2} = w_2, Y_{i2}(w_1, w_2) = Y_{i2}^{\text{obs}}\right)$

(a) For observations $i \in O(W_{i1, i}, Y_{i1}^{\text{obs}}) = O(0, 0)$ we have

$$P_r\left(G_i = 01 \mid \theta, W_{i1} = 0, Y_{i1}(w_1) = 0, W_{i2}, Y_{i2}(0, W_{i2}) = Y_{i2}^{\text{obs}}\right) \propto$$

$$\frac{\pi_{i01} \times f^{0, w_2}_{i01}(Y_{i2}^{\text{obs}}) \times \left(\left(h^{0}_{01}\right)^{W_{i2}} (1 - h^{0}_{00})^{(1-W_{i2})}\right)\right)}{\left(\pi_{i01} \times f^{0, w_2}_{i01}(Y_{i2}^{\text{obs}}) + \pi_{i00} \times f^{0, w_2}_{i00}(Y_{i2}^{\text{obs}})\right) \times \left(\left(h^{0}_{00}\right)^{W_{i2}} (1 - h^{0}_{00})^{(1-W_{i2})}\right)}$$

(b) For observations $i \in O(W_{i1, i}, Y_{i1}^{\text{obs}}) = O(0, 1)$ we have

$$P_r\left(G_i = 11 \mid \theta, W_{i1} = 0, Y_{i1}(w_1) = 1, W_{i2}, Y_{i2}(0, W_{i2}) = Y_{i2}^{\text{obs}}\right) \propto$$

$$\frac{\pi_{i11} \times f^{0, w_2}_{i11}(Y_{i2}^{\text{obs}}) \times \left(\left(h^{0}_{11}\right)^{W_{i2}} (1 - h^{0}_{10})^{(1-W_{i2})}\right)\right)}{\left(\pi_{i11} \times f^{0, w_2}_{i11}(Y_{i2}^{\text{obs}}) + \pi_{i10} \times f^{0, w_2}_{i10}(Y_{i2}^{\text{obs}})\right) \times \left(\left(h^{0}_{10}\right)^{W_{i2}} (1 - h^{0}_{10})^{(1-W_{i2})}\right)}$$

(c) For observations $i \in O(W_{i1, i}, Y_{i1}^{\text{obs}}) = O(1, 0)$ we have

$$P_r\left(G_i = 00 \mid \theta, W_{i1} = 1, Y_{i1}(w_1) = 0, W_{i2}, Y_{i2}(1, W_{i2}) = Y_{i2}^{\text{obs}}\right) \propto$$

$$\frac{\pi_{i00} \times f^{1, w_2}_{i00}(Y_{i2}^{\text{obs}}) \times \left(\left(h^{1}_{00}\right)^{W_{i2}} (1 - h^{1}_{01})^{(1-W_{i2})}\right)\right)}{\left(\pi_{i00} \times f^{1, w_2}_{i00}(Y_{i2}^{\text{obs}}) + \pi_{i10} \times f^{1, w_2}_{i10}(Y_{i2}^{\text{obs}})\right) \times \left(\left(h^{1}_{01}\right)^{W_{i2}} (1 - h^{1}_{01})^{(1-W_{i2})}\right)}$$

(d) For observations $i \in O(W_{i1, i}, Y_{i1}^{\text{obs}}) = O(1, 1)$ we have

$$P_r\left(G_i = 01 \mid \theta, W_{i1} = 1, Y_{i1}(w_1) = 1, W_{i2}, Y_{i2}(1, W_{i2}) = Y_{i2}^{\text{obs}}\right) \propto$$

$$\frac{\pi_{i01} \times f^{1, w_2}_{i01}(Y_{i2}^{\text{obs}}) \times \left(\left(h^{1}_{01}\right)^{W_{i2}} (1 - h^{1}_{01})^{(1-W_{i2})}\right)\right)}{\left(\pi_{i01} \times f^{1, w_2}_{i01}(Y_{i2}^{\text{obs}}) + \pi_{i11} \times f^{1, w_2}_{i11}(Y_{i2}^{\text{obs}})\right) \times \left(\left(h^{1}_{11}\right)^{W_{i2}} (1 - h^{1}_{11})^{(1-W_{i2})}\right)}$$

1.2.3 Sample the latent variables $G_i^*(g)$

(a) Sample $G_i^*(11)$ from $N(\alpha^{11}, 1)$ truncated to $(-\infty, 0)$ if $G_i = 11$ and to $(0, \infty)$ if $G_i \neq 11$.

(b) Sample $G_i^*(10)$ from $N(\alpha^{10}, 1)$ truncated to $(-\infty, 0)$ if $G_i = 10$ and to $(0, \infty)$ if $G_i \neq 10$.

(c) Sample $G_i^*(00)$ from $N(\alpha^{00}, 1)$ truncated to $(-\infty, 0)$ if $G_i = 00$ and to $(0, \infty)$ if $G_i \neq 00$.

1.2.4 Sample coefficients $\alpha^g$

(a) Let $N$ be the number of units, sample $\alpha^{11}$ from $N\left(\mu^{11}, \sigma^{2}_{\alpha^{11}}\right)$ where

$$\sigma^{2}_{\alpha^{11}} = \left(\frac{1}{\sigma^{2}_{\alpha^{11}}} + N\right)^{-1}, \quad \mu^{11} = \sigma^{2}_{\alpha^{11}} \left(\frac{\mu^{11}}{\sigma^{2}_{\alpha^{11}}} + \sum_{i=1}^{N} G_i^*(11)\right).$$
(b) Let $N^{00,01,10}$ be the number of units with $G_i^{(it+1)} \in \{00, 01, 10\}$. Sample $\alpha^{10}$ from $N(\mu_{\alpha^{10}}, \sigma_{\alpha^{10}})$ where

$$
\sigma_{\alpha^{10}}^2 = \left( \frac{1}{\sigma_{\alpha^{10}}^2} + N^{00,01,10} \right)^{-1}, \quad \mu_{\alpha^{10}} = \sigma_{\alpha^{10}}^2 \left( \frac{\mu_{\alpha^{10}}}{\sigma_{\alpha^{10}}^2} + \sum_{i} G_i^{(it+1)}(10) \right).
$$

(c) Finally let $N^{00,01}$ be the number of units with $G_i^{(it+1)} \in \{00, 01\}$. Sample $\alpha^{00}$ from $N(\mu_{\alpha^{00}}, \sigma_{\alpha^{00}})$ where

$$
\sigma_{\alpha^{00}}^2 = \left( \frac{1}{\sigma_{\alpha^{00}}^2} + N^{00,01} \right)^{-1}, \quad \mu_{\alpha^{00}} = \sigma_{\alpha^{00}}^2 \left( \frac{\mu_{\alpha^{00}}}{\sigma_{\alpha^{00}}^2} + \sum_{i} G_i^{(it+1)}(00) \right).
$$

1.2.5 Sample the latent treatment $W_{it}^*$

For each $i$ with $G_i = g$ and $W_{i1} = w_1$ sample the latent treatment $W_{it}^*$ from

$$
N \left( \gamma_{w1} + \gamma_{Y1(w1)}Y_{i1}^{(w1)}, 1 \right)
$$

truncated to $[0, +\infty)$ if $W_{i2} = 1$ and to $(-\infty, 0)$ if $W_{i2} = 0$.

1.2.6 Sample the coefficient $\gamma_{w_1}$

Let $1 = (1, \ldots, 1)'$ and let $Y_1^{obs} = (Y_{1,1,1}, \ldots, Y_{1,N_1})'$. Then let $Y_{1,w1}^{obs}$ be the sub-vector of $Y_1^{obs}$ and let $W_{2:w1}^*$ be the sub-vector of $W_2^*$ for units with $W_{i1} = w_1$. Sample the coefficients $\gamma_{w1}$ from

$$
N \left( \mu_{\gamma_{w1}}, \Sigma_{\gamma_{w1}} \right)
$$

for $w_1 \in \{0, 1\}$, where

$$
\Sigma_{\gamma_{w1}} = \left( \Sigma_{\gamma_{w1}}^{-1} + \left[1 | Y_{1:w1}^{obs} \right]' \left[1 | Y_{1:w1}^{obs} \right] \right)^{-1} \quad \text{and} \quad \mu_{\gamma_{w1}} = \Sigma_{\gamma_{w1}} \left( \Sigma_{\gamma_{w1}}^{-1} \mu_{\gamma_{w1}} + \left[1 | Y_{1:w1}^{obs} \right]' W_{2:w1}^* \right).
$$

1.2.7 Sample the coefficient $\beta_{w_1,w_2}$

Let $YY$ denote the matrix with $i$-th row equal to $(Y_{i1}(0), Y_{i1}(1), Y_{i1}(0) \times Y_{i1}(1))$. Let $Y_{1:w1,w2}^{obs}$ be the sub-matrix of $YY$ and let $Y_{2:w1,w2}^{obs}$ be the sub-vector of $Y_2^{obs}$ for units with $W_{i1} = w_1$ and $W_{i2} = w_2$. Sample the coefficients $\beta_{w1,w2}$ from

$$
N \left( \mu_{\beta_{w1,w2}}, \Sigma_{\beta_{w1,w2}} \right)
$$

for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$
\Sigma_{\beta_{w1,w2}} = \left( \Sigma_{\beta_{w1,w2}}^{-1} + \left[1 | YY_{w1,w2}^{obs} \right]' \left[1 | YY_{w1,w2}^{obs} \right] \right)^{-1}
$$

and

$$
\mu_{\beta_{w1,w2}} = \Sigma_{\beta_{w1,w2}} \left( \Sigma_{\beta_{w1,w2}}^{-1} \mu_{\beta_{w1,w2}} + \left[1 | YY_{w1,w2}^{obs} \right]' Y_{2:w1,w2}^{obs} \right).$$
1.2.8 Sample the parameters $\sigma^2_{w_1,w_2}$

Let $RY_2$ be the vector with the $i$-th element equal to

$$Y_{it}^{obs} - \left( \beta_{w_1,w_2} + \beta_{w_1,w_2}^Y Y_{i,1}(0) + \beta_{w_1,w_2}^Y Y_{i,1}(1) + \beta_{w_1,w_2}^{Y(0)} Y_{i,1}(0) Y_{i,1}(1) \right).$$

Let $RY_{2,w_1,w_2}$ be the sub vector of $RY_2$ for units with $W_{i1} = w_1$ and $W_{i2} = w_2$. Finally let $N^{w_1,w_2}$ be the number of units in the sample with $W_{i1} = w_1$ and $W_{i2} = w_2$. Sample the parameters $\sigma^2_{w_1,w_2}$ from

Scaling-$\chi^2(\nu_{w_1,w_2}, \sigma^2_{w_1,w_2})$

for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$\nu_{w_1,w_2} = \nu_{w_1,w_2} + N^{w_1,w_2}$$

and

$$\tau^2_{w_1,w_2} = \frac{\nu_{w_1,w_2} \sigma^2_{w_1,w_2} + RY'_{2,w_1,w_2} RY_{2,w_1,w_2}}{\nu_{w_1,w_2}}.$$

1.3 Bayesian Inference under SI/SIL: SI-2 Specification

1.3.1 Prior Distributions for Parameters

- $\alpha_{w_1} \sim N(m_{\alpha_{w_1}}, \sigma^2_{\alpha_{w_1}})$ with $m_{\alpha_{w_1}} = 0$ and $\sigma^2_{\alpha_{w_1}} = 100$ for $w_1 \in \{0, 1\}$;

- $\gamma_{w_1} \sim N(m_{\gamma_{w_1}}, \sigma^2_{\gamma_{w_1}})$ with $m_{\gamma_{w_1}} = (\mu_{\gamma_{w_1}}, \mu_{\gamma_{w_1}(w_1)}) = (0, 0)$ and

$$\sigma^2_{\gamma_{w_1}} = \sigma^2_{\gamma_{w_1}} I = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

- $\beta_{w_1,w_2} \sim N(m_{\beta_{w_1,w_2}}, \sigma^2_{\beta_{w_1,w_2}})$ with $m_{\beta_{w_1,w_2}} = (\mu_{\beta_{w_1,w_2}}, \mu_{\beta_{w_1,w_2}(w_1)}) = (0, 0)$ and

$$\sigma^2_{\beta_{w_1,w_2}} = \sigma^2_{\beta_{w_1,w_2}} I = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

- $\sigma^2_{w_1,w_2} \sim \text{Scaling-$\chi^2(\nu_{w_1,w_2}, \tau^2_{w_1,w_2})$}$ with $\nu_{w_1,w_2} = 0.002$ and $\tau^2_{w_1,w_2} = 1$, with $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$.

Let $(G^{(it)}, \theta^{(it)})$ denote the state of the chain at iteration $it$, where $\theta^{(it)} = (\alpha^{(it)}, \gamma^{(it)}, \beta^{(it)}, \sigma^{2,(it)}_{w_2})$. The state of the chain at iteration $it + 1$ is given by the following steps:

1.3.2 Sample the latent variables $Y^{*}_{i1}(w_1)$

For each $i$ with $W_{i1} = w_1$ sample the latent intermediate outcome $Y^{*}_{i1}(w_1)$ from

$$N(\alpha_{w_1}, 1)$$

truncated to $[0, +\infty)$ if $Y_{i1}(w_1) = 1$ and to $(-\infty, 0)$ if $Y_{i1}(w_1) = 0$. 7
1.3.3 Sample coefficients $\alpha_{w_1}$

Let $N_{w_1}$ be the number of units with $W_{i1} = w_1$. For $w_1 \in \{0, 1\}$, sample the coefficient $\alpha_{w_1}$ from $N\left(\mu_{\alpha_{w_1}}, \sigma^2_{\alpha_{w_1}}\right)$ where

$$
\sigma^2_{\alpha_{w_1}} = \left(\frac{1}{\sigma^2_{\alpha_{w_1}}} + N_{w_1}\right)^{-1}, \quad \mu_{\alpha_{w_1}} = \sigma^2_{\alpha_{w_1}} \left(\frac{1}{\sigma^2_{\alpha_{w_1}}} + \sum_{i: W_{i1} = w_1} Y_{i1}(w_1)\right).
$$

1.3.4 Sample the latent treatment $W_{i2}^*$

For each $i$ with $G_i = g$ and $W_{i1} = w_1$ sample the latent treatment $W_{i2}^*$ from $N\left(\gamma_{w_1} + \gamma_{Y_{1i}(w_1)} Y_{i1}(w_1), 1\right)$ truncated to $[0, +\infty)$ if $W_{i2} = 1$ and to $(-\infty, 0)$ if $W_{i2} = 0$.

1.3.5 Sample the coefficient $\gamma_{w_1}$

Let $1 = (1, \ldots, 1)'$ and let $Y_{1}^{obs} = (Y_{1}^{obs}, \ldots, Y_{N1}^{obs})'$. Then let $Y_{1}^{obs}$ be the sub-vector of $Y_{1}^{obs}$ and let $W_{2;w_1}^*$ be the sub-vector of $W_{2}^*$ for units with $W_{1i} = w_1$. Sample the coefficients $\gamma_{w_1}$ from $N\left(\mu_{\gamma_{w_1}}, \Sigma_{\gamma_{w_1}}\right)$ for $w_1 \in \{0, 1\}$, where

$$
\Sigma_{\gamma_{w_1}} = \left(\Sigma^{-1}_{\gamma_{w_1}} + [1|Y_{11}^{obs}]'(1|Y_{11}^{obs})\right)^{-1} \quad \text{and} \quad \mu_{\gamma_{w_1}} = \Sigma_{\gamma_{w_1}} \left(\Sigma^{-1}_{\gamma_{w_1}} \mu_{\gamma_{w_1}}' + [1|Y_{11}^{obs}]'W_{2;w_1}^*\right).
$$

1.3.6 Sample the coefficient $\beta_{w_1, w_2}$

Let $Y_{2;w_1, w_2}^{obs}$ be the sub-vector of $Y_{2}^{obs}$ for units with $W_{1i} = w_1$ and $W_{2i} = w_2$. Sample the coefficients $\beta_{w_1, w_2}$ from $N\left(\mu_{\beta_{w_1, w_2}}, \Sigma_{\beta_{w_1, w_2}}\right)$ for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$
\Sigma_{\beta_{w_1, w_2}} = \left(\Sigma^{-1}_{\beta_{w_1, w_2}} + [1|Y_{1w_1}^{obs}]'(1|Y_{1w_1}^{obs})\right)^{-1} \quad \text{and} \quad \mu_{\beta_{w_1, w_2}} = \Sigma_{\beta_{w_1, w_2}} \left(\Sigma^{-1}_{\beta_{w_1, w_2}} \mu_{\beta_{w_1, w_2}}' + [1|Y_{1w_1}^{obs}]'Y_{2;w_1, w_2}^{obs}\right).$$
1.3.7 Sample the parameters $\sigma^2_{w_1,w_2}$

Let $RY_2$ be the vector with the $i$-th element equal to

$$\left[ Y_{i2}^{obs} - \left( \beta_{w_1,w_2} Y_{i1}^{(w_1)} + \beta_{w_1,w_2}^2 Y_{i1}(w_1) \right) \right].$$

Let $RY_{2;w_1,w_2}$ be the sub vector of $RY_2$ for units with $W_{i1} = w_1$ and $W_{i2} = w_2$. Finally let $N_{w_1,w_2}$ be the number of unit in the sample with $W_{i1} = w_1$ and $W_{i2} = w_2$.

Sample the parameters $\sigma^2_{w_1,w_2}$ from

$$\text{Scale-inv-}\chi^2(\nu_{w_1,w_2}, s^2_{w_1,w_2})$$

for $w_1 \in \{0, 1\}$ and $w_2 \in \{0, 1\}$, where

$$\nu_{w_1,w_2} = \nu_{w_1,w_2}^0 + N_{w_1,w_2}$$

and

$$\tau^2_{w_1,w_2} = \frac{\nu_{w_1,w_2} s^2_{w_1,w_2} + RY_{2;w_1,w_2}^T RY_{2;w_1,w_2}}{\nu_{w_1,w_2}}.$$
# 2 True values for the simulated experiment’s parameters

Table S1: Real values for the data generating processes.

| parameter | LSI | SI |
|-----------|-----|----|
| $\alpha_{00}$ | 0.5244 | 0.5244 |
| $\alpha_{10}$ | 0.1357 | 0.1357 |
| $\alpha_{11}$ | 0.6745 | 0.6745 |
| $\beta_{00,0}$ | 3.00 | 3.00 |
| $\beta_{00,Y0}$ | 2.00 | 2.00 |
| $\beta_{00,Y1}$ | 6.50 | 6.50 |
| $\beta_{00,Y0Y1}$ | 1.50 | 1.50 |
| $\beta_{01,0}$ | 8.00 | 8.00 |
| $\beta_{01,Y0}$ | 3.00 | 3.00 |
| $\beta_{01,Y1}$ | 8.00 | 8.00 |
| $\beta_{01,Y0Y1}$ | 1.50 | 1.50 |
| $\beta_{10,0}$ | 8.00 | 8.00 |
| $\beta_{10,Y0}$ | 7.00 | 7.00 |
| $\beta_{10,Y1}$ | 2.00 | 2.00 |
| $\beta_{10,Y0Y1}$ | 1.50 | 1.50 |
| $\beta_{11,0}$ | 13.00 | 13.00 |
| $\beta_{11,Y0}$ | 10.00 | 10.00 |
| $\beta_{11,Y1}$ | 4.00 | 4.00 |
| $\beta_{11,Y0Y1}$ | 1.50 | 1.50 |
| $\gamma_{0,0}$ | -0.5244 | -0.5244 |
| $\gamma_{0,Y0}$ | 0.5244 | 0.5244 |
| $\gamma_{0,Y1}$ | 1.1989 | 0.00 |
| $\gamma_{0,Y0Y1}$ | 0.1419 | 0.00 |
| $\gamma_{1,0}$ | 0.2533 | 0.2533 |
| $\gamma_{1,Y0}$ | -1.0950 | 0.00 |
| $\gamma_{1,Y1}$ | 0.2711 | 0.2711 |
| $\gamma_{1,Y0Y1}$ | 0.0462 | 0.00 |
| $\sigma^2_{00}$ | 1.50 | 1.50 |
| $\sigma^2_{01}$ | 2.50 | 2.50 |
| $\sigma^2_{10}$ | 2.00 | 2.00 |
| $\sigma^2_{11}$ | 3.50 | 3.50 |

# 3 Results based on Saturated Marginal Structural Models

Results in this Section are obtained using saturated marginal structural models, estimated by means of inverse probability of treatment weighting in a frequentist fashion. Tables show means, standard errors and 95% confidence intervals. Tables S2 and S3 refer to the simulation results when LSI and SI/SIL hold, respectively. Table S4 gives results for the real case. As expected, results derived using MSMs are generally very similar to those under SI-2 specification, even if some discrepancies may arise (e.g., notably, result for $ATE_{01,10}$ for the real case data).
Table S2: Marginal Structural Model estimates: Simulated data under LSI.

| Estimand | True | Estimate | SE  | 95% Confidence Interval |
|----------|------|----------|-----|-------------------------|
| $ATE_{11.00}$ | 12.54 | 12.21    | 0.20| (11.82; 12.60)          |
| $ATE_{11.01}$ | 6.25  | 2.24     | 0.20| (1.86; 2.63)            |
| $ATE_{11.10}$ | 7.54  | 3.64     | 0.20| (3.26; 4.03)            |
| $ATE_{10.00}$ | 5.01  | 8.57     | 0.16| (8.25; 8.89)            |
| $ATE_{01.10}$ | 1.29  | 1.40     | 0.16| (1.08; 1.72)            |
| $ATE_{01.00}$ | 6.29  | 9.97     | 0.17| (9.65; 10.29)           |

Table S3: Marginal Structural Model estimates: Simulated data under SI/SIL.

| Estimand | True | Estimate | SE  | 95% Confidence Interval |
|----------|------|----------|-----|-------------------------|
| $ATE_{11.00}$ | 12.54 | 11.32    | 0.08| (11.17; 11.48)          |
| $ATE_{11.01}$ | 6.25  | 5.70     | 0.10| (5.51; 5.89)            |
| $ATE_{11.10}$ | 7.54  | 6.17     | 0.09| (5.99; 6.34)            |
| $ATE_{10.00}$ | 5.01  | 5.15     | 0.07| (5.02; 5.29)            |
| $ATE_{01.10}$ | 1.29  | 0.47     | 0.09| (0.29; 0.65)            |
| $ATE_{01.00}$ | 6.29  | 5.62     | 0.08| (5.46; 5.78)            |

Table S4: Marginal Structural Model estimates: Real case.

| Estimand | Estimate | SE  | 95% Confidence Interval |
|----------|----------|-----|-------------------------|
| $ATE_{11.00}$ | 6.96    | 2.68| (1.70; 12.21)           |
| $ATE_{11.01}$ | 4.41    | 2.73| (-0.95; 9.77)           |
| $ATE_{11.10}$ | 5.04    | 2.73| (-0.31; 10.38)          |
| $ATE_{10.00}$ | 1.92    | 0.51| (0.92; 2.92)            |
| $ATE_{01.10}$ | 0.63    | 0.74| (-0.83; 2.08)           |
| $ATE_{01.00}$ | 2.55    | 0.56| (1.46; 3.64)            |
4 Sensitivity Analysis with respect to Prior Specification

The simulation results in the main text are derived using proper but weakly informative priors, namely,

- Normal prior distributions with mean zero and variance 100 for the regression coefficients of both the principal stratum membership sub-model and the final outcome sub-models, and scaled-inverse-$\chi^2$ prior distributions with 0.002 degrees of freedom and scale parameter 1 for the variances of the final outcome models (see Section 1).

We investigate the robustness of the results with respect to the specification of prior distributions using both more informative priors as well as less informative priors (we always maintain the assumption that parameters are a priori independent). Specifically, we conduct inference in the simulation study by also using the following alternative specifications for the prior distributions:

- More informative priors: Normal prior distributions with mean zero and variance 25 for the regression coefficients in both the principal stratum membership sub-model and the final outcome sub-models, and scaled-inverse-$\chi^2$ prior distributions with 1 degrees of freedom and scale parameter 3 for the variances of the final outcome models

- Less informative priors: Normal prior distributions with mean zero and variance 400 for the regression coefficients in both the principal stratum membership sub-model and the final outcome sub-models, and scaled-inverse-$\chi^2$ prior distributions with 0.001 degrees of freedom and scale parameter 1 for the variances of the final outcome models

Table S5 and Figures S1-S3 and Table S6 and Figures S4-S6 show the results when data are generated under LSI and SI/LSI, respectively, including results we presented in the paper. The results are robust with respect to the specification of the prior distributions. Different prior specifications change the results only slightly, mainly affecting the posterior variability of the causal estimand: more informative prior distributions lead to larger posterior standard deviations (and larger 95% posterior credible intervals) than less informative prior distributions, especially when inference is conducted under LSI or SI/LSI with specification SI-1. However the substantive results do not change.

5 Additional Simulations

We investigate and compare the role of the alternative assumptions when they hold conditional on an unmeasured confounder, which can be (partially) explained by latent principal stratum membership, $G_i$. Specifically, we enlarge our simulation study by considering two additional sub-scenarios where data are generated under SI/LSI and LSI, respectively, conditioning on a binary covariate, $U_i$, such that (1) it is related to both treatment assignment at time $t = 2$, $W_{i2}$, and the final outcome, $Y_{i2}$; and (2) it is associated with principal stratum membership, $G_i$. The covariate $U_i$ is a confounder of the relationship between $W_{i2}$ and $Y_{i2}$, and it is associated with $G_i$. Formally, we assume that the following SI, SIL and LSI assumptions hold:

**Assumption S1 Sequential Ignorability (SI)**

\[
\begin{align*}
Pr(W_{i1} \mid U_i, Y_{i1}(0), Y_{i1}(1), Y_{i2}(0, 0), Y_{i2}(1, 0), Y_{i2}(0, 1), Y_{i2}(1, 1)) &= Pr(W_{i1}) \\
Pr(W_{i2} \mid U_i, W_{i1}, Y_{i1}^{obs}, Y_{i2}(0, 0), Y_{i2}(1, 0), Y_{i2}(0, 1), Y_{i2}(1, 1)) &= Pr(W_{i2} \mid U_i, W_{i1}, Y_{i1}^{obs})
\end{align*}
\]

**Assumption S2 Sequential Ignorability of Longitudinal Treatment Assignment (SIL)**

\[
Pr(W_i \mid U_i, Y_{i1}(0), Y_{i1}(1), Y_{i2}(0, 0), Y_{i2}(1, 0), Y_{i2}(0, 1), Y_{i2}(1, 1)) = Pr(W_{i1}) \times Pr(W_{i2} \mid W_{i1}, Y_{i1}^{obs}, U_i) .
\]
Table S5: Summary statistics of the posterior distributions of the causal estimands when LSI holds.

Prior distributions: $N(0, 25)$ and Scaled-Inverse-$\chi^2_1(3)$

| Estimand | true | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% |
|----------|------|------|----|-------|-------|------|----|-------|-------|------|----|-------|-------|
| $ATE_{11.00}$ | 12.54 | 12.40 | 0.30 | 11.74 | 13.07 | 12.59 | 0.49 | 11.68 | 13.54 | 12.21 | 0.19 | 11.82 | 12.58 |
| $ATE_{11.01}$ | 6.25  | 6.18  | 0.57 | 5.11  | 7.51  | 4.30  | 0.88 | 2.24  | 6.06  | 2.24  | 0.19 | 1.87  | 2.62  |
| $ATE_{11.10}$ | 7.54  | 7.52  | 0.47 | 6.60  | 8.75  | 4.92  | 0.69 | 3.33  | 6.12  | 3.63  | 0.18 | 3.27  | 3.99  |
| $ATE_{10.00}$ | 5.01  | 4.88  | 0.38 | 3.69  | 5.66  | 7.68  | 0.44 | 6.65  | 8.68  | 8.57  | 0.16 | 8.25  | 8.89  |
| $ATE_{01.10}$ | 1.29  | 1.35  | 0.62 | -0.05 | 2.64  | 0.61  | 0.63 | -0.90 | 1.84  | 1.39  | 0.16 | 1.07  | 1.71  |
| $ATE_{01.00}$ | 6.29  | 6.22  | 0.51 | 4.92  | 7.25  | 8.29  | 0.67 | 6.78  | 9.75  | 9.96  | 0.16 | 9.65  | 10.28 |

Prior distributions: $N(0, 100)$ and Scaled-Inverse-$\chi^2_0.02(1)$

| Estimand | true | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% |
|----------|------|------|----|-------|-------|------|----|-------|-------|------|----|-------|-------|
| $ATE_{11.00}$ | 12.54 | 12.41 | 0.19 | 11.91 | 12.76 | 12.17 | 0.17 | 11.74 | 12.52 | 12.21 | 0.19 | 11.84 | 12.58 |
| $ATE_{11.01}$ | 6.25  | 6.17  | 0.32 | 5.50  | 6.92  | 5.56  | 0.26 | 5.01  | 6.15  | 2.24  | 0.19 | 1.87  | 2.61  |
| $ATE_{11.10}$ | 7.54  | 7.52  | 0.25 | 6.96  | 7.99  | 6.97  | 0.23 | 6.43  | 7.46  | 3.64  | 0.18 | 3.27  | 4.00  |
| $ATE_{10.00}$ | 5.01  | 4.89  | 0.20 | 4.61  | 5.26  | 5.21  | 0.18 | 4.93  | 5.51  | 8.57  | 0.16 | 8.25  | 8.88  |
| $ATE_{01.10}$ | 1.29  | 1.33  | 0.32 | 0.44  | 2.14  | 1.41  | 0.28 | 0.79  | 2.05  | 1.40  | 0.16 | 1.09  | 1.71  |
| $ATE_{01.00}$ | 6.29  | 6.22  | 0.27 | 5.57  | 6.77  | 6.61  | 0.22 | 6.21  | 6.96  | 9.97  | 0.16 | 9.64  | 10.28 |

Prior distributions: $N(0, 400)$ and Scaled-Inverse-$\chi^2_0.001(1)$

| Estimand | true | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% | mean | sd | 2.5%  | 97.5% |
|----------|------|------|----|-------|-------|------|----|-------|-------|------|----|-------|-------|
| $ATE_{11.00}$ | 12.54 | 12.42 | 0.13 | 12.19 | 12.67 | 12.03 | 0.18 | 11.68 | 12.35 | 12.21 | 0.19 | 11.84 | 12.59 |
| $ATE_{11.01}$ | 6.25  | 6.20  | 0.17 | 5.91  | 6.46  | 4.42  | 0.20 | 4.03  | 4.78  | 2.24  | 0.19 | 1.87  | 2.61  |
| $ATE_{11.10}$ | 7.54  | 7.52  | 0.14 | 7.29  | 7.75  | 6.15  | 0.20 | 5.75  | 6.48  | 3.64  | 0.18 | 3.29  | 4.00  |
| $ATE_{10.00}$ | 5.01  | 4.90  | 0.13 | 4.70  | 5.11  | 5.88  | 0.13 | 5.69  | 6.09  | 8.57  | 0.16 | 8.25  | 8.88  |
| $ATE_{01.10}$ | 1.29  | 1.32  | 0.17 | 1.07  | 1.58  | 1.73  | 0.16 | 1.45  | 1.99  | 1.40  | 0.16 | 1.09  | 1.72  |
| $ATE_{01.00}$ | 6.29  | 6.22  | 0.14 | 6.03  | 6.44  | 7.61  | 0.15 | 7.31  | 7.89  | 9.97  | 0.16 | 9.66  | 10.29 |
Table S6: Summary statistics of the posterior distributions of the causal estimands when SI/SIL holds.

Prior distributions: $N(0, 100)$ and Scaled-Inverse-$\chi^2_{0.002}(1)$

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
|          |      | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% |
| $ATE_{11,00}$ | 12.54 | 12.52 | 0.25 | 12.00 | 12.98 | 12.62 | 0.26 | 12.13 | 13.08 | 12.42 | 0.18 | 12.06 | 17.77 |
| $ATE_{11,01}$ | 6.25  | 6.16 | 0.31 | 5.51  | 6.91  | 6.31 | 0.31 | 5.60  | 7.00  | 6.14 | 0.21 | 5.73  | 6.55  |
| $ATE_{11,10}$ | 7.54  | 7.49 | 0.27 | 6.85  | 8.12  | 7.56 | 0.28 | 6.88  | 8.22  | 7.53 | 0.19 | 7.14  | 7.90  |
| $ATE_{10,00}$ | 5.01  | 5.02 | 0.17 | 4.65  | 5.31  | 5.06 | 0.17 | 4.68  | 5.35  | 4.89 | 0.18 | 4.55  | 5.23  |
| $ATE_{01,10}$ | 1.29  | 1.33 | 0.25 | 0.62  | 1.90  | 1.24 | 0.24 | 0.66  | 1.85  | 1.39 | 0.21 | 0.99  | 1.79  |
| $ATE_{01,00}$ | 6.29  | 6.35 | 0.23 | 5.66  | 6.84  | 6.30 | 0.21 | 5.78  | 6.80  | 6.28 | 0.18 | 5.92  | 6.63  |

Prior distributions: $N(0, 25)$ and Scaled-Inverse-$\chi^2_{1}(3)$

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
|          |      | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% |
| $ATE_{11,00}$ | 12.54 | 12.65 | 0.46 | 11.44 | 13.63 | 12.58 | 0.47 | 11.38 | 13.56 | 12.41 | 0.18 | 12.06 | 12.76 |
| $ATE_{11,01}$ | 6.25  | 6.32 | 0.54 | 5.07  | 7.39  | 6.28 | 0.58 | 4.98  | 7.36  | 6.13 | 0.21 | 5.72  | 6.55  |
| $ATE_{11,10}$ | 7.54  | 7.38 | 0.50 | 6.13  | 8.39  | 7.54 | 0.51 | 6.31  | 8.57  | 7.52 | 0.20 | 7.13  | 7.90  |
| $ATE_{10,00}$ | 5.01  | 5.27 | 0.27 | 4.66  | 5.95  | 5.04 | 0.30 | 4.32  | 5.72  | 4.89 | 0.18 | 4.54  | 5.23  |
| $ATE_{01,10}$ | 1.29  | 1.06 | 0.39 | 0.25  | 1.82  | 1.26 | 0.46 | 0.35  | 2.13  | 1.39 | 0.21 | 0.97  | 1.80  |
| $ATE_{01,00}$ | 6.29  | 6.33 | 0.34 | 5.61  | 7.00  | 6.30 | 0.41 | 5.51  | 7.06  | 6.28 | 0.19 | 5.91  | 6.64  |

Prior distributions: $N(0, 400)$ and Scaled-Inverse-$\chi^2_{0.001}(1)$

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
|          |      | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% | mean | sd   | 2.5% | 97.5% |
| $ATE_{11,00}$ | 12.54 | 12.92 | 0.17 | 12.64 | 13.21 | 12.88 | 0.18 | 12.39 | 13.18 | 12.42 | 0.18 | 12.06 | 12.78 |
| $ATE_{11,01}$ | 6.25  | 6.71 | 0.21 | 6.35  | 7.11  | 6.56 | 0.20 | 6.22  | 6.90  | 6.14 | 0.22 | 5.72  | 6.57  |
| $ATE_{11,10}$ | 7.54  | 7.56 | 0.14 | 7.38  | 7.80  | 7.53 | 0.19 | 7.19  | 7.85  | 7.53 | 0.20 | 7.14  | 7.92  |
| $ATE_{10,00}$ | 5.01  | 5.36 | 0.13 | 5.12  | 5.59  | 5.36 | 0.12 | 5.14  | 5.57  | 4.89 | 0.18 | 4.54  | 5.23  |
| $ATE_{01,10}$ | 1.29  | 0.85 | 0.17 | 0.52  | 1.17  | 0.97 | 0.15 | 0.69  | 1.23  | 1.40 | 0.21 | 0.98  | 1.80  |
| $ATE_{01,00}$ | 6.29  | 6.21 | 0.13 | 5.97  | 6.44  | 6.32 | 0.13 | 6.09  | 6.56  | 6.28 | 0.19 | 5.93  | 6.65  |
Assumption S3  *Latent Sequential Ignorability (LSI)*

\[
Pr(W_i | U_i, Y_{i1}(0), Y_{i1}(1), Y_{i2}(0, 0), Y_{i2}(1, 0), Y_{i2}(0, 1), Y_{i2}(1, 1)) = Pr(W_{i1}) \times Pr(W_{i2} | W_{i1}, Y_{i1}(0), Y_{i1}(1), U_i).
\]

We consider \(U_i\) as an unmeasured confounder, and thus we conduct inference by ignoring it, that assuming that SI/SIL and LSI hold without conditioning on \(U_i\) (Assumptions 2, 3 and 4 in the main text).

The data generating processes described in Section 6.1 are modified as follows. The treatment at time 1, \(W_{i1}\), is randomly assigned with probability \(h_i = 0.5\).

**Model for principal stratum membership.**

\[
\begin{align*}
G_i(11) &= 1 \quad \text{if } G_i^*(11) \equiv \alpha_{11} + \alpha_{11} U_i + \epsilon_{i,11} \leq 0, \\
G_i(00) &= 1 \quad \text{if } G_i^*(11) > 0 \text{ and } G_i^*(00) \equiv \alpha_{00} + \alpha_{00} U_i + \epsilon_{i,00} \leq 0, \\
G_i(10) &= 1 \quad \text{if } G_i^*(11) > 0 G_i^*(00) > 0 \text{ and } G_i^*(10) \equiv \alpha_{10} + \alpha_{10} U_i + \epsilon_{i,10} \leq 0,
\end{align*}
\]

where \(\epsilon_{i,11} \sim N(0,1), \epsilon_{i,00} \sim N(0,1), \text{ and } \epsilon_{i,10} \sim N(0,1)\) independently. Therefore

\[
\pi_{i11} = 1 - \Phi(\alpha_{11} + \alpha_{11} U_i) \quad \pi_{i00} = \Phi(\alpha_{11} + \alpha_{11} U_i) \left[1 - \Phi(\alpha_{00} + \alpha_{00} U_i)\right] \\
\pi_{i10} = \Phi(\alpha_{11} + \alpha_{11} U_i) \left[1 - \Phi(\alpha_{10} + \alpha_{10} U_i)\right]
\]

and \(\pi_{i01} = 1 - \sum_{g \in \{11,00,10\}} \pi_{ig} = \Phi(\alpha_{11} + \alpha_{11} X_i) \Phi(\alpha_{00} + \alpha_{00} U_i) \Phi(\alpha_{10} + \alpha_{10} U_i),\) where \(\Phi(\cdot)\) is the cumulative distribution function of the standard Normal distribution.

**Model for the treatment assignment at time \(t = 2\).**

Under LSI (Assumption S3), we assume the following probit model for the treatment assignment at time \(t = 2:\)

\[
W_{i2} = 1 \text{ if } W_{i2} \equiv \gamma_{w1} + \gamma_{w1}(0)Y_{i1}(0) + \gamma_{w1}(1)Y_{i1}(1) + \gamma_{w1}(0)Y_{i1}(1)Y_{i1}(0)Y_{i1}(1) + \gamma_U U_i + \epsilon_{i,W2} > 0,
\]

under which, the treatment assignment probabilities at time \(t = 2\) are

\[
h_{i1}^{w1} = \begin{cases} 
\Phi(\gamma_{w1} + \gamma_U U_i) & \text{if } W_{i1} = w_1 \text{ and } G_i = 00; \\
\Phi(\gamma_{w1} + \gamma_{w1}(0) + \gamma_U U_i) & \text{if } W_{i1} = w_1 \text{ and } G_i = 10; \\
\Phi(\gamma_{w1} + \gamma_{w1}(1) + \gamma_U U_i) & \text{if } W_{i1} = w_1 \text{ and } G_i = 01; \\
\Phi(\gamma_{w1} + \gamma_{w1}(0) + \gamma_{w1}(1) + \gamma_{w1}(0)Y_{i1}(1) + \gamma_U U_i) & \text{if } W_{i1} = w_1 \text{ and } G_i = 11;
\end{cases}
\]

with \(w_1 = 0, 1\). Under SI/SIL (Assumptions S1 and S2), we impose:

\[
\gamma_{Y_i(0)} = \gamma_{Y_i(1)} = \gamma_{Y_i(0)Y_i(1)} = \gamma_{Y_i(1)Y_i(1)} = 0
\]
Table S7: True values for the parameters governing the association of \( U_i \) with \( W_{i2} \) and \( Y_{i2} \), and \( G_i \).

| Parameters | Assocation of \( U_i \) with \( G_i \) |
|------------|---------------------------------|
| \( \alpha_{U}^{11} \) | 0.20 0.45 |
| \( \alpha_{U}^{00} \) | 0.10 0.25 |
| \( \alpha_{U}^{10} \) | 0.02 0.15 |

Table S8: Summary statistics of the posterior distributions of the causal estimands when LSI (Assumption S3) holds.

| Estimand | true | LSI | SI | SI-2 |
|----------|------|-----|----|-----|
| \( \text{ATE}_{11,00} \) | 12.43 | 12.50 0.25 12.20 12.93 | 13.30 0.26 12.94 13.74 | 14.68 0.18 14.33 15.03 |
| \( \text{ATE}_{11,01} \) | 6.16 | 6.23 0.33 5.36 7.15 | 5.43 0.31 4.73 6.06 | 5.02 0.18 4.67 5.37 |
| \( \text{ATE}_{11,10} \) | 7.47 | 7.58 0.26 7.16 8.04 | 7.91 0.27 7.49 8.37 | 8.26 0.21 7.84 8.65 |
| \( \text{ATE}_{10,00} \) | 4.96 | 4.92 0.15 4.61 5.21 | 5.38 0.12 5.13 5.63 | 6.42 0.18 6.07 6.78 |
| \( \text{ATE}_{01,10} \) | 1.30 | 1.35 0.26 0.82 1.85 | 2.49 0.23 2.03 2.88 | 3.24 0.18 2.87 3.61 |
| \( \text{ATE}_{01,00} \) | 6.26 | 6.27 0.24 5.86 6.69 | 7.87 0.22 7.49 8.23 | 9.66 0.16 9.36 9.97 |
LSI with the omitted variable $U_i$ always leads to valid inference about the causal estimands of interests irrespective of the strength of association between $U_i$ and $G_i$. Conversely, assuming SI/SIL (without conditioning on $U_i$), when LSI given $U_i$ actually holds, yields to completely wrong inferences (see Table S8 and Figures S7 and S8).

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is not strong, both LSI and SI/SIL with a specification of type SI-1 (with the omitted variable $U_i$) lead to valid inferences about the causal estimands of interest, whereas SI/SIL with a specification of type SI-2 may yield to completely wrong inferences.

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is strong, results are less stringent. Nevertheless there is some evidence that LSI and SI/SIL with a specification of type SI-1 perform better than SI/SIL with a specification of type SI-2 in the presence of the unmeasured confounder $U_i$.  

### Table S9: Summary statistics of the posterior distributions of the causal estimands when SI/SIL holds.

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
| $ATE_{11.00}$ | 12.43 | 12.52 | 0.27 | 11.84 | 13.23 |
| $ATE_{11.01}$ | 6.16 | 6.05 | 0.31 | 5.31 | 6.82 |
| $ATE_{11.10}$ | 7.47 | 7.16 | 0.30 | 6.43 | 7.90 |
| $ATE_{10.00}$ | 4.96 | 5.37 | 0.19 | 4.86 | 5.73 |
| $ATE_{01.10}$ | 1.30 | 1.11 | 0.21 | 0.55 | 1.68 |
| $ATE_{01.00}$ | 6.26 | 6.48 | 0.19 | 6.02 | 6.82 |

LSI with the omitted variable $U_i$ always leads to valid inference about the causal estimands of interests irrespective of the strength of association between $U_i$ and $G_i$. Conversely, assuming SI/SIL (without conditioning on $U_i$), when LSI given $U_i$ actually holds, yields to completely wrong inferences (see Table S8 and Figures S7 and S8).

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is not strong, both LSI and SI/SIL with a specification of type SI-1 (with the omitted variable $U_i$) lead to valid inferences about the causal estimands of interest, whereas SI/SIL with a specification of type SI-2 may yield to completely wrong inferences.

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is strong, results are less stringent. Nevertheless there is some evidence that LSI and SI/SIL with a specification of type SI-1 perform better than SI/SIL with a specification of type SI-2 in the presence of the unmeasured confounder $U_i$.  

### Table S9: Summary statistics of the posterior distributions of the causal estimands when SI/SIL holds.

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
| $ATE_{11.00}$ | 12.17 | 13.15 | 0.31 | 12.08 | 13.85 |
| $ATE_{11.01}$ | 5.91 | 6.86 | 0.35 | 5.81 | 7.67 |
| $ATE_{11.10}$ | 7.40 | 7.33 | 0.36 | 6.30 | 8.20 |
| $ATE_{10.00}$ | 4.76 | 5.82 | 0.21 | 5.24 | 6.26 |
| $ATE_{01.10}$ | 1.49 | 0.47 | 0.27 | -0.22 | 1.14 |
| $ATE_{01.00}$ | 6.25 | 6.29 | 0.20 | 5.60 | 6.76 |

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
| $ATE_{11.00}$ | 12.70 | 12.06 | 0.35 | 11.64 | 13.26 |
| $ATE_{11.01}$ | 6.15 | 6.06 | 0.33 | 5.81 | 6.85 |
| $ATE_{11.10}$ | 7.32 | 7.54 | 0.33 | 6.43 | 8.26 |
| $ATE_{10.00}$ | 5.23 | 5.16 | 0.19 | 4.86 | 5.49 |
| $ATE_{01.10}$ | 1.17 | 1.48 | 0.25 | 0.81 | 2.09 |
| $ATE_{01.00}$ | 6.39 | 6.64 | 0.22 | 6.03 | 7.07 |

LSI with the omitted variable $U_i$ always leads to valid inference about the causal estimands of interests irrespective of the strength of association between $U_i$ and $G_i$. Conversely, assuming SI/SIL (without conditioning on $U_i$), when LSI given $U_i$ actually holds, yields to completely wrong inferences (see Table S8 and Figures S7 and S8).

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is not strong, both LSI and SI/SIL with a specification of type SI-1 (with the omitted variable $U_i$) lead to valid inferences about the causal estimands of interest, whereas SI/SIL with a specification of type SI-2 may yield to completely wrong inferences.

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is strong, results are less stringent. Nevertheless there is some evidence that LSI and SI/SIL with a specification of type SI-1 perform better than SI/SIL with a specification of type SI-2 in the presence of the unmeasured confounder $U_i$.  

### Table S9: Summary statistics of the posterior distributions of the causal estimands when SI/SIL holds.

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
| $ATE_{11.00}$ | 11.88 | 11.58 | 0.20 | 11.54 | 12.23 |
| $ATE_{11.01}$ | 5.47 | 5.15 | 0.20 | 5.10 | 5.86 |
| $ATE_{11.10}$ | 5.82 | 5.52 | 0.19 | 5.45 | 6.20 |
| $ATE_{10.00}$ | 6.06 | 5.60 | 0.17 | 5.73 | 6.41 |
| $ATE_{01.10}$ | 0.35 | 0.20 | 0.20 | 0.04 | 0.74 |
| $ATE_{01.00}$ | 6.41 | 6.07 | 0.17 | 6.07 | 6.75 |

LSI with the omitted variable $U_i$ always leads to valid inference about the causal estimands of interests irrespective of the strength of association between $U_i$ and $G_i$. Conversely, assuming SI/SIL (without conditioning on $U_i$), when LSI given $U_i$ actually holds, yields to completely wrong inferences (see Table S8 and Figures S7 and S8).

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is not strong, both LSI and SI/SIL with a specification of type SI-1 (with the omitted variable $U_i$) lead to valid inferences about the causal estimands of interest, whereas SI/SIL with a specification of type SI-2 may yield to completely wrong inferences.

When SI/SIL (conditional on $U_i$) is the true assumption underlying the data generation process and the association between $U_i$ and $G_i$ is strong, results are less stringent. Nevertheless there is some evidence that LSI and SI/SIL with a specification of type SI-1 perform better than SI/SIL with a specification of type SI-2 in the presence of the unmeasured confounder $U_i$.  

### Table S9: Summary statistics of the posterior distributions of the causal estimands when SI/SIL holds.

| Estimand | true | LSI | SI-1 | SI-2 |
|----------|------|-----|------|------|
| $ATE_{11.00}$ | 11.83 | 11.50 | 0.17 | 11.50 | 12.16 |
| $ATE_{11.01}$ | 5.67 | 5.29 | 0.19 | 5.29 | 6.04 |
| $ATE_{11.10}$ | 5.99 | 5.61 | 0.19 | 5.61 | 6.36 |
| $ATE_{10.00}$ | 5.84 | 5.49 | 0.17 | 5.49 | 6.16 |
| $ATE_{01.10}$ | 0.32 | 0.20 | 0.20 | -0.06 | 0.71 |
| $ATE_{01.00}$ | 6.16 | 5.84 | 0.17 | 5.84 | 6.50 |
Posterior density functions for the ATEs

Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S1: LSI scenario. Prior distributions: $N(0, 400)$ and Scaled-Inverse-$\chi^2_{0.001}(1)$.
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S2: LSI scenario. Prior distributions: $N(0, 100)$ and Scaled-Inverse-$\chi^2_{0.002}(1)$.
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE

**Posterior density functions for the assignment probabilities (inference conducted under LSI).**

\[
\begin{align*}
\Pr(W_2 = 1 | G=00, W_1=0) & \text{ vs } \Pr(W_2 = 1 | G=01, W_1=0) \\
\Pr(W_2 = 1 | G=00, W_1=1) & \text{ vs } \Pr(W_2 = 1 | G=10, W_1=1) \\
\Pr(W_2 = 1 | G=01, W_1=1) & \text{ vs } \Pr(W_2 = 1 | G=11, W_1=1)
\end{align*}
\]

Figure S3: LSI scenario. Prior distributions: \( N(0, 25) \) and Scaled-Inverse-\( \chi^2_1(3) \).
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

**Figure S4:** SI/SIL scenario. Prior distributions: $N(0, 400)$ and Scaled-Inverse-$\chi^2_{0.001}(1)$. 

---

Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

**Posterior density functions for the assignment probabilities (inference conducted under LSI).**

Pr($W_2 = 1 | G=00, W_1=0$) vs Pr($W_2 = 1 | G=01, W_1=0$)

Pr($W_2 = 1 | G=10, W_1=0$) vs Pr($W_2 = 1 | G=11, W_1=0$)

Pr($W_2 = 1 | G=00, W_1=1$) vs Pr($W_2 = 1 | G=10, W_1=1$)

Pr($W_2 = 1 | G=01, W_1=1$) vs Pr($W_2 = 1 | G=11, W_1=1$)
Posterior density functions for the ATEs

Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S5: SI/SIL scenario. Prior distributions: $N(0, 100)$ and Scaled-Inverse-$\chi^2_{0.002}(1)$. 

22
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S6: SI/SIL scenario. Prior distributions: $N(0, 25)$ and Scaled-Inverse-$\chi^2_1(3)$. 

\[ \text{Pr}(W_2 = 1 | G = 00, W_1 = 0) \text{ vs } \text{Pr}(W_2 = 1 | G = 01, W_1 = 0) \] 
\[ \text{Pr}(W_2 = 1 | G = 10, W_1 = 0) \text{ vs } \text{Pr}(W_2 = 1 | G = 11, W_1 = 0) \] 
\[ \text{Pr}(W_2 = 1 | G = 00, W_1 = 1) \text{ vs } \text{Pr}(W_2 = 1 | G = 10, W_1 = 1) \] 
\[ \text{Pr}(W_2 = 1 | G = 01, W_1 = 1) \text{ vs } \text{Pr}(W_2 = 1 | G = 11, W_1 = 1) \]
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE.

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S7: LSI scenario. Weak association of $U_i$ with $G_i$
Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S8: LSI scenario. Strong association of $U_i$ with $G_i$
Posterior density functions for the ATEs

Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Pr(W_2 = 1 | G = 00, W_1 = 0) vs Pr(W_2 = 1 | G = 01, W_1 = 0)

Pr(W_2 = 1 | G = 00, W_1 = 1) vs Pr(W_2 = 1 | G = 10, W_1 = 1)

Pr(W_2 = 1 | G = 01, W_1 = 0) vs Pr(W_2 = 1 | G = 11, W_1 = 0)

Pr(W_2 = 1 | G = 01, W_1 = 1) vs Pr(W_2 = 1 | G = 11, W_1 = 1)

Figure S9: SI/SIL scenario. Weak association of $U_i$ with $G_i$.
Posterior density functions for the ATEs

Inference using LSI (solid), SI-1 (dotted) and SI-2 (dashed). The vertical solid line indicates the true value of the ATE

Posterior density functions for the assignment probabilities (inference conducted under LSI).

Figure S10: SI/SIL scenario. Strong association of $U_i$ with $G_i$