SUPPLEMENTARY MATERIAL: SEPARABLE EFFECTS FOR CAUSAL INFERENCE IN THE PRESENCE OF COMPETING RISKS

MATS J. STENSRUD[1,2], JESSICA G. YOUNG[3], VANESSA DIDELEZ[4,5], JAMES M. ROBINS[1,6], MIGUEL A. HERNÁN[1,6]

[1] *Department of Epidemiology, Harvard T. H. Chan School of Public Health, USA*

[2] *Department of Biostatistics, University of Oslo, Norway*

[3] *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, USA*

[4] *Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany*

[5] *Faculty of Mathematics / Computer Science, University of Bremen, Germany*

[6] *Department of Biostatistics, Harvard T. H. Chan School of Public Health, USA*

Date: June 5, 2020.

Appendix A. Some intuition about the magnitude of the separable direct effects.

Consider the following scenarios:

- Scenario 1: A has a null direct effect on the competing event ($A \nrightarrow D_k$), and the separable direct effect is equal to the total effect.
- Scenario 2: A has a null direct effect on the event of interest ($A \nrightarrow Y_k$), and the indirect effect is equal to the total effect.
- Scenario 3: A has an average harmful (positive) total effect on both Y_k and D_k. The separable direct effects $\Pr(Y_{k+1}^{a_Y=1,a_D} = 1)$ vs. $\Pr(Y_{k+1}^{a_Y=0,a_D} = 1)$ are harmful (positive), and the separable indirect effects $\Pr(Y_{k+1}^{a_Y,a_D=1} = 1)$ vs. $\Pr(Y_{k+1}^{a_Y,a_D=0} = 1)$ are protective (negative).
- Scenario 4: A has an average harmful (positive) total effect on Y_k and a protective (negative) total effect on D_k, and the separable direct effects $\Pr(Y_{k+1}^{a_Y=1,a_D} = 1)$ vs. $\Pr(Y_{k+1}^{a_Y=0,a_D} = 1)$ are harmful (positive), and the separable indirect effects $\Pr(Y_{k+1}^{a_Y,a_D=1} = 1)$ vs. $\Pr(Y_{k+1}^{a_Y,a_D=0} = 1)$ are harmful (positive).

To provide some intuition about the magnitude of the separable effects across these scenarios, we conducted simulations under the following data generating process:

(1) Draw $L_1 \sim \text{Bernoulli}[p = 0.25]$.

(2) Draw $A_Y \sim \text{Bernoulli}[p = 0.5]$.

(3) Draw $A_D \sim \text{Bernoulli}[p = 0.5]$.

(4) Define $A = a$ if $A_Y = a$ and $A_D = a$.

(5) Set $D_0 = Y_0 = 0$.

(6) For each $k \in \{0, K\}$,

- if $D_k = Y_k = 0$,

  draw $D_{k+1} \sim$ Bernoulli$[p = \psi_k(A_Y, A_D, L_1, L_2)]$, where

$$\psi_k(A_Y, L_1) = \text{expit}(\omega_0 + \omega_{1,k}k + \omega_2 A_Y + \omega_3 L_1)$$

  if $D_{k+1} = 0$,

  draw $Y_{k+1} \sim$ Bernoulli$(p = \lambda_k(A_D, L_1))$, where

$$\lambda_k(A_D, L_1) = \text{expit}(\xi_0 + \xi_{1,k}k + \xi_2 A_D + \xi_3 L_1)$$

  if $D_{k+1} = 1$, set $Y_{k+1} = 0$.

- else, define $D_{k+1} = D_k$,$Y_{k+1} = Y_k$.

Scenario 1 is illustrated in Figure 1a, which was generated using the coefficients from the first row of Table 1.

Scenario 2 illustrated in Figure 1b, which was generated using the coefficients from the second row of Table 1.

Scenario 3 is illustrated in Figure 1c, which was generated using the coefficients from the third row of Table 1.

Scenario 4 is illustrated in Figure 1d, where data were generated from the forth row of Table 1.

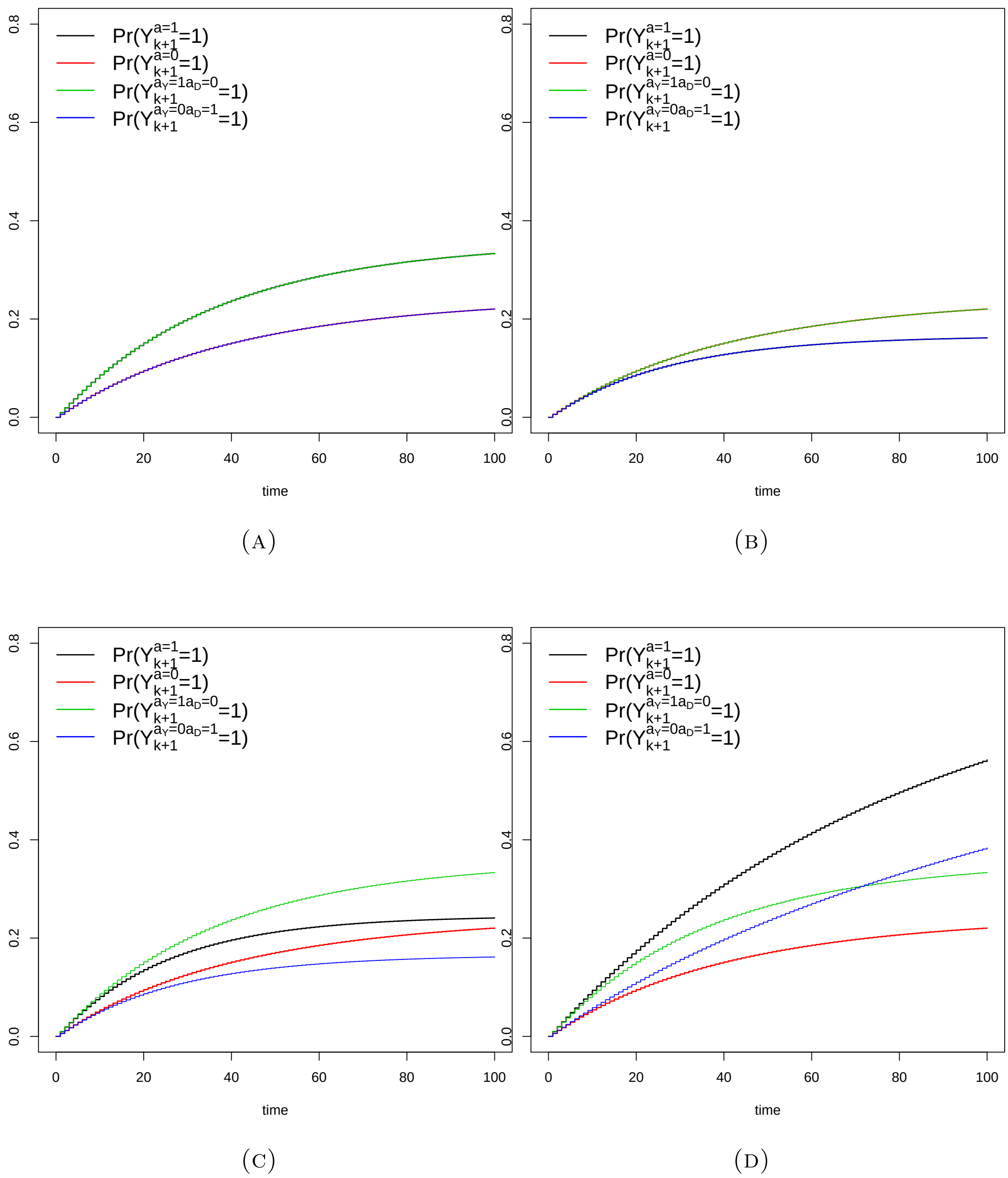

FIGURE 1. Counterfactual outcomes under the data generating mechanisms from Table 1. In the upper left panel, there is perfect overlap between the black and green curves, and of the red and blue curves. In the upper right panel, there is perfect overlap between the red and green curves, and of the black and blue curves.

Scenario	α_Y	ω_1	ω_2	ω_3	α_D	ξ_1	ξ_2	ξ_3
1	0.01	0	10	5	0.03	0	0	5
2	0.01	0	0	5	0.03	0	5	5
3	0.01	0	10	5	0.03	0	5	5
4	0.01	0	10	5	0.03	0	-5	5

TABLE 1. Coefficients for the data generating mechanism of the examples in Appendix A.

To provide additional intuition about the magnitude of the separable effects, it may be helpful to consider two hypothetical sets of individuals (Table 2).

First, define the set Q_k of individuals such that $i \in Q_k$ if i would experience the competing event at time $t_i < k$ under full treatment (that is, $A_Y = 1, A_D = 1$), and would experience the event of interest at a time s_i, where $t_i < s_i < k$, under the hypothetical treatment $A_Y = 1, A_D = 0$, see Table 2. Heuristically, this happens if the hypothetical treatment delays the competing event such that the event of interest is allowed to occur. If Q_k comprises a large fraction of the population, we would expect the total effect and the separable direct effect to be different at k, because competing events would make it impossible for the event of interest to occur under full treatment, but not under the hypothetical treatment.

Second, define the set of individuals R_k such that all individuals $j \in R_k$ experience the competing event at time $t_j < k$ under full treatment, but would either experience the competing event at s_j, where $s_j < k$, or not experience any event before k under the hypothetical treatment. That is, the subjects in R_k will not experience the event of interest before k under the hypothetical treatment, regardless of the time at which the competing event occurs. If R_k comprises a large fraction of the population, the total effect and the separable direct effect on the event of interest will be close at k.

TABLE 2. Outcomes at time k in subgroups Q_k and R_k.

Treatment	**Outcomes at k in Q_k**	**Outcomes at k in R_k**
$A_Y = 1, A_D = 1$	$(Y_k = 0, D_k = 1)$	$(Y_k = 0, D_k = 1)$
$A_Y = 1, A_D = 0$	$(Y_k = 1, D_k = 0)$	$(Y_k = 0, D_k = 1)$ or $(Y_k = 0, D_k = 0)$

Appendix B. Conditional Independencies that imply the dismissible component conditions.

We expressed the dismissible component conditions $\Delta 1$ and $\Delta 2$ in terms of equalities of hazard functions. We now show that these equalities are implied by certain counterfactual independencies that can be read directly off of successive single world transformation of a causal DAG.

Hypothetical trial. Suppose that each component of A is randomly assigned in a hypothetical 4-arm trial G. To indicate that the random variables are defined with respect to G, let $A_Y(G)$ and $A_D(G)$ be the value of A_Y and A_D observed under G, respectively. We assume that $A_Y(G)$ and $A_D(G)$ are randomized independently of each other to values in $\{0,1\}$, that is $A_Y(G) \perp\!\!\!\perp A_D(G)$. Assume no losses to follow-up. Define the independencies

$$Y_{k+1}(G) \perp\!\!\!\perp A_D(G) \mid A_Y(G), Y_k(G) = 0, D_{k+1}(G) = 0, L(G), \tag{1}$$

$$D_{k+1}(G) \perp\!\!\!\perp A_Y(G) \mid A_D(G), D_k(G) = 0, Y_k(G) = 0, L(G). \tag{2}$$

B.1. **Conditions that ensure $\Delta 1$ and $\Delta 2$.** Since $A_Y(G)$ and $A_D(G)$ are randomly assinged, conditional exchangeability is satisfied in the trial G, such that

$$\bar{Y}_{K+1}^{a_Y,a_D}(G), \bar{D}_{K+1}^{a_Y,a_D}(G) \perp\!\!\!\perp A_Y(G), A_D(G) \mid L(G),$$

where $a_Y, a_D \in \{0,1\}$. In the special case where $a_Y = a_D$, this conditional exchangeability condition is the same as the conditional exchangeability condition in the main text.

Furthermore, we assume consistency in G, that is, if $A_Y = a_Y$ and $A_D = a_D$ then

$$Y_{k+1}^{a_Y,a_D}(G) = Y_{k+1}(G)$$
$$D_{k+1}^{a_Y,a_D}(G) = D_{k+1}(G),$$

where $a_Y, a_D \in \{0,1\}$. This consistency condition is identical to the consistency condition in the main text when $a_Y = a_D$.

We assume positivity in G, that is, for all $l \in \mathcal{L}$,

$$\Pr(L(G) = l) > 0 \implies$$
$$\Pr(A_Y(G) = a_Y, A_D(G) = a_D \mid L(G) = l) > 0, \text{ for } a_Y, a_D \in \{0,1\}, \tag{3}$$

which holds by design in G.

Let $a_Y = 0$, $a_D = 1$ (an analogous argument holds when $a_Y = 1$, $a_D = 0$). Using exchangeability and consistency we find that, for all $l \in \mathcal{L}$,

$$\Pr(Y_{k+1}(G) = 1 \mid Y_k(G) = 0, D_{k+1}(G) = 0, A_Y(G) = 0, A_D(G) = 1, L(G) = l)$$
$$= \Pr(Y_{k+1}^{a_Y=0,a_D=1}(G) = 1 \mid Y_k^{a_Y=0,a_D=1}(G) = 0, D_{k+1}^{a_Y=0,a_D=1}(G) = 0, A_Y(G) = 0, A_D = 1, L(G) = l)$$
$$\text{consistency, pos.}$$
$$= \Pr(Y_{k+1}^{a_Y=0,a_D=1}(G) = 1 \mid Y_k^{a_Y=0,a_D=1}(G) = 0, D_{k+1}^{a_Y=0,a_D=1}(G) = 0, L(G) = l) \qquad \text{exchangeability}$$

(4)

Similarly, using (1), exchangeability and consistency we find

$$\Pr(Y_{k+1}(G) = 1 \mid Y_k(G) = 0, D_{k+1}(G) = 0, A_Y(G) = 0, A_D(G) = 1, L(G) = l)$$

$$= \Pr(Y_{k+1}(G) = 1 \mid Y_k(G) = 0, D_{k+1}(G) = 0, A_Y(G) = 0, L(G) = l) \qquad \text{due to (1)}$$

$$= \Pr(Y_{k+1}(G) = 1 \mid Y_k(G) = 0, D_{k+1}(G) = 0, A_Y(G) = 0, A_D(G) = 0, L(G) = l) \qquad \text{due to (1)}$$

$$= \Pr(Y_{k+1}^{a_Y=0,a_D=0}(G) = 1 \mid Y_k^{a_Y=0,a_D=0}(G) = 0, D_{k+1}^{a_Y=0,a_D=0}(G) = 0, A_Y(G) = A_D(G) = 0, L(G) = l)$$

consistency, pos.

$$= \Pr(Y_{k+1}^{a_Y=0,a_D=0}(G) = 1 \mid Y_k^{a_Y=0,a_D=0}(G) = 0, D_{k+1}^{a_Y=0,a_D=0}(G) = 0, L(G) = l) \qquad \text{exchangeability}$$

(5)

The derivations in (4) and (5) show that $\Delta 1$ is satisfied if condition (1) holds, assuming conditional exchangeability, positivity and consistency. We can use exactly the same argument to show that condition $\Delta 2$ holds under conditional exchangeability, positivity, consistency and condition (2). Conditions (1) and (2) are helpful in practice because these independences can be evaluated in causal graphs. In particular, these conditions hold in Figure 2, where we have described a trial in which A_Y and A_D are randomly assigned such that $\Pr(A_Y = a_Y, A_D = a_D) > 0$ for all $a_D, a_Y \in \{0, 1\}$.

Note that conditions (3) and (4) in the main text, which are part of the decomposition assumption, are required for the independencies (1) and (2) to hold.

Appendix C. Proof of identifiability

We assume a Finest Fully Randomized Causally Interpretable Structured Tree Graph (FFRCISTG) model [1]. The aim is to identify $P\left(Y_k^{a_Y,a_D,\bar{c}=0} = 1\right)$ as a function of the factual data, in which A is randomized. To do this, we will initially

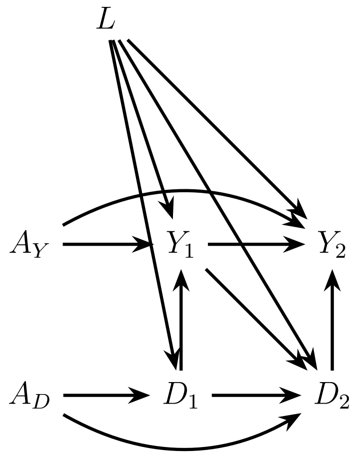

FIGURE 2. Directed acyclic graph describing a trial (G) in which A_Y and A_D are randomized. Here, $\Delta 1$ and $\Delta 2$ hold.

consider a scenario in which both A_Y and A_D are randomized, that is, we consider a 4 arm trial G, as described in Appendix B. Hereafter we omit the string '(G)' after the random variables, e.g. $A_Y(G) = A_Y$, to avoid clutter. We will provide a proof for the scenario with a measured pretreatment covariate L and censoring C_k. The results will immediately hold in simpler scenarios, e.g. by defining L or C_k to be empty.

C.1. **Identifiabilty conditions in the presence of censoring.** First, we generalize the identifiability conditions to allow for censoring. Assume that subjects may be lost to follow-up, and that the losses to follow-up can depend on A_Y, A_D and L, as suggested in Figure 6. Further, assume that the losses to follow-up are independent of future counterfactual events ('independent censoring'). To be more precise, we consider a setting in which we intervened such that no subject was lost to follow-up. Let $C_k \in \{0, 1\}$ be an indicator of loss to follow-up by k. Let $D_k^{a_Y, a_D, \bar{c}=0}$ and $Y_k^{a_Y, a_D, \bar{c}=0}$ be the counterfactual values of Y_k and D_k when A_Y is set to a^*, A_D is set to a, and follow-up is ensured at all times.

In a continuous time setting, it is usually assumed that two events cannot occur at the same point in time. In our discrete time setting with pretreatment covariates

L and censoring C_k, we define a temporal order

$$(L, A_D, A_Y, C_1, D_1, Y_1, C_2, D_2, Y_2, ..., C_{K+1}, D_{K+1}, Y_{K+1}).$$

For all $k \in \{0, K\}$ we consider the following conditions. First, we extend the exchangeability conditions from Section 5.1,

$$\mathbf{E1}: \bar{Y}_{K+1}^{a,\bar{c}=0}, \bar{D}_{K+1}^{a,\bar{c}=0} \perp\!\!\!\perp A \mid L$$

$$\mathbf{E2}: \underline{Y}_{k+1}^{a,\bar{c}=0}, \underline{D}_{k+1}^{a,\bar{c}=0} \perp\!\!\!\perp C_{k+1} \mid Y_k = D_k = \bar{C}_k = 0, L, A.$$

Here, as in Section 5.1, E1 holds when $A \equiv A_Y \equiv A_D$ are randomized. E2 requires that losses to follow-up are independent of future counterfactual events, given the measured past. This condition is similar to the 'independent censoring' condition that is assumed to hold in classical randomized trials [2].

Furthermore, we require a consistency condition such that if $A_Y = a_Y$, $A_D = a_D$ and $\bar{C}_k = 0$, then $Y_k = Y_k^{a_Y,a_D,\bar{c}=0}$ and $D_k = D_k^{a_Y,a_D,\bar{c}=0}$, and still we only observe scenarios where $a_Y = a_D$. The consistency condition ensures that if an individual has a data history consistent with the intervention under a counterfactual scenario, then the observed outcome is equal to the counterfactual outcome.

Similar to Section 5.1, the exchangeability and consistency conditions are conventional in the causal inference literature. We also require an extra positivity condition in the presence of censoring, that is,

$$\Pr(A = a, Y_k = 0, D_k = 0, \bar{C}_k = 0, L = l) > 0 \implies$$

$$\Pr(C_{k+1} = 0 \mid Y_k = 0, D_k = 0, \bar{C}_k = 0, L = l, A = a) > 0,$$

for $a = \{0, 1\}$, which ensures that for any possible history of treatment assignments and covariates among those who are event-free and uncensored at k, some subjects will remain uncensored at $k + 1$.

Finally, we rely on two dismissible component conditions which generalize the conditions in Section 5, by allowing for a hypothetical intervention to eliminate censoring at all times.

Dismissible component conditions: For all $l \in \mathcal{L}$,

$$\mathbf{\Delta 1_c} : \Pr(Y_{k+1}^{a_Y, a_D=1, \bar{c}=0} = 1 \mid Y_t^{a_Y, a_D=1, \bar{c}=0} = 0, D_{k+1}^{a_Y, a_D=1, \bar{c}=0} = 0, L = l)$$
$$= \Pr(Y_{k+1}^{a_Y, a_D=0, \bar{c}=0} = 1 \mid Y_t^{a_Y, a_D=0, c=0} = 0, D_{k+1}^{a_Y, a_D=0, c=0} = 0, L = l)$$

$$\mathbf{\Delta 2_c} : \Pr(D_{k+1}^{a_Y=1, a_D, \bar{c}=0} = 1 \mid Y_k^{a_Y=1, a_D, \bar{c}=0} = 0, D_k^{a_Y=1, a_D, \bar{c}=0} = 0, L = l)$$
$$= \Pr(D_{k+1}^{a_Y=0, a_D, \bar{c}=0} = 1 \mid Y_k^{a_Y=0, a_D, c=0} = 0, D_k^{a_Y=0, a_D, c=0} = 0, L = l).$$

Under these conditions, $\Pr(Y_{K+1}^{a_Y, a_D, \bar{c}=0} = 1)$ is identified from (10).

C.2. **Proof of identifiability.** We consider the counterfactual outcomes in a setting where $a_Y = 0$ and $a_D = 1$ (analogous arguments holds for the setting where $a_Y = 1$ and $a_D = 0$), and we use laws of probability as well as $\Delta 1_c$ and $\Delta 2_c$ to express

$$
\begin{aligned}
&\Pr(Y_{K+1}^{a_Y=0,a_D=1,\bar{c}=0}=1)\\
=&\sum_l \Big[\Pr(Y_{K+1}^{a_Y=0,a_D=1,\bar{c}=0}=1 \mid L=l)\Big]\Pr(L=l)\\
=&\sum_l \Big[\sum_{s=0}^{K}\Pr(Y_{s+1}^{a_Y=0,a_D=1,\bar{c}=0}=1 \mid D_{s+1}^{a_Y=0,a_D=1,\bar{c}=0}=Y_{s}^{a_Y=0,a_D=1,\bar{c}=0}=0, L=l)\\
&\prod_{j=0}^{s}\big[\Pr(D_{j+1}^{a_Y=0,a_D=1,\bar{c}=0}=0 \mid D_{j}^{a_Y=0,a_D=1,\bar{c}=0}=Y_{j}^{a_Y=0,a_D=1,\bar{c}=0}=0, L=l)\\
&\times\Pr(Y_{j}^{a_Y=0,a_D=1,\bar{c}=0}=0 \mid D_{j}^{a_Y=0,a_D=1,\bar{c}=0}=Y_{j-1}^{a_Y=0,a_D=1,\bar{c}=0}=0, L=l)\big]\Big]\Pr(L=l)\\
=&\sum_l \Big[\sum_{s=0}^{K}\Pr(Y_{s+1}^{a_Y=0,a_D=0,\bar{c}=0}=1 \mid D_{s+1}^{a_Y=0,a_D=0,\bar{c}=0}=Y_{s}^{a_Y=0,a_D=0,\bar{c}=0}=0, L=l)\\
&\prod_{j=0}^{s}\big[\Pr(D_{j+1}^{a_Y=1,a_D=1\bar{c}=0}=0 \mid D_{j}^{a_Y=1,a_D=1\bar{c}=0}=Y_{j}^{a_Y=1,a_D=1\bar{c}=0}=0, L=l)\\
&\times\Pr(Y_{j}^{a_Y=0,a_D=0,\bar{c}=0}=0 \mid D_{j}^{a_Y=0,a_D=0,\bar{c}=0}=Y_{j-1}^{a_Y=0,a_D=0,\bar{c}=0}=0, L=l)\big]\Big]\Pr(L=l),\\
=&\sum_l \Big[\sum_{s=0}^{K}\Pr(Y_{s+1}^{a=0,\bar{c}=0}=1 \mid D_{s+1}^{a=0,\bar{c}=0}=Y_{s}^{a=0,\bar{c}=0}=0, L=l)\\
&\prod_{j=0}^{s}\big[\Pr(D_{j+1}^{a=1\bar{c}=0}=0 \mid D_{j}^{a=1\bar{c}=0}=Y_{j}^{a=1\bar{c}=0}=0, L=l)\\
&\times\Pr(Y_{j}^{a=0,\bar{c}=0}=0 \mid D_{j}^{a=0,\bar{c}=0}=Y_{j-1}^{a=0,\bar{c}=0}=0, L=l)\big]\Big]\Pr(L=l),
\end{aligned}
$$

(6)

where $Y_{-1}^{a_Y,a_D,\bar{c}=0}$ and $Y_{-1}^{a_Y,\bar{c}=0}$ are empty.

For $s \geq 0$ and all l such that $\Pr(D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = 0, L = l) > 0$, let us consider the term

$$\begin{aligned}
&\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = 0, L = l) \\
=&\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_0 = 0, L = l) \\
=&\frac{\Pr(Y_{s+1}^{a,\bar{c}=0} = 1, \bar{D}_{s+1}^{a,\bar{c}=0} = \bar{Y}_s^{a,\bar{c}=0} = 0 \mid Y_0 = D_0 = \bar{C}_0 = 0, A = a, L = l)}{P(\bar{D}_{s+1}^{a,\bar{c}=0} = \bar{Y}_s^{a,\bar{c}=0} = 0 \mid Y_0 = D_0 = \bar{C}_0 = 0, A = a, L = l)},
\end{aligned}$$

where we use the fact that all subjects are event-free and uncensored at $k = 0$ in the 2nd line, and laws of probability and E1 in the 3rd line. Then, we use positivity and E2 to find

$$\begin{aligned}
&\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_0 = 0, A = a, L = l) \\
=&\frac{\Pr(Y_{s+1}^{a,\bar{c}=0} = 1, \bar{D}_{s+1}^{a,\bar{c}=0} = \bar{Y}_s^{a,\bar{c}=0} = 0 \mid Y_0 = D_0 = \bar{C}_1 = 0, A = a, L = l)}{P(\bar{D}_{s+1}^{a,\bar{c}=0} = \bar{Y}_s^{a,\bar{c}=0} = 0 \mid Y_0 = D_0 = \bar{C}_1 = 0, A = a, L = l)} \\
=&\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_1 = 0, A = a, L = l).
\end{aligned}$$

(7)

Similarly, if $s = 1$ we use consistency, a new step like (7), and consistency to find that

$$\begin{aligned}
&\Pr(Y_2^{a,\bar{c}=0} = 1 \mid D_2^{a,\bar{c}=0} = Y_1^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_1 = 0, A = a, L = l) \\
=&\Pr(Y_2^{a,\bar{c}=0} = 1 \mid D_2^{a,\bar{c}=0} = Y_1 = D_1 = \bar{C}_1 = 0, A = a, L = l) \\
=&\Pr(Y_2^{a,\bar{c}=0} = 1 \mid D_2^{a,\bar{c}=0} = Y_1 = D_1 = \bar{C}_2 = 0, A = a, L = l) \\
=&\Pr(Y_2 = 1 \mid Y_1 = D_2 = \bar{C}_2 = 0, A = a, L = l).
\end{aligned}$$

If $s > 1$, we use consistency to find

$$\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_1 = 0, A = a, L = l)$$

$$= \Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_1 = D_1 = \bar{C}_1 = 0, A = a, L = l).$$

(8)

Then, we repeat the steps in (7) and (8) to find that for all $s \in (1, 2, ..., K+1)$,

$$\Pr(Y_{s+1}^{a,\bar{c}=0} = 1 \mid D_{s+1}^{a\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_0 = D_0 = \bar{C}_0 = 0, A = a, L = l)$$

$$= \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = \bar{C}_{s+1} = 0, A = a, L = l).$$

(9)

Similarly, for $D_{s+1}^{a,\bar{c}=0}$ we could follow the same steps as for $Y_{s+1}^{a,\bar{c}=0}$ to express

$$\Pr(D_{s+1}^{a,\bar{c}=0} = 1 \mid D_s^{a,\bar{c}=0} = Y_s^{a,\bar{c}=0} = Y_k = D_k = \bar{C}_k = 0, A = a, L = l)$$

$$= \Pr(D_{s+1} = 1 \mid D_s = Y_s = \bar{C}_{s+1} = 0, A = a, L = l).$$

(10)

Using the results in (6), (9) and (10), we find that

$$\Pr(Y_{K+1}^{a_Y,a_D,\bar{c}=0} = 1)$$

$$= \sum_l \Big[\sum_{s=0}^{K} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = \bar{C}_{s+1} = 0, A = a_Y, L = l)$$

$$\prod_{j=0}^{s} \big[\Pr(D_{j+1} = 0 \mid D_j = Y_j = \bar{C}_{j+1} = 0, A = a_D, L = l)$$

$$\times \Pr(Y_j = 0 \mid D_j = Y_{j-1} = \bar{C}_j = 0, A = a_Y, L = l)\big]\Big] \Pr(L = l).$$

In words, we have derived that $\Pr(Y_{K+1}^{a_Y,a_D,\bar{c}=0} = 1)$ is identified from a trial in which only subjects with $(A_Y = A_D = A)$ are observed, i.e. in a trial in which A is randomized. Hence, in practice we only need data from the treatment arms in which $A \equiv A_Y \equiv A_D \in \{0,1\}$.

Appendix D. Proof of weighted representations

For the ease of exposition, define

$$W'_{C,k}(a_Y) = \frac{1}{\prod_{j=0}^{k} \Pr(C_{j+1} = 0 \mid \bar{C}_j = D_j = Y_j = 0, L = l, A = a_D)}.$$

Consider the expression

$$\begin{aligned}
&\mathbb{E}[W_{C,k}(a_Y)W_{D,k}(a_Y, a_D)Y_{k+1}(1 - Y_k)(1 - D_{k+1}) \mid A = a_Y] \\
=&\mathbb{E}[W'_{C,k}(a_Y)W_{D,k}(a_Y, a_D)Y_{k+1}(1 - Y_k)(1 - D_{k+1})(1 - \bar{C}_{k+1}) \mid A = a_Y] \\
=&\sum_{l} \sum_{\bar{y}_{k+1}} \sum_{\bar{d}_{k+1}} [\Pr(\bar{y}_{k+1}, d_{k+1}, c_{k+1}, l \mid A = a_Y)W'_{C,k}(a)W_{D,k}(a_Y, a_D) \\
&\times y_{k+1}(1 - y_k)(1 - d_{k+1})(1 - c_{k+1})] \\
=&\sum_{l} [\Pr(Y_{k+1} = 1, Y_k = D_{k+1} = \bar{C}_{k+1} = 0, l \mid A = a_Y)W'_{C,k}(a_Y)W_{D,k}(a_Y, a_D)] \\
=&\sum_{l} [\Pr(Y_{k+1} = 1 \mid Y_k = D_{k+1} = \bar{C}_{k+1} = 0, L = l, A = a_Y) \\
&\times \Pr(D_{k+1} = 0 \mid \bar{C}_{k+1} = \bar{D}_k = \bar{Y}_k = 0, L = l, A = a_Y) \\
&\times \Pr(C_{k+1} = 0 \mid \bar{D}_k = \bar{Y}_k = \bar{C}_k = 0, L = l, A = a_Y) \\
&\times \Pr(\bar{Y}_k = \bar{D}_k = \bar{C}_k = 0, L = l \mid a_Y) \\
&\times W'_{C,k}(a_Y)W_{D,k}(a_Y, a_D)]
\end{aligned}$$

where we use the definition of expected value, the fact that Y_k and D_k are binary, and laws of probability.

We use laws of probability to express $\Pr(\bar{Y}_k = \bar{D}_k = \bar{C}_k = 0, L = l \mid A = a_Y)$ as

$$\Pr(Y_k = 0 \mid \bar{C}_k = D_k = Y_{k-1} = 0, L = l, A = a_Y)$$

$$\times \Pr(D_k = 0 \mid \bar{C}_k = D_{k-1} = Y_{k-1} = 0, L = l, A = a_Y)$$

$$\times \Pr(C_k = 0 \mid D_{k-1} = Y_{k-1} = \bar{C}_{k-1} = 0, L = l, A = a_Y)$$

$$\times \Pr(\bar{Y}_{k-1} = \bar{D}_{k-1} = 0, \bar{C}_{k-1} = 0, L = l \mid A = a_Y),$$

where any variable indexed with a number $m < 0$ are defined to be empty.

Arguing iteratively for $k-1, k-2, ..., 0$ we find that

$$\mathbb{E}[W'_{C,k}(a_Y) W_{D,k}(a_Y, a_D) Y_{k+1}(1 - Y_k)(1 - D_{k+1})(1 - C_{k+1}) \mid A = a_Y]$$

$$= \sum_l \Big[\Pr(Y_{k+1} = 1 \mid Y_k = D_{k+1} = \bar{C}_{k+1} = 0, L = l, A = a_Y)$$

$$\prod_{j=0}^{k} \big\{ \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L = l, A = a_Y)$$

$$\times \Pr(Y_j = 0 \mid \bar{C}_j = D_j = Y_{j-1} = 0, L = l, A = a_Y)$$

$$\times \Pr(C_{j+1} = 0 \mid \bar{D}_j = \bar{Y}_j = \bar{C}_j = 0, L = l, a_Y) \big\}$$

$$\times \Pr(L = l) W'_{C,k}(a_Y) W_{D,k}(a_Y, a_D) \Big],$$

We plug in the expression for $W'_{C,k}(a_Y)$ to get

$$
\begin{aligned}
&= \sum_{\bar{l}} [\Pr(Y_{k+1} = 1 \mid Y_k = D_{k+1} = \bar{C}_{k+1} = 0, L = l, A = a_Y) \\
&\times \prod_{j=0}^{k} \Big\{ \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L = l, A = a_Y) \\
&\quad \times \Pr(Y_j = 0 \mid \bar{C}_j = D_j = Y_{j-1} = 0, L = l, A = a_Y) \Big\} \\
&\quad \times \Pr(L = l) W_{D,k}(a_Y, a_D)].
\end{aligned}
$$

We plug in the expression for the weights $W_{D,k}(a_Y, a_D)$ to get

$$
\begin{aligned}
&= \sum_{\bar{l}} [\Pr(Y_{k+1} = 1 \mid Y_k = D_{k+1} = \bar{C}_{k+1} = 0, L = l, A = a_Y) \\
&\times \prod_{j=0}^{k} \Big\{ \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L = l, A = a_D) \\
&\quad \times \Pr(Y_j = 0 \mid \bar{C}_j = D_j = Y_{j-1} = 0, L = l, A = a_Y) \Big\}, \\
&\times \Pr(L = l)],
\end{aligned}
$$

and the final expression is equal to (10).

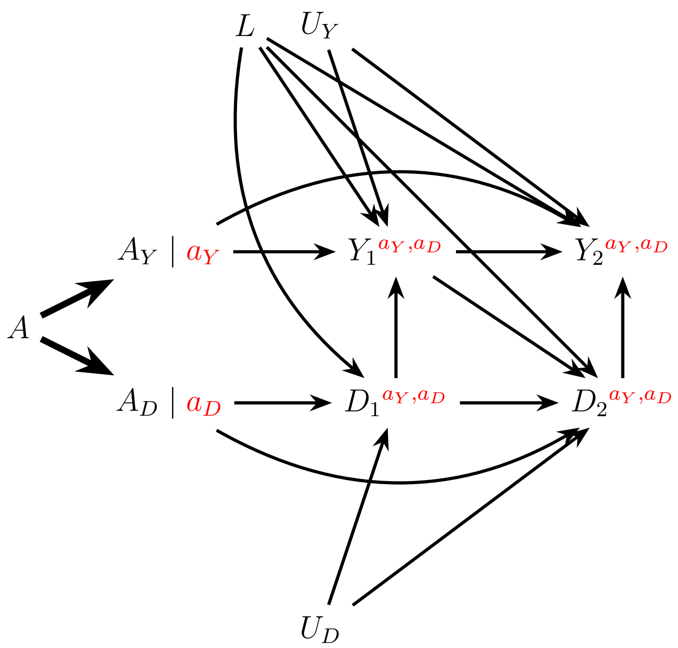

FIGURE 3. Single world intervention template (SWIT) that describes a scenario with interventions on A_Y, A_D and $\bar{C}_k$. Even if U_Y and U_D are unmeasured, $\Delta 1$ and $\Delta 2$ hold.

APPENDIX E. EXPLORING THE DISMISSIBLE COMPONENT CONDITIONS

By considering causal graphs, we provide some insight into the interpretation of assumptions $\Delta 1$ and $\Delta 2$.

E.1. **Scenario in which the dismissible component conditions are satisfied.** Consider the study from Appendix B in which A_Y and A_D were randomized without loss to follow-up, which ensures positivity and exchangeability. Furthermore, we assume that the usual assumptions about consistency is satisfied; if $A_Y = a_Y$,$A_D = a_D$, then $Y_k = Y_k^{a_Y,a_D}$.

Assume that the causal structure in the single world intervention template (SWIT) of Figure 5 holds. Here, A_Y is d-separated from both $Y_k^{a_Y,a_D}$ and $D_k^{a_Y,a_D}$ for $k \in 1,2$. Similarly A_D is d-separated from both $Y_k^{a_Y,a_D}$ and $D_k^{a_Y,a_D}$. Hence, under the assumptions about positivity and consistency, we can identify the following joint law from the g-formula,

$$
\begin{aligned}
&\Pr(Y_2^{a_Y,a_D}=1, Y_1^{a_Y,a_D}=0, D_2^{a_Y,a_D}=0, D_1^{a_Y,a_D}=0 \mid L)\\
=&\Pr(D_1=0 \mid A_Y=a_Y, A_D=a_D, L)\Pr(Y_1=0 \mid D_1=0, A_Y=a_Y, A_D=a_D, L)\\
\times&\Pr(D_2=0 \mid D_1=0, Y_1=0, A_Y=a_Y, A_D=a_D, L)\\
\times&\Pr(Y_2=1 \mid D_2=0, D_1=0, Y_1=0, A_Y=a_Y, A_D=a_D, L)\\
=&\Pr(D_1=0 \mid A_D=a, L)\Pr(Y_1=0 \mid D_1=0, A_Y=a_Y, L)\\
\times&\Pr(D_2=0 \mid D_1=0, Y_1=0, A_D=a_D, L)\Pr(Y_2=1 \mid D_2=0, D_1=0, Y_1=0, A_Y=a_Y, L),
\end{aligned}
$$

(11)

where the last equality follows due to conditional independences that we read off the causal graph. Similarly, we can identify

$$
\begin{aligned}
&\Pr(Y_1^{a_Y,a_D}=0, D_2^{a_Y,a_D}=0, D_1^{a_Y,a_D}=0 \mid L)\\
=&\Pr(D_1=0 \mid A_D=a_D, L)\Pr(Y_1=0 \mid D_1=0, A_Y=a_Y, L)\\
\times&\Pr(D_2=0 \mid D_1=0, Y_1=0, A_D=a_D, L).
\end{aligned}
$$

(12)

Using laws of total probability,

$$
\begin{aligned}
&\Pr(Y_2^{a_Y,a_D}=1 \mid Y_1^{a_Y,a_D}=0, D_2^{a_Y,a_D}=0, D_1^{a_Y,a_D}=0, L)\\
=&\frac{\Pr(Y_2^{a_Y,a_D}=1, Y_1^{a_Y,a_D}=0, D_2^{a_Y,a_D}=0, D_1^{a_Y,a_D}=0 \mid L)}{\Pr(Y_1^{a_Y,a_D}=0, D_2^{a_Y,a_D}=0, D_1^{a_Y,a_D}=0 \mid L)}\\
=&\Pr(Y_2=1 \mid D_2=0, D_1=0, Y_1=0, A_Y=a_Y, L).
\end{aligned}
$$

(13)

Hence,

$$
\begin{aligned}
&= \Pr(Y_2^{a_Y, a_D=1} = 1 \mid Y_1^{a_Y, a_D=1} = 0, D_2^{a_Y, a_D=1} = 0, D_1^{a_Y, a_D=1} = 0, L) \\
&= \Pr(Y_2^{a_Y, a_D=0} = 1 \mid Y_1^{a_Y, a_D=0} = 0, D_2^{a_Y, a_D=0} = 0, D_1^{a_Y, a_D=0} = 0, L),
\end{aligned}
$$

that is $\Delta 1$ is satisfied at $k = 2$. Using the same argument, we can derive that $\Delta 2$ is satisfied for $k = 2$, and both $\Delta 1$ and $\Delta 2$ will be satisfied for $k = 1$. That is, Figure 5 implies that $\Delta 1$ and $\Delta 2$ hold. Furthermore, we could use exactly the same derivations to find that $\Delta 1$ and $\Delta 2$ hold in Figure 3, even if U_Y and U_D are unmeasured.

E.2. **Scenario in which the dismissible component conditions are not necessarily satisfied.** Consider the SWIT in Figure 4, which only differs from Figure 5 in the variable U_Y that is an unmeasured common cause of Y_1 and D_1. Here we read off Figure 4 to find that

$$\Pr(Y_1^{a_Y,a_D} = 1 \mid D_1^{a_Y,a_D} = 0, L)$$
$$= \Pr(Y_1 = 1 \mid D_1 = 0, A_Y = a_Y, A_D = a_D, L),$$

(14)

However, we cannot conclude from the graph that

$$\Pr(Y_1 = 1 \mid D_1 = 0, A_Y = a_Y, A_D = 1, L)$$
$$= \Pr(Y_1 = 1 \mid D_1 = 0, A_Y = a_Y, A = 0, L)$$

(15)

because there is an open collider path $a_D \rightarrow D_1 \leftarrow U_{YD} \rightarrow Y_1$. Hence, we cannot conclude that the graph in Figure 4 implies $\Delta 1$, and our results do not allow us to identify $\Pr(Y_1^{a_Y,a_D} = 1)$ in this scenario. The unmeasured common cause U_{YD} of Y_k and $D_{k'}$ for $k, k' \in (0, 1, ..., K + 1)$ leads to violation of $\Delta 1$ and $\Delta 2$.

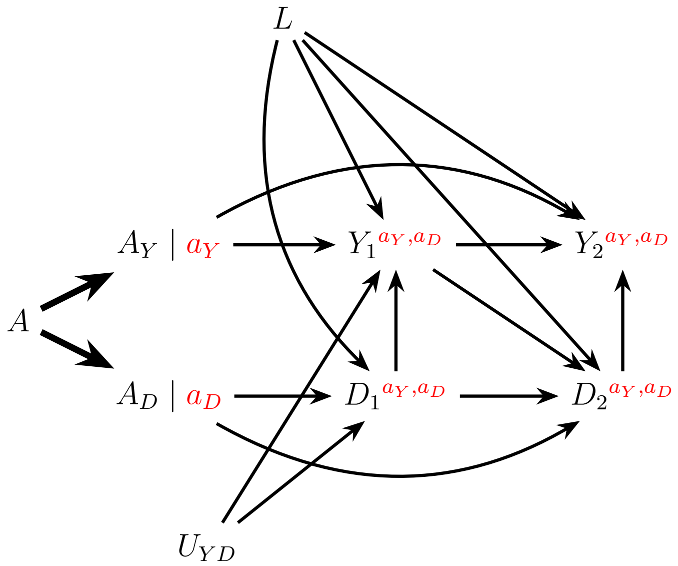

FIGURE 4. Single world intervention template (SWIT) of a scenario in which $\Delta 1$ and $\Delta 2$ are not implied by the graph.

Appendix F. Simulations

Here we present simulations from 5 scenarios to illustrate the finite sample performance of the separable effects. We consider settings where the dismissible component conditions are satisfied, but also settings where these conditions are violated. Furthermore, we consider coverage under violation of the parametric model assumptions.

In each scenario, we simulated two randomized experiments in which 400 and 2000 subjects were randomly assigned to treatment $A \in \{0,1\}$, respectively. To assess finite sample behavior, we calculated confidence intervals for 3 time points by simulating each experiment 500 times, and for each of these experiments we created non-parametric percentile bootstrap confidence intervals from 500 bootstrap samples.

The true cumulative incidences from the simulation scenarios are shown in Figure 5. Generally, our simulations confirm that the g-formula and IPW estimators perform satisfactory when the identifiability conditions are satisfied.

F.1. **Data generating mechanism.** For each individual, the data were generated from the following algorithm, where we have omitted i subscripts to indicate inidivuals:

(1) Draw $L_1 \sim \text{Bernoulli}[p = 0.25]$.

(2) Draw $L_2 \sim \text{Bernoulli}[p = 0.2L_1 + 0.8(1 - L_1)]$.

(3) Draw $A \sim \text{Bernoulli}[p = 0.5]$, and define $A_Y \equiv A_D \equiv A$.

(4) Set $D_0 = Y_0 = 0$.

(5) For each $k \in \{0, K\}$,

- if $D_k = Y_k = 0$,

  draw $D_{k+1} \sim \text{Bernoulli}[p = \alpha_D \psi_k(A_Y, A_D, L_1, L_2)]$, where

$$\begin{aligned}\psi_k(A_Y, A_D, L_1, L_2) = \text{expit}(&\omega_0 + \omega_{1,k}k + \omega_2 A_Y + \omega_3 A_D + \omega_4 L_1 + \omega_5 L_2 \\ &+ \omega_6 A_Y L_1 + \omega_7 A_D L_1)\end{aligned} \tag{16}$$

  if $D_{k+1} = 0$,

  draw $Y_{k+1} \sim \text{Bernoulli}(p = \alpha_Y \lambda_k(A_Y, A_D, L_1, L_2))$, where

$$\begin{aligned}\lambda_k(A_Y, A_D, L_1, L_2) = \text{expit}(&\xi_0 + \xi_{1,k}k + \xi_2 A_Y + \xi_3 A_D + \xi_4 L_1 + \xi_5 L_2 \\ &+ \xi_6 A_Y L_1 + \xi_7 A_D L_1).\end{aligned} \tag{17}$$

  if $D_{k+1} = 1$, set $Y_{k+1} = 0$.

- else, define $D_{k+1} = D_k$,$Y_{k+1} = Y_k$.

The coefficients in each of the scenarios are found in Table 3 and the true cumulative incidence curves of $Y_{k+1}, k \in \{0, 99\}$ is found in Figure 5.

Scenario	α_Y	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	α_D	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
1	0.01	0	10	0	5	0	0	0	0.03	0	0	-2	5	0	0	0
2	0.01	0	10	0	-2	5	0	0	0.03	0	0	-2	5	-2	0	0
3	0.01	0	10	0	5	-10	5	0	0.03	0	0	-2	5	-10	0	0
4	0.01	0	10	5	5	0	0	0	0.03	0	0	-2	5	0	0	0
5	0.01	0	10	0	-10	0	0	0	0.03	0	0	-2	0	0	0	0

TABLE 3. Data generating mechanism for the 5 simulation scenarios.

F.2. **Scenario 1: Dismissible component conditions hold and no model mis-specification.** Data were generated from the simple setting described by the first row in Table 3; that is, there is a causal effect of (i) A_Y on Y_k, (ii) A_D on D_k, and (iii) L_1 on both Y_k and D_k. Here, both the dismissible component conditions hold conditional on L_1.

To estimate the separable effects, we fitted the following models

$$\text{logit}[\widehat{\Pr}(Y_k = 1 \mid D_k = Y_{k-1} = 0, A, L_1, L_2)] = \theta_{0,k} + \theta_1 A + \theta_2 L_1 + \theta_3 L_2 \tag{18}$$

$$\text{logit}[\widehat{\Pr}(D_k = 1 \mid D_{k-1} = Y_{k-1} = 0, A, L_1, L_2)] = \beta_{0,k} + \beta_1 A + \beta_2 L_1, \tag{19}$$

which are correctly specified, even if model (18) includes a term θ_3 that is redundant. Thus, we would expect all our estimators to have nominal coverage, and this is confirmed in Table 4; here, coverage is derived from estimated 95% confidence intervals based on the parametric g-formula estimator (g-formula) and the weighted estimators ($\hat{\nu}_{1,a_Y,a_D,k}$ and $\hat{\nu}_{2,a_Y,a_D,k}$) for the trial with $n = 400$ subjects.

		$n = 400$		
Parameter	Estimator	$k = 100$	$k = 75$	$k = 25$
$\Pr(Y_k^{a_Y=1,a_D=1} = 1)$	g-formula	0.95	0.94	0.93
	non-parametric	0.95	0.94	0.95
$\Pr(Y_k^{a_Y=0,a_D=0} = 1)$	g-formula	0.94	0.93	0.92
	non-parametric	0.94	0.95	0.95
$\Pr(Y_k^{a_Y=1,a_D=0} = 1)$	g-formula	0.95	0.96	0.94
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.94	0.95	0.95
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.96	0.95	0.95
$\Pr(Y_k^{a_Y=0,a_D=1} = 1)$	g-formula	0.93	0.93	0.94
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.92	0.90	0.95
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.94	0.94	0.92

TABLE 4. Scenario 1.

Scenario 2: Dismissible component conditions hold and minor model mis-specification. In this scenario, there are causal effects of both L_1 and L_2 on Y_k and D_k (second row in Table 3). Both the dismissible component conditions hold conditional on L_1 and L_2. We used regression models (18) and (19) for model fitting.

Note that in this setting (18) is correctly specified, but (19) is mis-specified because it does not include a term for L_2. Thus, we would expect that the IPW estimator that uses the correctly specified regression model ($\hat{\nu}_{2,a_Y,a_D,k}$) is unbiased, but the parametric g-formula estimator and the other IPW estimator ($\hat{\nu}_{1,a_Y,a_D,k}$) are biased because (19) is mis-specified. The results in Table 5, however, suggest that all estimators have close to nominal coverage. This may be explained by the fact that the model mis-specification is minor, and the magnitude of the separable effects is small (see Figure 5).

		$n = 400$		
Parameter	Estimator	$k = 100$	$k = 75$	$k = 25$
$\Pr(Y_k^{a_Y=1,a_D=1} = 1)$	g-formula	0.91	0.92	0.91
	non-parametric	0.95	0.96	0.93
$\Pr(Y_k^{a_Y=0,a_D=0} = 1)$	g-formula	0.94	0.94	0.93
	non-parametric	0.93	0.93	0.93
$\Pr(Y_k^{a_Y=1,a_D=0} = 1)$	g-formula	0.96	0.94	0.91
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.93	0.95	0.93
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.91	0.92	0.88
$\Pr(Y_k^{a_Y=0,a_D=1} = 1)$	g-formula	0.94	0.93	0.93
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.90	0.91	0.93
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.93	0.94	0.94

TABLE 5. Scenario 2.

Scenario 3: Dismissible component conditions hold and model mis-specification. In this scenario, both the dismissible component conditions hold conditional on L_1 and L_2. Unlike Scenarios 1 and 2, we fitted the following regression models to the simulated data,

$$\text{(20)} \quad \text{logit}[\widehat{\Pr}(Y_k = 1 \mid D_k = Y_{k-1} = 0, A, L_1, L_2)] = \theta_{0,k} + \theta_1 A + \theta_2 L_1$$

$$\text{(21)} \quad \text{logit}[\widehat{\Pr}(D_k = 1 \mid D_{k-1} = Y_{k-1} = 0, A, L_1, L_2)] = \beta_{0,k} + \beta_1 A + \beta_2 L_1 + \beta_3 L_2.$$

Here, (20) is mis-specified because it does not include a term for L_2, but (21) is correctly specified; thus the correctness of the model specifications are opposite from Scenario 2. Also, L_2 exerts larger effects on Y_k and D_k in this setting compared to Scenario 2.

The results in Table 6 illustrate that the IPW estimator $\hat{\nu}_{1,a_Y,a_D,k}$ is unbiased because it only relies on a model that is correctly specified, but the parametric g-formula estimator and the other IPW estimator ($\hat{\nu}_{2,a_Y,a_D,k}$) are biased – in particular, for shorter follow-up times – because they rely on mis-specified regression models.

		$n = 400$		
Parameter	Estimator	$k = 100$	$k = 75$	$k = 25$
$\Pr(Y_k^{a_Y=1,a_D=1} = 1)$	g-formula	0.93	0.95	0.91
	non-parametric	0.93	0.93	0.94
$\Pr(Y_k^{a_Y=0,a_D=0} = 1)$	g-formula	0.93	0.86	0.48
	non-parametric	0.94	0.93	0.94
$\Pr(Y_k^{a_Y=1,a_D=0} = 1)$	g-formula	0.93	0.94	0.93
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.94	0.94	0.93
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.91	0.72	0.56
$\Pr(Y_k^{a_Y=0,a_D=1} = 1)$	g-formula	0.82	0.74	0.45
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.95	0.95	0.94
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.84	0.72	0.33

TABLE 6. Scenario 3.

Scenario 4: Dismissible component conditions fail and model misspecification. The dismissible component condition $\Delta 2$ fails in this scenario due to the non-zero coefficient $\omega_3 = 5$; there is a direct effect $A_Y \to D_k$ for $k \in \{0, 100\}$. Yet we fitted regression models (18) and (19) to the simulated data.

The simulations suggest that none of the estimators has nominal coverage for $\Pr(Y_{k+1}^{a_Y=0,a_D=1} = 1)$. However, since dismissible component condition $\Delta 1$ holds we can identify $\Pr(Y_{k+1}^{a_Y=1,a_D=0} = 1)$, as suggested by the nominal coverage for this quantity in Table 7. Yet we cannot interpret a contrast $\Pr(Y_{k+1}^{a_Y=0,a_D=1} = 1)$ vs $\Pr(Y_{k+1}^{a_Y=1,a_D=1} = 1)$ as the *separable direct effect* of A, due to the violation of the dismissible component condition.

		$n = 400$		
Parameter	Estimator	$k = 100$	$k = 75$	$k = 25$
$\Pr(Y_k^{a_Y=1,a_D=1} = 1)$	g-formula	0.96	0.94	0.93
	non-parametric	0.95	0.94	0.94
$\Pr(Y_k^{a_Y=0,a_D=0} = 1)$	g-formula	0.93	0.93	0.92
	non-parametric	0.93	0.93	0.95
$\Pr(Y_k^{a_Y=1,a_D=0} = 1)$	g-formula	0.96	0.97	0.94
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.94	0.96	0.94
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.97	0.96	0.96
$\Pr(Y_k^{a_Y=0,a_D=1} = 1)$	g-formula	0.05	0.05	0.07
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.31	0.26	0.34
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.05	0.04	0.12

TABLE 7. Scenario 4.

Scenario 5: Dismissible component conditions hold and no model misspecification. In this scenario, L_1 exerts (strong) causal effects on Y_k but not on D_k. Thus, all the dismissible component conditions hold marginally. To illustrate that we obtain unbiased estimates even if L_1 is not included in any of the regression models, we fitted the parsimonious models,

$$\text{logit}[\widehat{\Pr}(Y_k = 1 \mid D_k = Y_{k-1} = 0, A)] = \theta_{0,k} + \theta_1 A. \tag{22}$$

$$\text{logit}[\widehat{\Pr}(D_k = 1 \mid D_{k-1} = Y_{k-1} = 0, A)] = \beta_{0,k} + \beta_1 A, \tag{23}$$

and the results in Table 8 show that all estimators have nominal coverage, even if L_1 is not included in the models.

Parameter	Estimator	$n = 400$ $k = 100$	$k = 75$	$k = 25$
$\Pr(Y_k^{a_Y=1,a_D=1} = 1)$	g-formula	0.95	0.94	0.94
	non-parametric	0.95	0.95	0.95
$\Pr(Y_k^{a_Y=0,a_D=0} = 1)$	g-formula	0.94	0.94	0.93
	non-parametric	0.95	0.94	0.94
$\Pr(Y_k^{a_Y=1,a_D=0} = 1)$	g-formula	0.96	0.95	0.94
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.97	0.96	0.95
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.95	0.95	0.94
$\Pr(Y_k^{a_Y=0,a_D=1} = 1)$	g-formula	0.93	0.94	0.94
	$\hat{\nu}_{1,a_Y,a_D,k}$	0.94	0.93	0.94
	$\hat{\nu}_{2,a_Y,a_D,k}$	0.94	0.94	0.94

TABLE 8. Scenario 5.

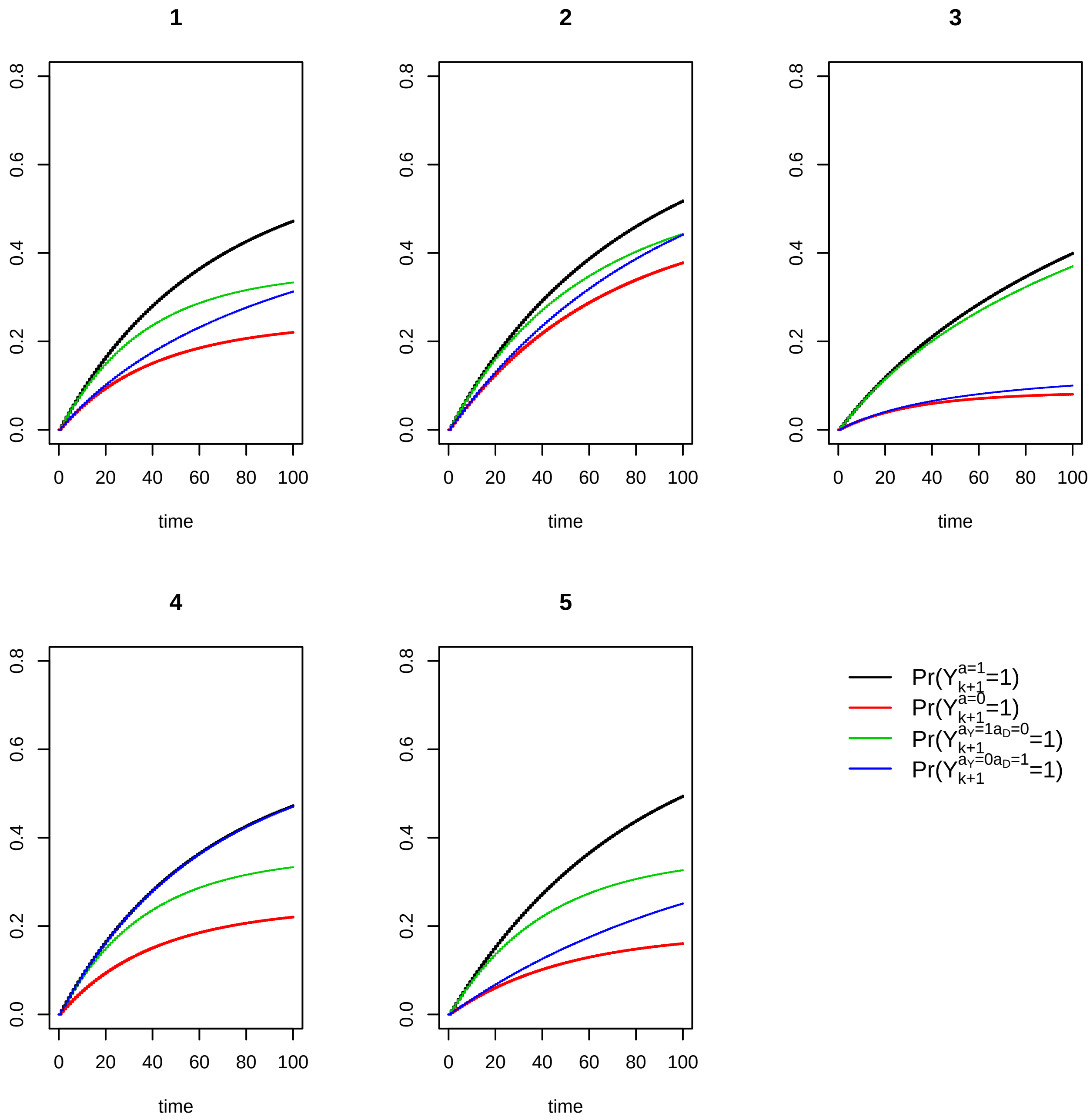

FIGURE 5. True cumulative incidence curves for scenarios 1-5.

References

[1] James M Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.

[2] Jessica G. Young, Mats J. Stensrud, Eric J. Tchetgen Tchetgen, and Miguel A. Hernán. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, 2020.