Separable Pathway Effects of Semi-Competing Risks via Multi-State Models

Yuhao Deng
Joint work with Yi Wang, Xiang Zhan and Xiao-Hua Zhou

Peking University; University of Michigan

November 2, 2023

Outline

Outline

Motivating Data

- Allogeneic stem cell transplantation is a well applied therapy to treat acute lymphoblastic leukemia, including two sorts of transplant modalities: human leukocyte antigens matched sibling donor transplantation (MSDT) and haploidentical stem cell transplantation from family (Haplo-SCT).
- MSDT has long been regarded as the first choice of transplantation because MSDT leads to lower transplantation related mortality, also known as non-relapse mortality (NRM).
- In recent years, some benefits of Haplo-SCT have been noticed that patients with positive pre-transplantation minimum residual disease (MRD) undergoing Haplo-SCT have better prognosis in relapse.
- However, the mechanism of how transplant modalities can affect overall survival needs further exploration.

Semi-Competing Risks

- In many clinical trials focusing on time-to-event outcomes, the primary (terminal) event may be affected by occurrences of intermediate (non-terminal) events.
- The terminal event can truncate the non-terminal event, but not vice versa (Fine et al., 2001).
- Some individuals have observations on both intermediate and primary events, while others only have observations on primary events.
- For example, to study the effectiveness of stem cell transplantation, mortality is a primary event and relapse of leukemia is an intermediate event.

Cumulative Incidences

- Let T be the time to the primary event, and R be the time to the intermediate event if any.
- The cumulative incidence of the primary event $F(t) := P(T \leq t)$ is usually adopted as the estimand in survival analysis.
- The hazard of developing primary events at time $t > 0$, denoted by $\lambda(t) := -d \log\{P(T \geq t)\}/dt$, may be different between those with intermediate events and those without intermediate events.
- Therefore, the hypothetical cumulative incidence of the primary event by appropriately adjusting the intermediate event is of more interest.

Methods for Semi-Competing Risks Data

- Identification of the dependence between T and R is a fundamental problem.
- Joint survivor function:

$$P(T \geq t, R \geq r),$$

  defined on a wedge $\{(t, r) : t \geq r \geq 0\}$
- Using copulas to model the joint survivor function (Clayton, 1978).
- Modelling the data generating process under illness-death models (Xu et al. 2010).

Multi-State (Illness-Death) Models

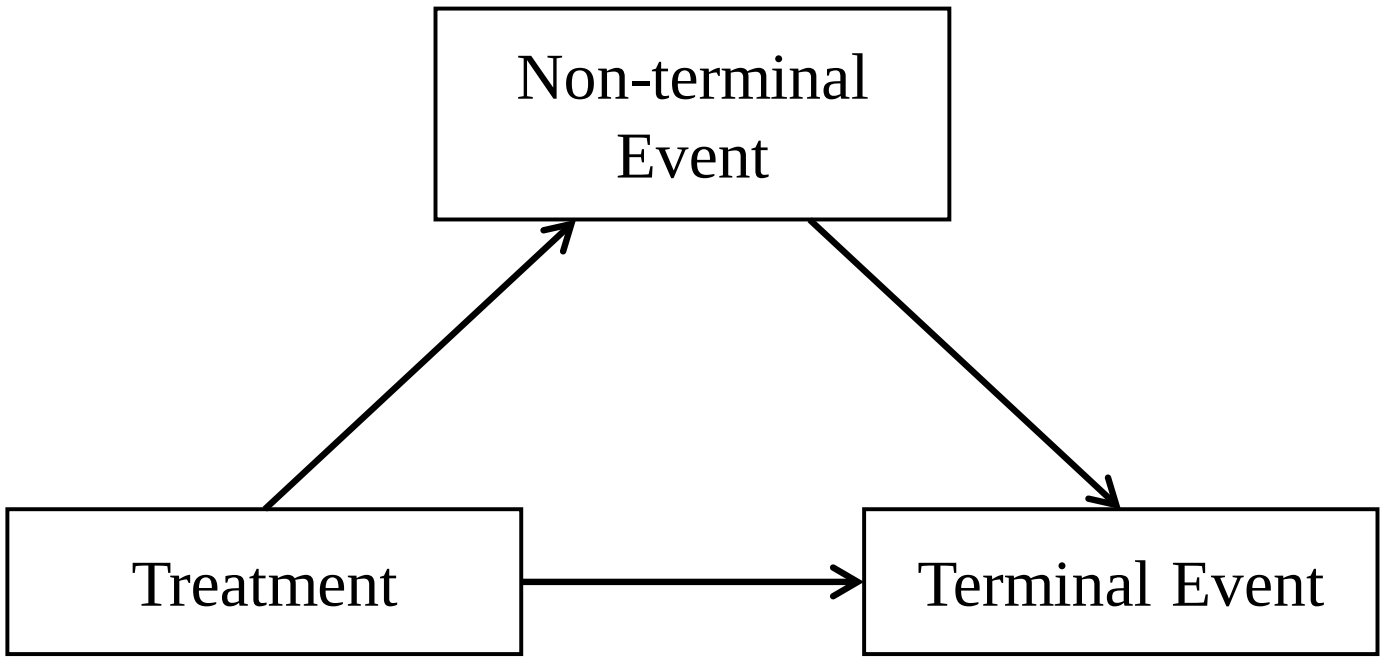

Figure: Multi-state (illness-death) models.

Causal Inference to Semi-Competing Risks

- To study the treatment effect of an intervention A on the primary event, we shall adopt the potential outcomes framework.
- A: Binary treatment.
- T^a: Potential time to primary event.
- R^a: Potential time to intermediate event.
- C^a: Potential cenosring time.
- X: Baseline covariates.
- Under the stable unit treatment value assumption (SUTVA), the potential outcomes are well defined.
- The treatment effect should be defined by contrasting well-defined distributions of T^1 and T^0, appropriately adjusting R^1 and R^0.

Causal Inference to Semi-Competing Risks

- Fundemantal problem of causal inference: only one of the two sets of potential outcomes is observable.
- Fundamental problem of time-to-event data analysis: large failure times are censored.
- Potential event times and event indicators:

$$\tilde{T}^a = \min\{T^a, C^a\},\ \delta_T^a = I\{T^a \leq C^a\},$$
$$\tilde{R}^a = \min\{R^a, C^a\},\ \delta_R^a = I\{R^a \leq C^a\}.$$

- Consistency:

$$\tilde{T} = \tilde{T}^A,\ \delta_T = \delta_T^A,\ \tilde{R} = \tilde{R}^A,\ \delta_R = \delta_R^A.$$

Causal Inference to Semi-Competing Risks

- Principal stratification: to restrict the target population on a principal stratum

$$\{R^1 = \infty, R^0 = \infty\},$$

where no intermediate events would happen regardless of treatments (Xu et al., 2022; Nevo and Gorfine, 2022; Gao et al., 2022).

- Mediation analysis: to define mediated potential outcomes on event (counting) processes, so the natural direct effect

$$P(T^{1,0} \leq t)/P(T^{0,0} \leq t)$$

is controlling the prevalence of intermediate events (Huang, 2021).

Separable Effects

- An interventionist approach to mediation analysis attempts to decompose the initial treatment into two segments, each of which only has a direct effect on a single event (Robins and Richardson, 2010; Stensrud et al., 2022).
- Challenge: Since intermediate events may further develop terminal events, the treatment effect on terminal events relies on both the hazard of intermediate events and the heterogeneous cause-specific hazards of terminal events.
- As a result, the total treatment effect can be decomposed into more than two separable effects.

Primary Events Developed from Different Sources

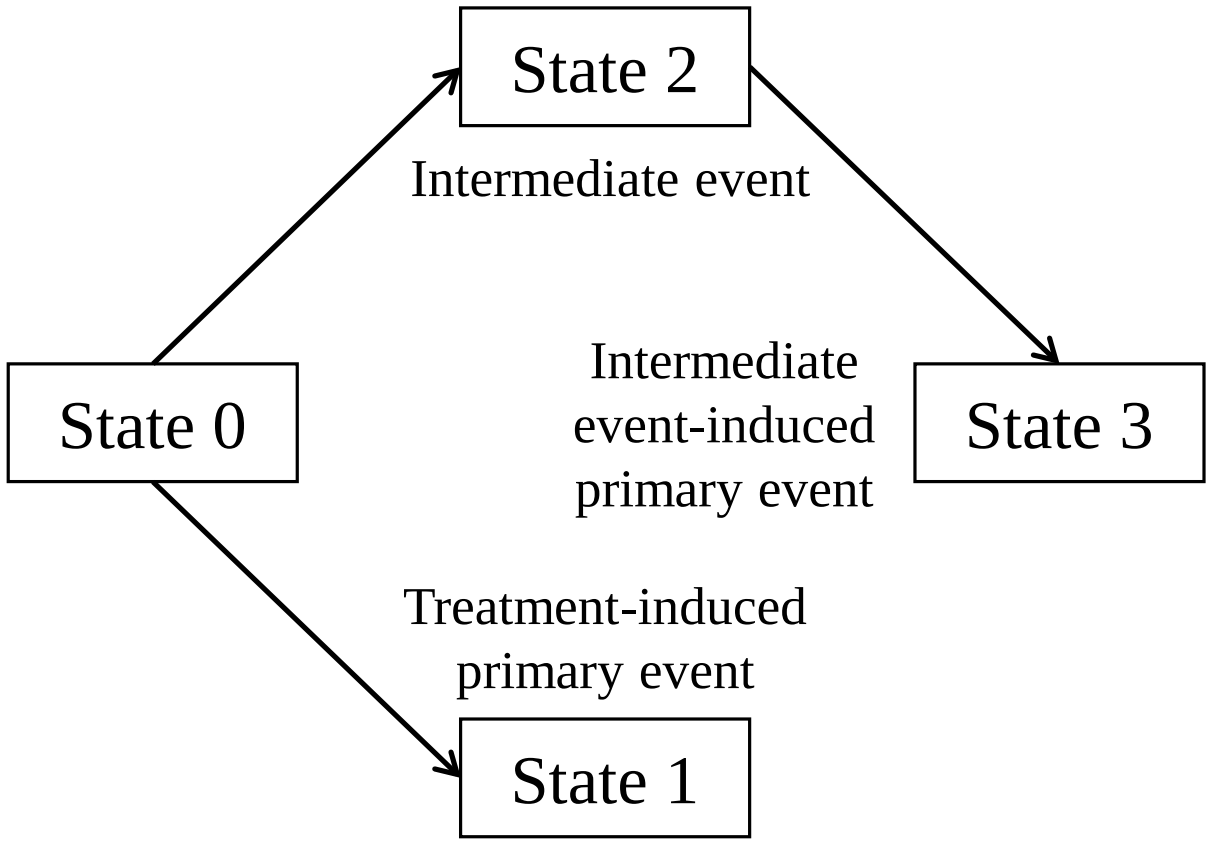

Figure: A multi-state model. Primary events include direct outcome events (treatment-induced) and indirect outcome events (intermediate event-induced).

Outline

Assumptions

- We need the following three assumptions. These assumptions are common in causal inference and survival analysis.
- Assumption 1 (Ignorability): $(T^a, R^a, C^a) \perp\!\!\!\perp A \mid X$.
- Assumption 2 (Random censoring): $I(t \leq C^a < t + dt) \perp\!\!\!\perp (T^a, R^a) \mid \mathcal{F}(t)$.
- Assumption 3 (Positivity): $c < P(A = a \mid X) < 1 - c$ for a constant $c > 0$, $P(T^a > \tau, C^a > \tau \mid A = a, \mathcal{F}(t)) > 0$ for any $0 \leq t \leq \tau$ and $a = 0, 1$.

Potential Hazards (Transition Rates)

- Potential hazards (transition rates) of direct outcome event (State 1), intermediate event (State 2) and indirect outcome event through intermediate event occurring at time r (State 3) at time t:

$$d\Lambda_1^a(t \mid \mathcal{F}(t)) := P(t \leq T^a < t + dt, T^a < R^a \mid T^a \geq t, R^a \geq t, \mathcal{F}(t)),$$
$$d\Lambda_2^a(t \mid \mathcal{F}(t)) := P(t \leq R^a < t + dt, R^a \leq T^a \mid T^a \geq t, R^a \geq t, \mathcal{F}(t)),$$
$$d\Lambda_3^a(t \mid \mathcal{F}(t)) := P(t \leq T^a < t + dt, R^a \leq t \mid T^a \geq t, R^a \leq t, \mathcal{F}(t)).$$

Decomposing Treatment

- Suppose the treatment can be decomposed into three segments (A_1, A_2, A_3), where A_j only has an effect on the hazard of developing into State j.
- Then, potential hazards $\{d\Lambda_j^a(\cdot) : j = 1, 2, 3\}$ can be written as functions of separable treatments $a = (a_1, a_2, a_3)$.
- In a realized trial, observed separable treatments $a_1 = a_2 = a_3$, equal to the actual treatment.
- In a hypothetical world, separable treatments can take different values.

Hazards in Multi-State Models

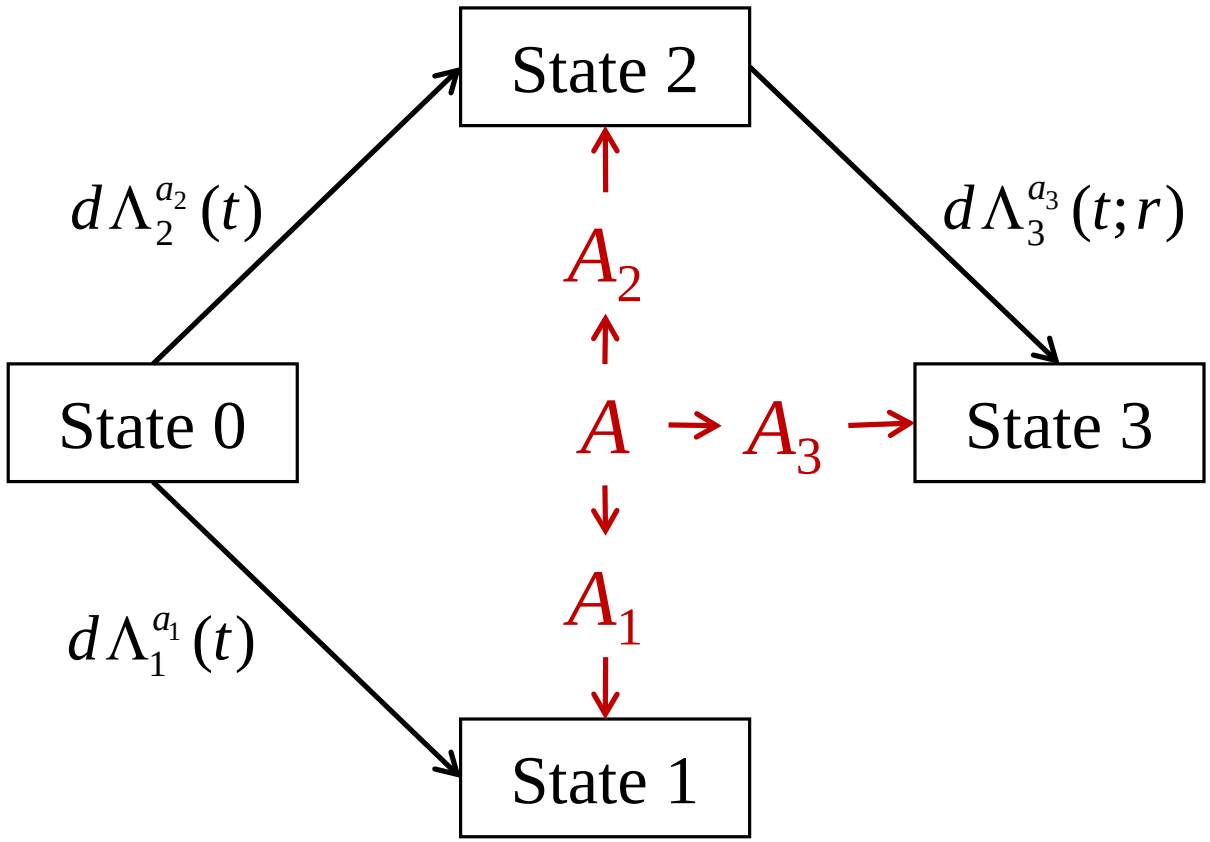

Figure: A multi-state model. Primary events include direct outcome events (treatment-induced) and indirect outcome events (intermediate event-induced).

Dismissible Treatment Components

- Assumption 4 (Dismissible treatment components):

$$d\Lambda_j^{(a_1,a_2,a_3)}(t \mid \mathcal{F}(t)) = d\Lambda_j^{a_j}(t \mid \mathcal{F}(t)).$$

- When $a_1 = a_2 = a_3$, this assumption is naturally satisfied because no hypothetical worlds are involved.
- An example this assumption holds is the additive hazards model.

Full Isolation

- Full isolation refers to the case where the effects of treatment components are separable at the population level.
- $\mathcal{F}(t)$ only involves prior path of states rather than covariates.

$$
\begin{aligned}
d\Lambda_1^a(t \mid \mathcal{F}(t)) &= d\Lambda_1^a(t \mid \emptyset) := d\Lambda_1^a(t), \\
d\Lambda_2^a(t \mid \mathcal{F}(t)) &= d\Lambda_2^a(t \mid \emptyset) := d\Lambda_2^a(t), \\
d\Lambda_3^a(t \mid \mathcal{F}(t)) &= d\Lambda_3^a(t \mid R^a = r) := d\Lambda_3^a(t; r).
\end{aligned}
$$

Counterfactual Cumulative Incidences

- Denote counterfactual cumulative incidences for three states by $F_1^a(t) = P(T^a \leq t, T^a < R^a)$, $F_2^a(t) = P(R^a \leq t, R^a \leq T^a)$ and $F_3^a(t) = P(T^a \leq t, R^a \leq T^a)$, respectively.
- The counterfactual cumulative incidence of primary event becomes $F^a(t) = P(T^a \leq t) = F_1^a(t) + F_3^a(t)$.
- The counterfactual cumulative incidences

$$F_1^{a=(a_1,a_2,a_3)}(t) = \int_0^t \exp\{-\Lambda_1^{a_1}(s) - \Lambda_2^{a_2}(s)\} d\Lambda_1^{a_1}(s),$$

$$F_2^{a=(a_1,a_2,a_3)}(t) = \int_0^t \exp\{-\Lambda_1^{a_1}(s) - \Lambda_2^{a_2}(s)\} d\Lambda_2^{a_2}(s),$$

$$F_3^{a=(a_1,a_2,a_3)}(t) = \int_0^t \exp\{-\Lambda_1^{a_1}(s) - \Lambda_2^{a_2}(s)\}[1 - \exp\{-\Lambda_3^{a_3}(t;s)\}] d\Lambda_2^{a_2}(s).$$

Definition of Separable Pathway Effects

- The total treatment effect is decomposed as

$$
\begin{aligned}
& F^{a=(1,1,1)}(t) - F^{a=(0,0,0)}(t) \\
&= \{F^{a=(1,0,0)}(t) - F^{a=(0,0,0)}(t)\} + \{F^{a=(1,1,1)}(t) - F^{a=(1,0,0)}(t)\} \\
&:= \mathrm{SPE}_{0\to 1}(t;0,0) + \mathrm{SPE}_{0\to 3}(t;1) \\
&= \{F^{a=(1,0,0)}(t) - F^{a=(0,0,0)}(t)\} + \{F^{a=(1,1,0)}(t) - F^{a=(1,0,0)}(t)\} \\
&\quad + \{F^{a=(1,1,1)}(t) - F^{a=(1,1,0)}(t)\} \\
&:= \mathrm{SPE}_{0\to 1}(t;0,0) + \mathrm{SPE}_{0\to 2}(t;1,0) + \mathrm{SPE}_{2\to 3}(t;1,1).
\end{aligned}
$$

Identification of Potential Hazards

Theorem 1 (Identification)

Under Assumptions 1–4 and full isolation, for $0 \le t \le \tau$*, the potential transition hazards*

$$d\Lambda_1^a(t) = \frac{E\{I(t \le \tilde{T} < t + dt, \delta_T(1-\delta_R) = 1, A = a)/P(A = a \mid X)\}}{E\{I(\tilde{T} \ge t, \tilde{R} \ge t, A = a)/P(A = a \mid X)\}},$$

$$d\Lambda_2^a(t) = \frac{E\{I(t \le \tilde{R} < t + dt, \delta_R = 1 \mid A = a)/P(A = a \mid X)\}}{E\{I(\tilde{T} \ge t, \tilde{R} \ge t, A = a)/P(A = a \mid X)\}},$$

$$d\Lambda_3^a(t; r) = \frac{E\{I(t \le \tilde{T} < t + dt, \tilde{R} = r, \delta_T\delta_R = 1, A = a)/P(A = a \mid X)\}}{E\{I(\tilde{T} \ge t, \tilde{R} = r, \delta_R = 1, A = a)/P(A = a \mid X)\}}.$$

The separable pathway effects are identifiable.

Outline

Weighted Counting Processes

- Let $w_i(a_j) = I\{A_i = a_j\}/P(A_i = a_j \mid X_i)$.
- Define weighted counting processes, at-risk processes and residuals with respect to $d\Lambda_1^{a_1}(t)$ and $d\Lambda_2^{a_2}(t)$ as follows:

$$N_1(t; a_1) = \sum_{i=1}^{n} w_i(a_1) I\{\tilde{T}_i \le t, \tilde{R}_i > t, \delta_i^T = 1\},$$

$$N_2(t; a_2) = \sum_{i=1}^{n} w_i(a_2) I\{\tilde{R}_i \le t, \tilde{T}_i \ge t, \delta_i^R = 1\},$$

$$Y_j(t; a_j) = \sum_{i=1}^{n} w_i(a_j) I\{\tilde{T}_i \ge t, \tilde{R}_i \ge t\},$$

$$Y_j^w(t; a_j) = \sum_{i=1}^{n} w_i(a_j)^2 I\{\tilde{T}_i \ge t, \tilde{R}_i \ge t\},$$

$$M_j(t; a_j) = \int_0^t \left\{ dN_j(s; a_j) - Y_j(s; a_j) d\Lambda_j^{a_j}(s) \right\}, \; j = 1, 2.$$

Weighted Counting Processes

- To yield well-defined estimators for $\Lambda_3^{a_3}(t; s)$, processes $N_3(t; r, a_3)$ and $Y_3(t; r, a_3)$ should be refined so that $Y_3(t; r, a_3)$ is nonzero and

$$M_3(t; r, a_3) = \int_r^t \{dN_3(s; r, a_3) - Y_3(s; r, a_3) d\Lambda_3^{a_3}(s; r)\}$$

  is a martingale with respect to some filter.
- To ensure that $Y_3(t; r, a_3)$ is left-continuous, we assume that the intermediate event happens just before the primary event if $R_i = T_i$.
- For example, we can assume Markovness or semi-Markovness for the transition from State 2 to State 3.

Markovness

- Under Markov assumption $d\Lambda_3^{a_3}(t;r) = d\Lambda_{3,\text{ma.}}^{a_3}(t)$, let

$$N_3(t;r,a_3) = N_{3,\text{ma.}}(t;a_3) = \sum_{i=1}^{n} w_i(a_3) I\{\tilde{T}_i \le t, \delta_i^T \delta_i^R = 1\},$$

$$Y_3(t;r,a_3) = Y_{3,\text{ma.}}(t;a_3) = \sum_{i=1}^{n} w_i(a_3) I\{\tilde{T}_i \ge t, \tilde{R}_i \le t, \delta_i^R = 1\},$$

$$Y_3^w(t;r,a_3) = Y_{3,\text{ma.}}^w(t;a_3) = \sum_{i=1}^{n} w_i(a_3)^2 I\{\tilde{T}_i \ge t, \tilde{R}_i \le t, \delta_i^R = 1\},$$

$$M_3(t;r,a_3) = M_{3,\text{ma.}}(t;a_3) = \int_0^t \left\{ dN_{3,\text{ma.}}(s;a_3) - Y_{3,\text{ma.}}(s;a_3) d\Lambda_{3,\text{ma.}}^{a_3}(s) \right\}.$$

Semi-Markovness

- Under semi-Markov assumption $d\Lambda_3^{a_3}(t;r) = d\Lambda_{3,\text{sm.}}^{a_3}(t-r)$, let

$$N_3(t;r,a_3) = N_{3,\text{sm.}}(u;a_3) = \sum_{i=1}^{n} w_i(a_3) I\{\tilde{T}_i - \tilde{R}_i \leq u, \delta_i^T \delta_i^R = 1\},$$

$$Y_3(t;r,a_3) = Y_{3,\text{sm.}}(u;a_3) = \sum_{i=1}^{n} w_i(a_3) I\{\tilde{T}_i - \tilde{R}_i \geq u, \delta_i^R = 1\},$$

$$Y_3^w(t;r,a_3) = Y_{3,\text{sm.}}^w(u;a_3) = \sum_{i=1}^{n} w_i(a_3)^2 I\{\tilde{T}_i - \tilde{R}_i \geq u, \delta_i^R = 1\},$$

$$M_3(t;r,a_3) = M_{3,\text{sm.}}(u;a_3) = \int_0^u \left\{ dN_{3,\text{sm.}}(s;a_3) - Y_{3,\text{sm.}}(s;a_3) d\Lambda_{3,\text{sm.}}^{a_3}(s) \right\}.$$

Estimators for Cumulative Hazards

- The generalized Nelson-Aalen estimators for cumulative hazards are

$$\hat{\Lambda}_1^{a_1}(t) = \int_0^t \frac{dN_1(s; a_1)}{Y_1(s; a_1)}, \ \hat{\Lambda}_2^{a_2}(t) = \int_0^t \frac{dN_2(s; a_2)}{Y_2(s; a_2)},$$
$$\hat{\Lambda}_3^{a_3}(t; r) = \int_r^t \frac{dN_3(s; r, a_3)}{Y_3(s; r, a_3)}.$$

- The residuals $\{M_j(t; a_j) : j = 1, 2, 3(\text{ma. and sm.})\}$ are martingales with respect to filters

$$\mathcal{F}_1^{a_j}(t) = \{w_i(a_1), I(T_i^{a_1} \geq s, R_i^{a_1} \geq s, C_i^{a_1} \geq s) : s \leq t, i = 1, \ldots, n\},$$
$$\mathcal{F}_2^{a_j}(t) = \{w_i(a_2), I(T_i^{a_2} \geq s, R_i^{a_2} \geq s, C_i^{a_2} \geq s) : s \leq t, i = 1, \ldots, n\},$$
$$\mathcal{F}_{3,\text{ma.}}^{a_3}(t) = \{w_i(a_3), I(T_i^{a_3} \geq s, T_i^{a_3} \geq R_i^{a_3}, C_i^{a_3} \geq s) : s \leq t, i = 1, \ldots, n\},$$
$$\mathcal{F}_{3,\text{sm.}}^{a_3}(t) = \{w_i(a_3), I(T_i^{a_3} - R_i^{a_3} \geq s, C_i^{a_3} - R_i^{a_3} \geq s) : s \leq t, i = 1, \ldots, n\},$$

with $\text{var}\{dM_j(t; a_j) \mid \mathcal{F}_j^{a_j}(t)\} = Y_j^w(t; a_j) d\Lambda_j^{a_j}(t)$.

Weak Convergence of Estimators

Theorem 2 (Asymptotic distribution)

Under Assumptions 1–4 and $Y_j^{-1}(\tau; a_j) = o_p(n^{-1/2})$ *for* $j = 1, 2, 3$*(ma. and sm.), the empirical process* $n^{1/2}\{\hat{F}^a(\cdot) - F^a(\cdot)\}$ *converges to*

$$n^{1/2}\{G_1^a(\cdot) + G_2^a(\cdot) + G_3^a(\cdot)\},$$

whose limiting distribution is a Gaussian process.

Asymptotic Convergence: Markovness

Corollary 3 (Asymptotic distribution under Markovness)

Under Assumptions 1–4, if the hazard $d\Lambda_3^{a_3}(t;r)$ *is Markov, i.e.,* $d\Lambda_3^{a_3}(t;r) = d\Lambda_{3,ma.}^{a_3}(t)$ *for every* $r \in [0,\tau]$*, then*

$$n^{1/2}\{\hat{F}^a(t) - F^a(t)\} \xrightarrow{d} N\{0, \sigma^2(t)\}$$

on $t \in [0,\tau]$*, with*

$$\begin{aligned}\sigma^2(t) = \lim E\Bigg\{ &\int_0^t [1 - F^a(t) - \{F_2^a(s) - F_3^a(s)\}\exp\{\Lambda_3^{a_3}(s) - \Lambda_3^{a_3}(t)\}]^2 \frac{nY_1^w(s;a_1)}{Y_1(s;a_1)^2} d\Lambda_1^{a_1}(s) \\ &+ \int_0^t [1 - F^a(t) - \{1 - F^a(s)\}\exp\{\Lambda_3^{a_3}(s) - \Lambda_3^{a_3}(t)\}]^2 \frac{nY_2^w(s;a_2)}{Y_2(s;a_2)^2} d\Lambda_2^{a_2}(s) \\ &+ \int_0^t [\{F_2^a(s) - F_3^a(s)\}\exp\{\Lambda_3^{a_3}(s) - \Lambda_3^{a_3}(t)\}]^2 \frac{nY_3^w(s;a_3)}{Y_3(s;a_3)^2} d\Lambda_3^{a_3}(s) \Bigg\}.\end{aligned}$$

Asymptotic Convergence: Semi-Markovness

Corollary 4 (Asymptotic distribution under semi-Markovness)

Under Assumptions 1–4, if the hazard $d\Lambda_3^{a_3}(t;r)$ *is semi-Markov, i.e.,* $d\Lambda_3^{a_3}(t;r) = d\Lambda_{3,sm.}^{a_3}(t-r)$ *for every* $r \in [0,\tau]$*, then*

$$n^{1/2}\{\hat{F}^a(t) - F^a(t)\} \xrightarrow{d} N\{0, \sigma^2(t)\}$$

on $t \in [0,\tau]$*, with*

$$\begin{aligned}\sigma^2(t) = \lim E\Bigg\{ & \int_0^t \left[1 - F_1^a(t) - F_2^a(t) + \int_s^t \exp\{-\Lambda_3^{a_3}(t-u)\} dF_2^a(u)\right]^2 \frac{nY_1^w(s;a_1)}{Y_1(s;a_1)^2} d\Lambda_1^{a_1}(s) \\ & + \int_0^t \Bigg[\{1 - F_1^a(u) - F_2^a(u)\} \exp\{-\Lambda_3^{a_3}(t-u)\}\Big|_s^t \\ & \qquad + \int_s^t \exp\{-\Lambda_3^{a_3}(t-u)\} dF_2^a(u)\Bigg]^2 \frac{nY_2^w(s;a_2)}{Y_2(s;a_2)^2} d\Lambda_2^{a_2}(s) \\ & + \int_0^t \left[\int_0^{t-s} \exp\{-\Lambda_3^{a_3}(t-u)\} dF_2^a(u)\right]^2 \frac{nY_3^w(s;a_3)}{Y_3(s;a_3)^2} d\Lambda_3^{a_3}(s)\Bigg\}.\end{aligned}$$

Uniform Convergence for Mixtures

Theorem 5 (Uniform convergence)

Suppose $d\Lambda_3^{a_3}(t;r)$ *is a linear combination of* $d\Lambda_{3,ma.}^{a_3}(t)$ *and* $d\Lambda_{3,sm.}^{a_3}(t-r)$,

$$d\Lambda_3^{a_3}(t;r) = (1-\kappa)d\Lambda_{3,ma.}^{a_3}(t) + \kappa d\Lambda_{3,sm.}^{a_3}(t-r)$$

where $\kappa \in [0,1]$ *is a prespecified parameter. Under Assumptions 1–4,*

$$\sup_{t\in[0,\tau]} \left|\hat{F}^a(t) - F^a(t)\right| \xrightarrow{p} 0.$$

Hypotheses

- To detect on which pathways treatment effects exist, we shall conduct hypothesis tests on separable pathway effects $\text{SPE}_{0\to1}$, $\text{SPE}_{0\to2}$ and $\text{SPE}_{2\to3}$.
- We consider testing three potential hazards here:

$$H_0^1 : d\Lambda_1^1(t) = d\Lambda_1^0(t), \forall t \le \tau \text{ vs. } H_1^1 : d\Lambda_1^1(t) \neq d\Lambda_1^0(t), \exists t \le \tau,$$
$$H_0^2 : d\Lambda_2^1(t) = d\Lambda_2^0(t), \forall t \le \tau \text{ vs. } H_1^2 : d\Lambda_2^1(t) \neq d\Lambda_2^0(t), \exists t \le \tau,$$
$$H_0^3 : d\Lambda_3^1(t) = d\Lambda_3^0(t), \forall t \le \tau \text{ vs. } H_1^3 : d\Lambda_3^1(t) \neq d\Lambda_3^0(t), \exists t \le \tau.$$

- When the null hypothesis holds, the corresponding separable pathway effect is zero.

Test Statistics

- Define the weighted logrank statistics with a left-continuous weight $\omega(t)$

$$U_j = \int_0^\tau \omega(t) \frac{Y_j(t;1)dN_j(t;0) - Y_j(t;0)dN_j(t;1)}{Y_j(t;1) + Y_j(t;0)}, \ j = 1, 2, 3(\text{ma. or sm.}).$$

Theorem 6 (Weighted logrank tests)

Under the null hypothesis H_0^j, we have $n^{-1/2}U_j \xrightarrow{d} N(0, \sigma_j^2)$, where

$$\sigma_j^2 = E\left\{ \int_0^\tau \omega(t)^2 \frac{Y_j(t;1)^2 Y_j^w(t;0) + Y_j(t;0) Y_j^w(t;1)}{n\{Y_j(t;1) + Y_j(t;0)\}^2} d\Lambda_j^{a_j}(t) \right\}, \ j = 1, 2, 3(\textit{ma. or}$$

Outline

Partial Isolation

- Partial isolation refers to the case where the effects of treatment components are separable given all history involving covariates and path of states.
- First we assume that $\mathcal{F}(t)$ consists of baseline covariates and current status (Markovian).
- Identification is straightforward by conditioning on baseline covariates.

Estimation Model

- Cox regression is adopted to model three hazard functions for the direct primary outcome event, intermediate event and all-cause primary outcome event respectively,

$$\begin{aligned}\Lambda_1^a(t;x) &= \Lambda_{01}(t)\exp(\beta_{10}+\beta_{1x}x+\beta_{1a}a),\\ \Lambda_2^a(t;x) &= \Lambda_{02}(t)\exp(\beta_{20}+\beta_{2x}x+\beta_{2a}a),\\ \Lambda^a(t;x) &= \Lambda_0(t)\exp(\beta_{00}+\beta_{0x}x+\beta_{0a}a),\end{aligned}$$

  where $\Lambda_{0j}(t)$ $(j=1,2,0)$ is an unknown baseline hazard function.

- The hazard for the indirect event is solved from

$$\Lambda_3^a(t;x)=\int_0^t \frac{dF^a(s;x)-dF_1^a(s;x)}{F_1^a(s;x)+F_2^a(s;x)-F^a(s;x)},$$

  where $F_1^a(t;x)=\int_0^t \exp\{-\Lambda_1^a(s;x)-\Lambda_2^a(s;x)\}d\Lambda_1^a(s;x)$, $F_2^a(t;x)=\int_0^t \exp\{-\Lambda_1^a(s;x)-\Lambda_2^a(s;x)\}d\Lambda_2^a(s;x)$ and $F^a(t;x)=1-\exp\{-\Lambda^a(s;x)\}$.

Time-Varying Covariates

- In the presence of time-varying covariates $X(t)$, we let $\mathcal{F}^{-}(t) = \{X(s), S(s) : s < t\}$ be the information prior to time t including time-varying covariates and prior statuses, and $\mathcal{F}(t) = \mathcal{F}^{-}(t) \cup X(t)$.
- Let $\lambda_j^a(t; \mathcal{F}(t))$ be the hazard of transiting to State $j \in \{1, 2, 3\}$ conditional on the information $\mathcal{F}(t)$ under the hypothetical treatment $a = (a_1, a_2, a_3)$.
- Assumption 4': Dismissible components conditional on $\mathcal{F}(t)$,

$$\lambda_j^a(t; \mathcal{F}(t)) = \lambda_j^{a_j}(t; \mathcal{F}(t)), \ j = 1, 2, 3.$$

Conditional Isolation

- Assumption 2': Positivity,

$$p(\mathcal{F}(t) \mid A) \cdot P(C^A > t \mid A, \mathcal{F}(t)) > 0, \ \forall\ 0 \leq t \leq \tau.$$

- Assumption 3': Censoring does not depend on future information,

$$I(t \leq C^a < t + dt) \perp\!\!\!\perp \mathcal{F}(u) \mid A = a, \mathcal{F}(t), \ u > t, \ \forall\ 0 \leq t \leq \tau.$$

- Thus, the hazard of transition $\lambda_j(t; a_j, \mathcal{F}(t))$ and the hazard of censoring $\lambda_C(t; a_j, \mathcal{F}(t))$ can be identified in the $A = a_j$ treatment group.

Dismissible Covariates

- Assuumption 5: The time-varying covariates $X(t)$ consist of three dismissible components.
- The transition density of covariates $X(t)$ at time t under the hypothetical treatment $a = (a_1, a_2, a_3)$

$$\begin{aligned} p(X(t); a, \mathcal{F}^-(t)) = & \, p(X_2(t) \mid A = a_2, \mathcal{F}^-(t)) \\ & \cdot p(X_1(t) \mid A = a_1, \mathcal{F}^-(t), X_2(t)) \\ & \cdot p(X_3(t) \mid A = a_3, \mathcal{F}^-(t), X_2(t), X_1(t)). \end{aligned}$$

- The counterfactual cumulative incidences can be identified by a continuous generalization of g-formula.

Identification Results 1

- Under Assumptions 1, 2', 3', 4' and 5,

$$F_1^{a=(a_1,a_2,a_3)}(t) = \int_0^t \int_{\{\mathcal{X}(u)\}_0^s} \exp\{-\Lambda_1(s; a_1, \mathcal{F}(s)) - \Lambda_2(s; a_2, \mathcal{F}(s))\}\lambda_1(s; a_1, \mathcal{F}(s)) \prod_0^s \{p(y; a, \mathcal{F}^-(u))dy\}ds,$$

$$F_2^{a=(a_1,a_2,a_3)}(t) = \int_0^t \int_{\{\mathcal{X}(u)\}_0^s} \exp\{-\Lambda_1(s; a_1, \mathcal{F}(s)) - \Lambda_2(s; a_2, \mathcal{F}(s))\}\lambda_2(s; a_2, \mathcal{F}(s)) \prod_0^s \{p(y; a, \mathcal{F}^-(u))dy\}ds,$$

$$F_3^{a=(a_1,a_2,a_3)}(t) = F_2^{a=(a_1,a_2,a_3)}(t) - \int_0^t \int_{\{\mathcal{X}(u)\}_0^s} \exp\{-\Lambda_1(s; a_1, \mathcal{F}(s)) - \Lambda_2(s; a_2, \mathcal{F}(s)) - \Lambda_3(t; a_3, \mathcal{F}(t)) + \Lambda_3(s; a_3, \mathcal{F}(s))\}\lambda_2(s; a_2, \mathcal{F}(s)) \prod_0^s \{p(y; a, \mathcal{F}^-(u))dy\}ds.$$

Identification Results 2

- Another way to identify the counterfactual cumulative incidences is by weighting.
- Let

$$W_1(t; j, a, a_k) = \frac{\exp\{-\Lambda_j(t; a_j, \mathcal{F}(t)\}}{\exp\{-\Lambda_j(t; a_k, \mathcal{F}(t)\}},$$

$$W_2(t; a, a_k) = \prod_0^t \left\{ \frac{P(A = a_2 \mid X_2(t), \mathcal{F}^-(s))}{P(A = a_k \mid X_2(t), \mathcal{F}^-(s))} \frac{P(A = a_k \mid \mathcal{F}^-(s))}{P(A = a_2 \mid \mathcal{F}^-(s))} \right.$$
$$\cdot \frac{P(A = a_1 \mid X_2(s), X_1(s), \mathcal{F}^-(s))}{P(A = a_k \mid X_2(s), X_1(s)\mathcal{F}^-(s))} \frac{P(A = a_k \mid X_2(s), \mathcal{F}^-(s))}{P(A = a_1 \mid X_2(s), \mathcal{F}^-(s))}$$
$$\left. \cdot \frac{P(A = a_3 \mid X_3(s), X_2(s), X_1(s), \mathcal{F}^-(s))}{P(A = a_k \mid X_3(s), X_2(s), X_1(s)\mathcal{F}^-(s))} \frac{P(A = a_k \mid X_3(s), X_2(s), \mathcal{F}^-(s))}{P(A = a_3 \mid X_3(s), X_2(s), \mathcal{F}^-(s))} \right\},$$

$$W_3(t; a_k) = \frac{I(A = a_k)}{\exp\{-\Lambda_C(t; a_k, \mathcal{F}(t))\}}.$$

Identification Results 2

- Under Assumptions 1, 2', 3', 4' and 5,

$$F_1^{a=(a_1,a_2,a_3)}(t) = E\Bigg[\int_0^t W_1(s;2,a,a_1)W_2(s;a,a_1)W_3(s;a_1) \cdot I(s \le \tilde{T} < s+ds, \delta_T(1-\delta_R)=1) \mid A=a_1\Bigg],$$

$$F_2^{a=(a_1,a_2,a_3)}(t) = E\Bigg[\int_0^t W_1(s;1,a,a_2)W_2(s;a,a_2)W_3(s;a_2) \cdot I(s \le \tilde{R} < s+ds, \delta_R=1) \mid A=a_2\Bigg],$$

$$F_3^{a=(a_1,a_2,a_3)}(t) = F_2^{a=(a_1,a_2,a_3)}(t) - E\Bigg[\int_0^t W_1(s;1,a,a_2)W_1(t;3,a,a_2)W_1(s;3,a,a_2)^{-1} \cdot W_2(s;a,a_2)W_3(s;a_2)I(s \le \tilde{T} < s+ds, \delta_T\delta_R=1) \mid A=a_2\Bigg].$$

Outline

Simulation Setup

- We generate a sample with $n = 500$ independent units.
- For $i \in \{1, \ldots, n\}$, a dichotomized covariate X_i equals 1 or 0.5 with equal probability.
- A unit receives treatment $A_i = 1$ with probability $P(X_i/3 + 0.25)$.
- Consider three settings of hazards:
  1. Setting 1: $d\Lambda_1^a(t;x) = 0.15(x+a)dt$, $d\Lambda_2^a(t;x) = 0.1(x+a)dt$, $d\Lambda_3^a(t;x) = 0.2dt$. Both the Markov assumption and semi-Markov assumption are satisfied.
  2. Setting 2: $d\Lambda_1^a(t;x) = 0.04(x+a)tdt$, $d\Lambda_2^a(t;x) = 0.02(x+a)tdt$, $d\Lambda_3^a(t;x) = 0.05tdt$. Only the Markov assumption is satisfied.
  3. Setting 3: $d\Lambda_1^a(t;x) = 0.04(x+a)tdt$, $d\Lambda_2^a(t;x) = 0.02(x+a)tdt$, $d\Lambda_3^a(t;r,x) = 0.1(t-r)dt$. Only the semi-Markov assumption is satisfied.
- Assumptions 1–4 hold under these three settings.

Estimated Cumulative Incidences

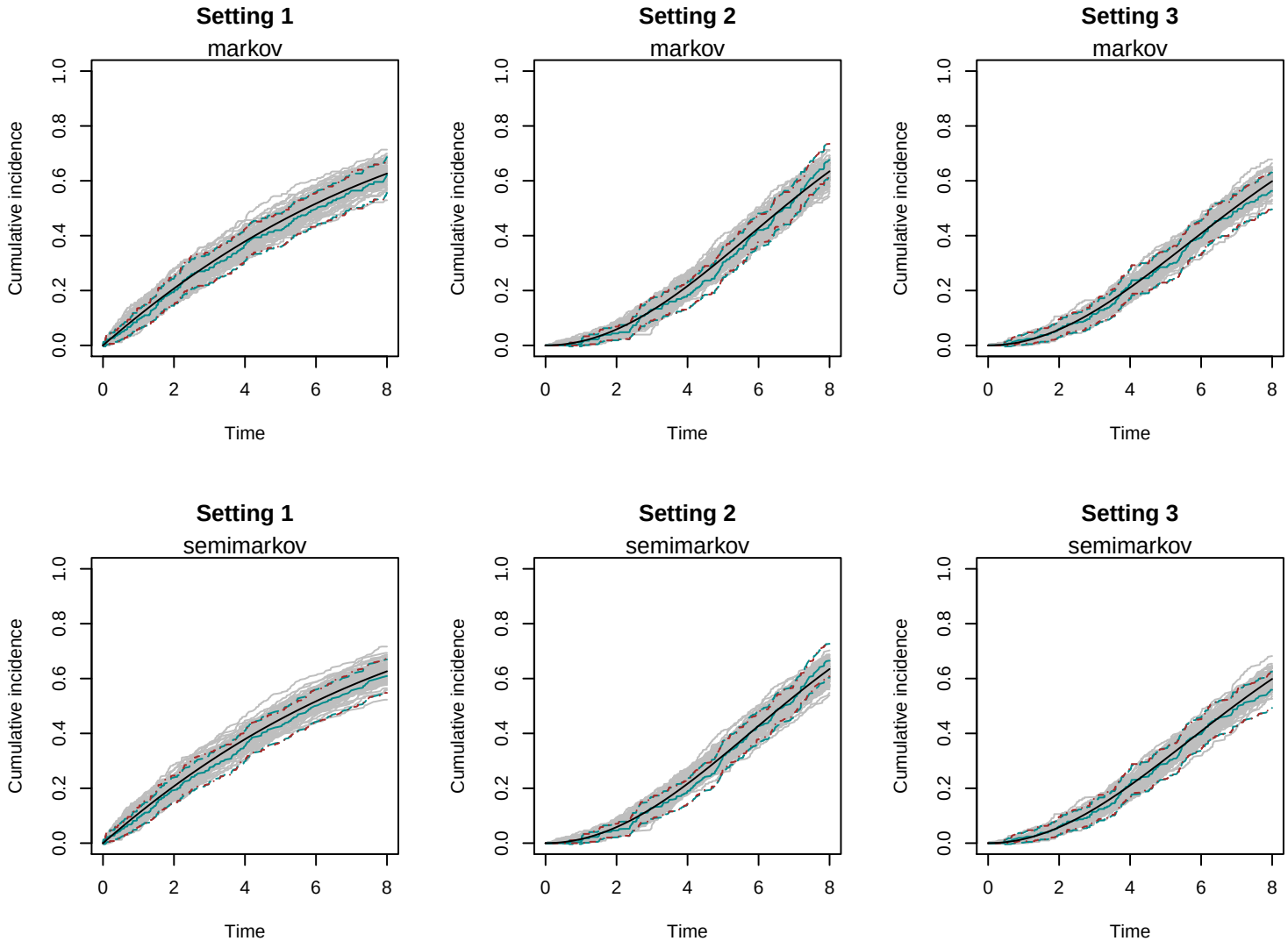

Figure: Estimated cumulative incidence functions for $F^{a=(0,0,0)}(t)$.

Estimated Cumulative Incidences

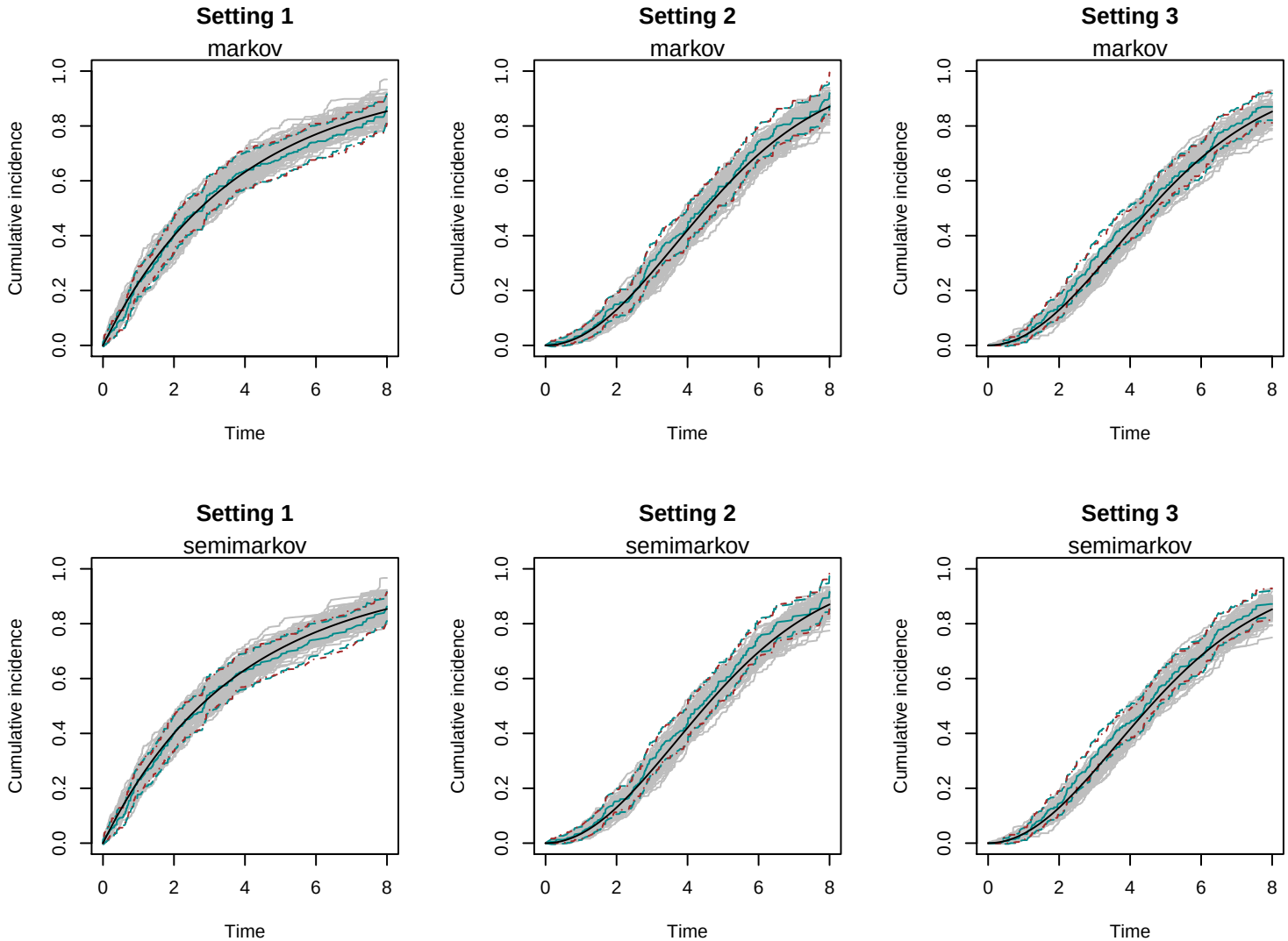

Figure: Estimated cumulative incidence functions for $F^{a=(1,0,0)}(t)$.

Estimated Cumulative Incidences

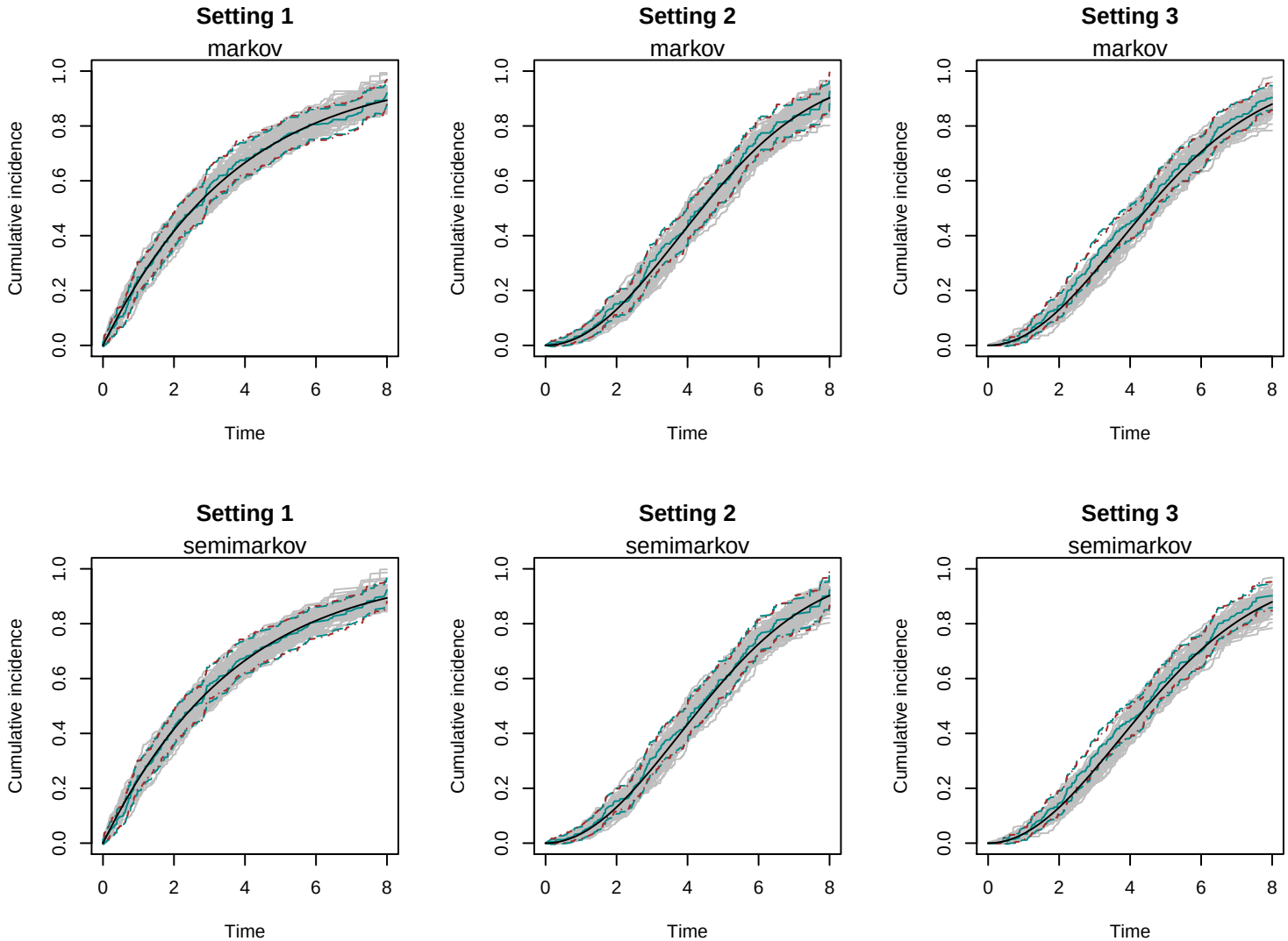

Figure: Estimated cumulative incidence functions for $F^{a=(1,0,1)}(t)$.

Estimated Cumulative Incidences

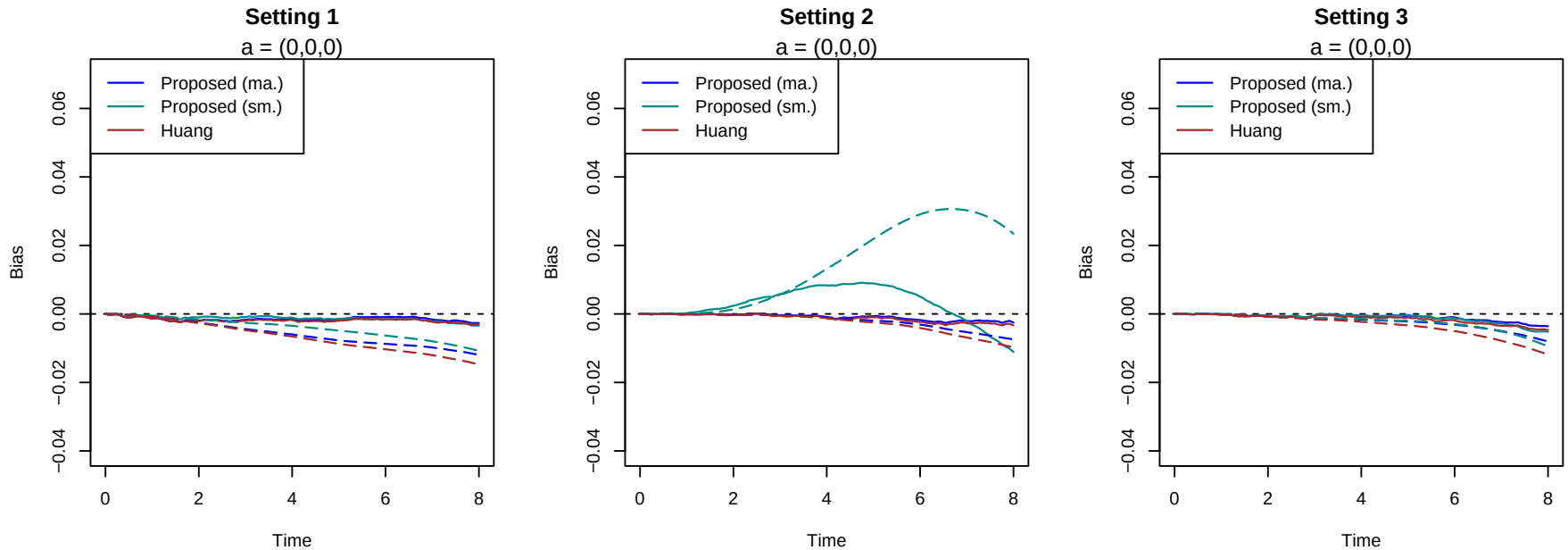

Figure: Pointwise bias and intergrated bias of estimated cumulative incidences of the proposed method using Markov, using semi-Markov, and Huang's methodin estimating $F^{a=(0,0,0)}(t)$.

Estimated Cumulative Incidences

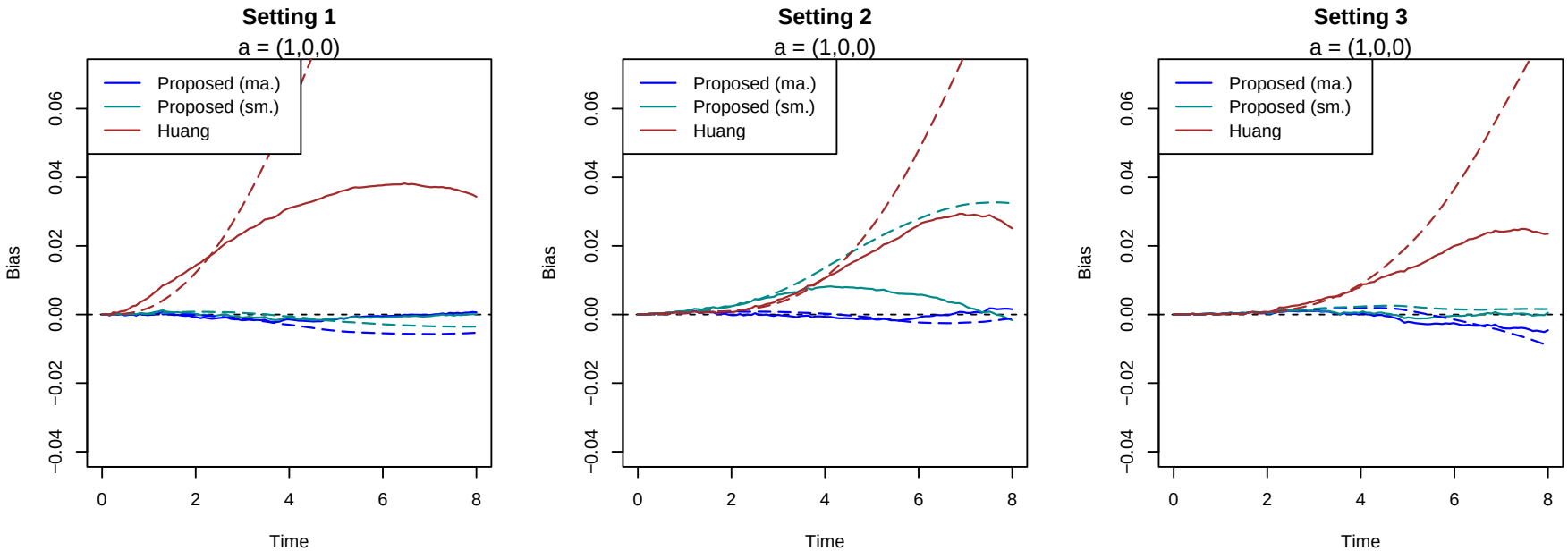

Figure: Pointwise bias and intergrated bias of estimated cumulative incidences of the proposed method using Markov, using semi-Markov, and Huang's method in estimating $F^{a=(1,0,0)}(t)$.

Estimated Cumulative Incidences

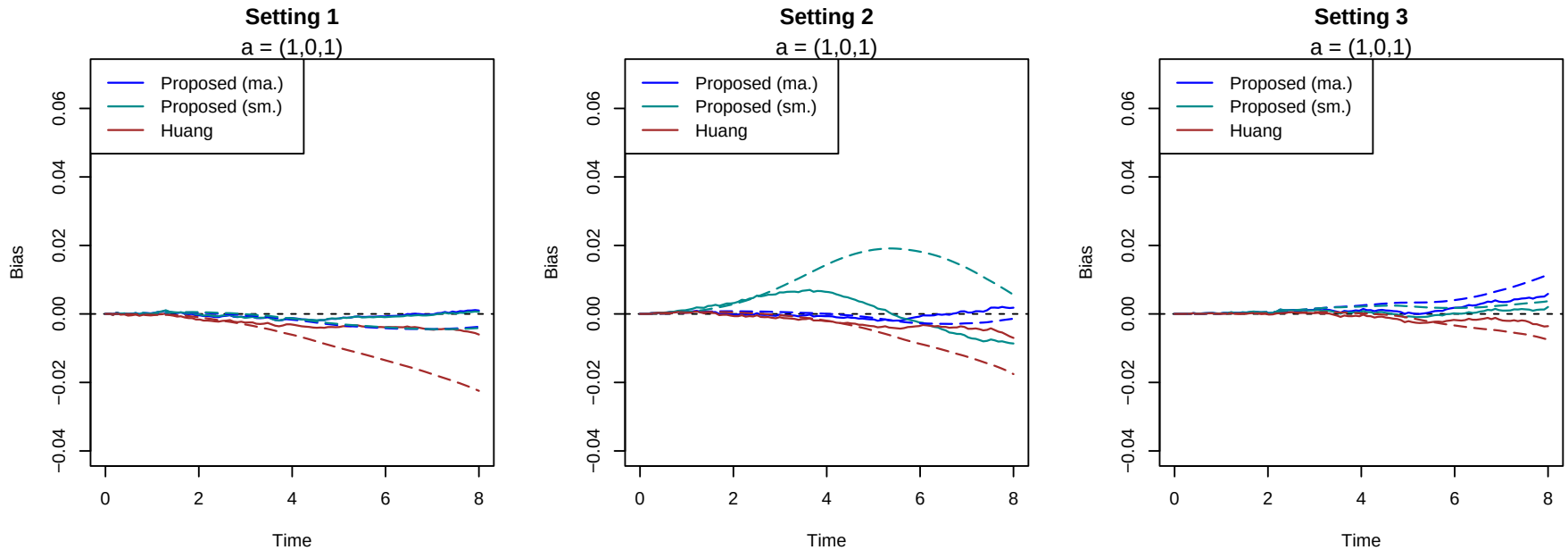

Figure: Pointwise bias and intergrated bias of estimated cumulative incidences of the proposed method using Markov, using semi-Markov, and Huang's methodin estimating $F^{a=(1,0,1)}(t)$.

Empirical Type I Error Rate and Power

Table: The empirical type I error rate (power) of tests

Setting	Hypotheses satisfied	$n = 100$				$n = 500$			
		H_0^1	H_0^2	H_0^3 (ma.)	H_0^3 (sm.)	H_0^1	H_0^2	H_0^3 (ma.)	H_0^3 (sm.)
1	None	0.8539	0.6822	0.3399	0.3580	1.0000	0.9997	0.9488	0.9567
	H_0^1	**0.0453**	0.7757	0.3709	0.3767	**0.0407**	1.0000	0.9675	0.9703
	H_0^2	0.9067	**0.0434**	0.2856	0.3051	1.0000	**0.0446**	0.8774	0.8854
	H_0^3	0.8539	0.6822	**0.0559**	**0.0545**	1.0000	0.9997	**0.0515**	**0.0518**
	H_0^2, H_0^3	0.9067	**0.0434**	**0.0568**	**0.0586**	1.0000	**0.0446**	**0.0508**	**0.0524**
	All	**0.0415**	**0.0474**	**0.0542**	**0.0522**	**0.0412**	**0.0430**	**0.0503**	**0.0502**
2	None	0.8828	0.5918	0.2674	0.1958	1.0000	0.9993	0.8966	0.7338
	H_0^1	**0.0436**	0.6976	0.3003	0.2564	**0.0435**	1.0000	0.9356	0.8760
	H_0^2	0.9190	**0.0450**	0.2305	0.1890	1.0000	**0.0449**	0.7922	0.6501
	H_0^3	0.8828	0.5918	**0.0537**	0.0636	1.0000	0.9993	**0.0504**	0.0780
	H_0^2, H_0^3	0.9190	**0.0450**	**0.0586**	0.0594	1.0000	**0.0449**	**0.0516**	0.0575
	All	**0.0457**	**0.0476**	**0.0537**	0.0544	**0.0471**	**0.0463**	**0.0495**	0.0516
3	None	0.8784	0.5912	0.1977	0.2327	1.0000	0.9985	0.8652	0.8325
	H_0^1	**0.0415**	0.6957	0.1438	0.2404	**0.0449**	0.9999	0.7501	0.8519
	H_0^2	0.9136	**0.0463**	0.1479	0.2080	1.0000	**0.0502**	0.6648	0.7104
	H_0^3	0.8784	0.5912	0.0399	**0.0594**	1.0000	0.9985	0.0810	**0.0558**
	H_0^2, H_0^3	0.9136	**0.0463**	0.0415	**0.0606**	1.0000	**0.0502**	0.0510	**0.0521**
	All	**0.0458**	**0.0450**	0.0317	**0.0580**	**0.0426**	**0.0494**	0.0294	**0.0536**

Outline

Allogeneic Stem Cell Transplantation

- Allogeneic stem cell transplantation is a well applied therapy to treat leukemia, including two sorts of transplant modalities: human leukocyte antigens matched sibling donor transplantation (MSDT) and haploidentical stem cell transplantation from family (Haplo-SCT).
- MSDT has long been regarded as the first choice of transplantation because MSDT leads to lower transplantation related mortality, also known as non-relapse mortality (NRM).
- Another source of mortality is due to relapse, known as relapse related mortality (RRM).
- In recent years, some benefits of Haplo-SCT have been noticed that patients with positive pre-transplantation minimum residual disease (MRD) undergoing Haplo-SCT have better prognosis in relapse.
- The mechanism of how transplant modalities can affect overall survival needs further exploration.

Data

- A total of $n = 303$ patients with positive MRD undergoing allogeneic stem cell transplantation are included in our study.
- Among these patients, 65 received MSDT ($A_i = 1$) and 238 received Haplo-SCT ($A_i = 0$).
- Let R_i be the time of relapse and T_i be the time of death after transplantation.
- In the MSDT group, 47.7% individuals were observed to experience relapse and 53.8% mortality. In the Haplo-SCT group, 29.0% individuals were observed to experience relapse, and 36.6% mortality.
- Age, disease status (CR1 or CR>1), diagnosis (T-ALL or B-ALL) are found to be related with relapse or mortality.

Real Data Application

- In the multi-state model, States 1, 2 and 3 refer to NRM, relapse and RRM, respectively.
- The total effect of mortality can be decomposed into three separable pathway effects: one on NRM $\text{SPE}_{0\to1}(t;0,0)$, one on relapse $\text{SPE}_{0\to2}(t;1,0)$, and the other on RRM $\text{SPE}_{2\to3}(t;1,1)$.
- We maintain the semi-Markov assumption, because RRM usually happens soon after relapse, making it reasonable to assume that the hazard of RRM after relapse relies on how long it passed after relapse rather than the duration from transplantation to relapse.

Estimated Separable Pathway Effects

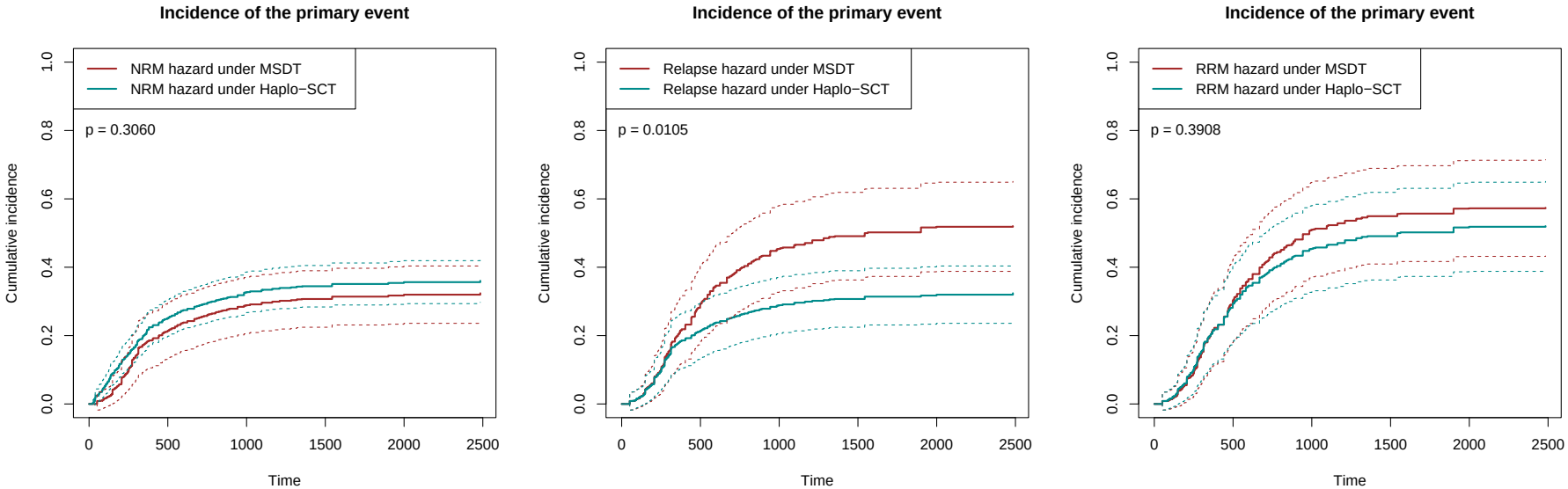

Figure: Counterfactual cumulative incidences of mortality, compared between (1) $F^{a=(1,0,0)}(t)$ and $F^{a=(0,0,0)}(t)$, (2) $F^{a=(1,1,0)}(t)$ and $F^{a=(1,0,0)}(t)$, (3) $F^{a=(1,1,1)}(t)$ and $F^{a=(1,1,0)}(t)$.

Estimated Separable Pathway Effects, Partial Isolation

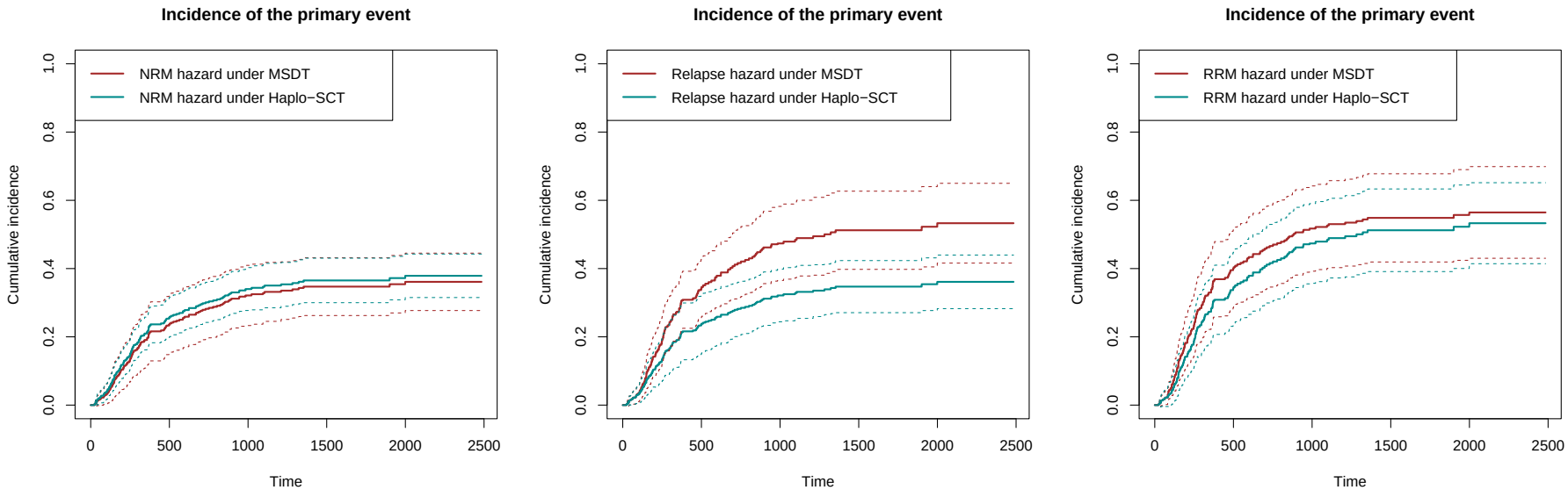

Figure: Counterfactual cumulative incidences of mortality, compared between (1) $F^{a=(1,0,0)}(t)$ and $F^{a=(0,0,0)}(t)$, (2) $F^{a=(1,1,0)}(t)$ and $F^{a=(1,0,0)}(t)$, (3) $F^{a=(1,1,1)}(t)$ and $F^{a=(1,1,0)}(t)$.

Sensitivity Analysis on Semi-Markovness

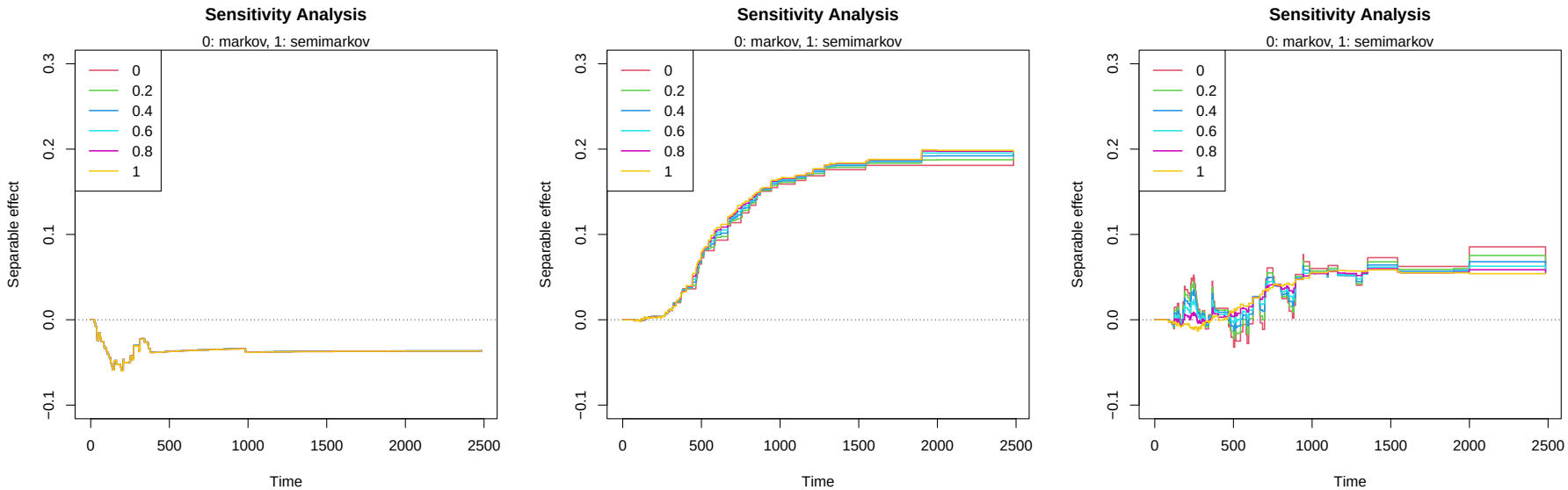

Figure: Sensitivity analysis of separable pathway effects, compared between (1) $F^{a=(1,0,0)}(t)$ and $F^{a=(0,0,0)}(t)$, (2) $F^{a=(1,1,0)}(t)$ and $F^{a=(1,0,0)}(t)$, (3) $F^{a=(1,1,1)}(t)$ and $F^{a=(1,1,0)}(t)$.

P-Values of Tests

Table: Some hypothesis tests in the leukemia data

Test	Interpretation	*p*-value
Total	The total treatment effect on mortality	0.0315
$\text{SPE}_{0\to1}$	The treatment effect on transition rates from transplantation to NRM, i.e., the separable pathway effect via NRM	0.3060
$\text{SPE}_{0\to2}$	The treatment effect on transition rates from transplantation to relapse, i.e., the separable pathway effect via relapse	0.0105
$\text{SPE}_{2\to3}$	The treatment effect on transition rates from relapse to RRM, i.e., the separable pathway effect via RRM (assuming semi-Markov)	0.3098

Results

- We find that Haplo-SCT lowers the overall mortality by reducing the risk of relapse compared with MSDT.
- The differences between Haplo-SCT and MSDT on NRM and RRM are not significant.
- Since Halpo-SCT is more accessible than MSDT, Haplo-SCT can serve as an alternative to MSDT.

Outline

Concluding Remarks

- We studied the identification and estimation of the counterfactual cumulative incidence of the primary (terminal) event when there is an intermediate (non-terminal) event.
- Covariates are incorporated in the hazards by inverse probability weighting.
- We define population-level separable pathway effects based on counterfactual cumulative incidences.
- Confidence intervals and hypothesis testings are available for the counterfactual cumulative incidences and separable pathway effects.
- The concept of separable pathway effects provides an opportunity to understand the causal pathways of treatment effects on the terminal event.

Possible Extensions

- Semiparametrically efficient estimation under full isolation and partial isolation.
- Identification under partial isolation or with time-varying covariates has been proved. But estimation and inference are challenging.

Acknowledgements

- We thank Dr. Yingjun Chang for discussing the background of real-data application.
- Funding information: National Key Research and Development Program of China, Grant No. 2021YFF0901400; National Natural Science Foundation of China, Grant No. 12026606, 12226005. This work is also partly funded by Novo Nordisk A/S.

Main References

[1] Fine, J. P., H. Jiang, and R. Chappell (2001). On semi-competing risks data. Biometrika 88 (4), 907-919.
[2] Gao, F., F. Xia, and K. C. G. Chan (2022). Defining and estimating subgroup mediation effects with semi-competing risks data. Statistica Sinica.
[3] Huang, Y.-T. (2021). Causal mediation of semicompeting risks. Biometrics 77 (4), 1143–1154.
[4] Nevo, D. and M. Gorne (2022). Causal inference for semi-competing risks data. Biostatistics 23 (4), 1115–1132.
[5] Robins, J. M. and T. S. Richardson (2010). Alternative graphical causal models and the identification of direct effects. Causality and psychopathology: Finding the determinants of disorders and their cures 84, 103–158.
[6] Stensrud, M. J., J. G. Young, V. Didelez, J. M. Robins, and M. A. Hernan (2022). Separable effects for causal inference in the presence of competing events. Journal of the American Statistical Association 117 (537), 175–183.
[7] Xu, J., J. D. Kalbeisch, and B. Tai (2010). Statistical analysis of illness-death processes and semicompeting risks data. Biometrics 66 (3), 716–725.
[8] Xu, Y., D. Scharfstein, P. Muller, and M. Daniels (2022). A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semicompeting risks. Biostatistics 23 (1), 34-49.

Welcome for discussion!