Checklist

- [x] Introduction
- [x] Construction of counterfactuals
- [x] Equivalence with sequential criteria
- [x] Full vs partial sequential exchangeability.
- [x] Comparison of the identifying functionals ((Ryalen, 2024) and (Rytgaard et al., 2022)).
- [ ] Uniqueness of identifying formula (Works with more than two treatment levels?)
- [ ] Right-censoring (Need EIF under right-censoring + construction of counterfactuals?)
- [ ] Comparison with CAR

Identification and Estimation of Causal Effects under Treatment-Assigned-At-Visit Interventions in Continuous Time

Johan S. Ohlendorff
johan.ohlendorff@sund.ku.dk
University of Copenhagen

Pål Ryalen
pal.ryalen@medisin.uio.no
University of Oslo

Kjetil Røysland
kjetil.roysland@medisin.uio.no
University of Oslo

Anders Munch
a.munch@sund.ku.dk
University of Copenhagen

Thomas A. Gerds
tag@biostat.ku.dk
University of Copenhagen

Abstract

Marginal structural models (MSMs) offers an approach to estimating longitudinal causal effects in discrete time. However, their reliance on fixed-grid data limits their applicability to irregularly spaced observations. Rytgaard et al. (2022) investigated a continuous-time jump process setting where treatment intervention was applied to assigned treatment, but not the timing of visits. This article provides a formal proof of the causal interpretation of the estimands, constructs counterfactual distributions, discusses the uniqueness of the identifying formula, and compares the findings with recent work by Ryalen (2024).

Keywords causal inference · continuous-time · coarsening at random · treatment-assigned interventions

1 Introduction

Robins' theory of causal inference for complex longitudinal data structures offers a robust framework for identifying causal effects when treatments and covariates are measured at discrete time points (Robins (1986)). This approach seeks to estimate the counterfactual mean outcome under a specific treatment intervention. This pursuit led to the development of the g-formula, a method that identifies the counterfactual mean under specific conditions, via the observed data distribution (Robins (1986), Robins et al. (2000), Gill & Robins (2001)). Crucially, the g-formula relies on several key assumptions, including consistency, sequential exchangeability, and positivity.

However, many real-world scenarios involve continuous-time processes where treatments and covariates can change at subject-specific times, representing a significant challenge for traditional discrete-time approaches. To address this, Lok (2008) extended Robins' framework to continuous time using nested structural models, enabling causal inference in continuous-time settings through the estimation of a counterfactual outcome process, particularly within survival analysis contexts. Despite these advancements, these models still place strong structural assumptions.

Subsequently, it was discussed in Gill & Robins (2023) how to obtain a general continuous-time g-formula within counting process settings, extending the discrete-time g-formula of Gill & Robins (2023) to accommodate marginal structural models in continuous time. Rytgaard et al. (2022) explored a specific intervention where treatment timing remains constant but treatment decisions are made at each visit, providing a continuous-time g-formula without a rigorous proof and accompanying estimation techniques. More recently, Ryalen (2024) developed a potential outcomes framework for causal inference in continuous time, establishing conditions for the identifiability of the counterfactual mean outcome under continuous-time treatment interventions. Despite similarities in their approaches, the identification formulae may not generally be the same. In many practical cases, the identification formulae do however coincide. Here, uniqueness of the identifying functional is also discussed.

We establish formal conditions under which the g-formula in Rytgaard et al. (2022) identifies the counterfactual mean outcome under a continuous-time treatment intervention. While that work introduced a g-formula and a sequential exchangeability condition, it did not provide a proof of identifiability. Here, we present such a proof under slightly modified but intuitively related conditions and derive an equivalent martingale formulation. Our approach fits naturally within the potential outcomes framework of Ryalen (2024).

Notably, our exchangeability condition admits a natural extension to a full exchangeability assumption, which is equivalent to coarsening at random (CAR) – an extension not available in Ryalen (2024). We also provide an example illustrating that full exchangeability is strictly stronger than standard exchangeability. (Finally, we extend Theorem 25.40 of van der Vaart (1998) to show that CAR and only CAR implies a saturated model for the observed data distribution, as the original theorem's conditions are not directly applicable in the continuous-time setting.)

1.1 A (hypothetical) motivating application

Consider a longitudinal study as an example. A subject's baseline measurements $L(0)$ and initial treatment assignment $A(0)$ are recorded. The subject is then followed over the interval $[0, T]$ with visits to the clinic/hospital. At each visitation event, the patient either gets their blood measurements taken (ℓ event), or the doctor may decide on treatment based on the patients history so far (a event). We define $N^a(t)$ and $N^\ell(t)$ as the number of treatment and covariate visits, respectively, up to time t and let $A(t)$ and $L(t)$ denote the measurements and treatment decision at time t. The outcome, Y_t is typically a function of the patient's history up to time t, $Y_t = \sigma\big(L(\cdot \wedge t), N^\ell(\cdot \wedge t)\big)$ or it may consist of a separate component such as a primary event, say $Y_t = N^y$.

To illustrate this further, consider a randomized trial. A common scenario is that patients receive treatment at each doctor's visit. However, in real-world practice, patients may experience adverse effects, leading the doctor to discontinue treatment. We're interested in understanding what would have happened **had** the patient continued to receive treatment. We represent this counterfactual outcome as $\tilde{Y}_t$. Therefore, we wish to estimate $\mathbb{E}\big[\tilde{Y}_t\big]$. However, this counterfactual outcome is not observed directly for all subjects, presenting a key challenge.

2 Notation and setup

Let $(\Omega, \mathcal{F}, P)$ be a measure space. We assume that all measurements are assumed to take place over a time interval $[0, T]$ where $t = 0$ denotes baseline and $T > 0$ denotes the time to end-of-study. First, we formulate the setting of Rytgaard et al. (2022). We observe trajectories of the process $\zeta(t) = \big(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t)\big)$, where $\zeta(t)$ is a multivariate jump process on $[0, T]$.

We have $\Delta L(t) \neq 0$ only if $\Delta N^\ell(t) \neq 0$ and $\Delta A(t) \neq 0$ only if $\Delta N^a(t) \neq 0$. To simplify things a bit, we suppose that

$$P\Big(\cap_{t \in [0,T]} (A(t) \in \mathcal{A})\Big) = P\Big(\cap_{t \in [0,T]} (L(t) \in \mathcal{L})\Big) = 1$$

and

$$\mathcal{A} = \Big\{a_1, ..., a_{d_a}\Big\} \subseteq \mathbb{R},$$
$$\mathcal{L} = \Big\{l_1, ..., l_{d_l}\Big\} \subseteq \mathbb{R}^k.$$

This means that all considered processes are jump processes. It is implicitly assumed that $\big(N^d, N^y, N^a, N^\ell\big)$ forms a multivariate counting process (Andersen et al. (1993)). Importantly, we also make the assumption of no explosion of N which is that $P\Big(\sum_{x=a,\ell,d,y} N^x(T) < \infty\Big) = 1$. Now we can let, for $n \geq 1$

$$T_{(n)} = \inf\Big\{t > T_{(n-1)} : N^y(t) + N^a(t) + N^\ell(t) > n\Big\} \text{ with } T_{(0)} := 0.$$

(as is convention with point processes, we let $T_{(n)} = \infty$ if $T_{(n)} > T$). These values are possibly infinite; then we can let

$$Z_{(n)} := \left(N^y\left(T_{(n)}\right), N^a\left(T_{(n)}\right), N^\ell\left(T_{(n)}\right), A\left(T_{(n)}\right), L\left(T_{(n)}\right)\right).$$

Then, the marked point process given by $\Phi = \left(T_{(n)}, Z_{(n)}\right)_{n\geq 1}$ generates the same natural filtration as the process $\zeta(t)$ (Theorem 2.5.10 of Last & Brandt (1995)), that is $\mathcal{F}_t = \sigma((L(0), A(0)), \Phi_t) = \sigma(\zeta(s), s \leq t)^*$. Let further $\Delta_{(k)}$ be given by $\Delta_{(k)} = x$ if $\Delta N^x\left(T_{(k)}\right) = 1$ (with the empty mark if $T_{(k)} = \infty$. Let $\bar{\Phi} = \left(T_{(n)}, \bar{Z}_{(n)}\right)_{n\geq 1}$ be a marked point process given by

$$\bar{Z}_n = \left(\Delta_{(n)}, A\left(T_{(n)}\right), L\left(T_{(n)}\right)\right).$$

Then, since there is a measurable bijection $\bar{\Phi}$ and Φ, we have that $\mathcal{F}_t = \sigma\left((L(0), A(0)), \bar{\Phi}_t\right)$, and we can work with $\bar{\Phi}$ instead of Φ. Intuitively, this means that the information obtained from the multivariate jump process is the same as that obtained from the marked point process at time t. Importantly, we shall work within a so-called canonical setting which allows us to write down explicit formulae for the compensators in terms of the mark and event time distributions (Last & Brandt (1995)). We can let N^{a,a_j} be given by

$$N^{a,a_j}(t) := \sum_k \mathbb{1}\left\{T_{(k)} \leq t, \Delta_{(k)} = a, A\left(T_{(k)}\right) = a_j\right\}$$

Let $\Lambda^{a,a_j}(t)$ denote the P-$\mathcal{F}_t$-compensator of N^{a,a_j} and $\Lambda^a(t) = \sum_{j=1}^k \Lambda^{a,a_j}(t)$ denote the total P-$\mathcal{F}_t$-compensator of N^a. By Lemma 1, we can find kernels $\pi_t(\mathrm{d}x)$ such that

$$\Lambda^{a,a_j}(\mathrm{d}t) = \pi_t\left(\left\{a_j\right\}\right)\Lambda^a(\mathrm{d}t). \tag{1}$$

Lemma 1: Let $\mathcal{F}_t$ denote the natural filtration of $\bar{\Phi}$. Then, the (predictable) stochastic kernel π from $\mathcal{A} \times \mathcal{L} \times N_{\{a,y,\ell,d\}\times\mathcal{A}\times\mathcal{L}} \times \mathbb{R}^+$ to $\mathcal{A}$

$$\pi\left(a(0), l(0), \bar{\varphi}, t, \left\{a_j\right\}\right) = \sum_k \mathbb{1}\{\rho_{k-1}(\bar{\varphi}) < t \leq \rho_k(\bar{\varphi})) P\left(A\left(T_{(k)}\right) = a_j \mid T_{(k)} = t, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}} = \bar{\varphi}^{k-1}\right) \tag{2}$$

satisfies Equation 1 where $\bar{\varphi}^k$ simply consists of the random variables $\left(L(0), A(0), ..., L\left(T_{(k)}\right), A\left(T_{(k)}\right), \Delta_{(k)}, T_{(k)}\right)$. Any kernel π from $\mathcal{A} \times \mathcal{L} \times N_{\{a,y,\ell,d\}\times\mathcal{A}\times\mathcal{L}} \times \mathbb{R}^+$ to $\mathcal{A}$ fulfilling Equation 1 identically satisfies

$$\pi\left(A(0), L(0), \bar{\Phi}, T_{(k)}, \left\{a_j\right\}\right) = P\left(A\left(T_{(k)}\right) = a_j \mid T_{(k)}, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right)$$

P-a.s. whenever $T_{(k)} < \infty$.

Proof: Note that

$$\begin{aligned}
\mu_k(\mathrm{d}(t,\delta,a,l)) &:= \frac{P\left(\left(T_{(k)}, \bar{Z}_{(k)}\right) \in \mathrm{d}(t,\delta,a,l) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&= P\left(\left(A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right) \in \mathrm{d}(a,l) \mid T_{(k)} = t, \Delta_{(k)} = \delta, \mathcal{F}_{T_{(k-1)}}\right) \frac{P\left(\left(T_{(k)}, \Delta_{(k)}\right) \in \mathrm{d}(t,\delta) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \\
&= \Big\{P\left(A\left(T_{(k)}\right) \in \mathrm{d}a \mid T_{(k)} = t, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right)\delta_{L\left(T_{(k-1)}\right)}(\mathrm{d}l)\mathbb{1}\{\delta = a\} \\
&\quad + P\left(L\left(T_{(k)}\right) \in \mathrm{d}l \mid T_{(k)} = t, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right)\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a)\mathbb{1}\{\delta = \ell\} \\
&\quad + \delta_{A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)}(\mathrm{d}(a,l))\mathbb{1}\{\delta = y\}\Big\} \frac{P\left(\left(T_{(k)}, \Delta_{(k)}\right) \in \mathrm{d}(t,\delta) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)}
\end{aligned}$$

According to Theorem 4.1.11 (ii) of Last & Brandt (1995), we have

*As make the assumption of no explosions, the *minimal* jump process and the jump process are not different ensuring a unique measurable correspondence between the jump process and marked point process since the visitation counting process are included in ζ (Theorem 2.5.10-2.5.11 of Last & Brandt (1995))

$$\Lambda(\mathrm{d}(t,\delta,a,l)) = \sum_{k=1}^{\infty} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right)\mu_k(\mathrm{d}(t,\delta,a,l))$$

is the P-$\mathcal{F}_t$ compensator associated with the point process $\bar{\Phi}$. From this, we obtain versions of the compensators on the right-hand side and the left hand side of Equation 1 integrating over sets of the form $(0,t] \times \{a\} \times \{a_j\} \times \mathcal{L}$ and $(0,t] \times \{a\} \times \mathcal{A} \times \mathcal{L}$. Thus we obtain the first desired statement.

NOTE: Fix notation. Now use this with Theorem 4.3.2 of Last & Brandt (1995). Then, we have (almost surely)

$$\begin{aligned}
&P\left(\Delta_{(k)} \in \mathrm{d}\delta \mid T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)\left\{P\left(A\left(T_{(k)}\right) \in \mathrm{d}a \mid T_{(k)}, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right)\delta_{L\left(T_{(k-1)}\right)}(\mathrm{d}l)\mathbb{1}\{\delta = a\}\right.\\
&\left.+P\left(L\left(T_{(k)}\right) \in \mathrm{d}l \mid T_{(k)}, \Delta_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right)\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a)\mathbb{1}\{\delta = \ell\} + \delta_{A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)}(\mathrm{d}(a,l))\mathbb{1}\{\delta = y\}\right\}\\
&= \kappa\left(T_{(k)}, \mathrm{d}(\delta,a,\ell)\right) \qquad (3)\\
&= \kappa_{\Delta}\left(T_{(k)}, \mathrm{d}\delta\right)\left\{\pi'\left(T_{(k)}, \delta, \mathrm{d}a\right)\delta_{L\left(T_{(k-1)}\right)}(\mathrm{d}l)\mathbb{1}\{\delta = a\}\right.\\
&\left.+\kappa_L\left(T_{(k)}, \delta, \mathrm{d}\ell\right)\delta_{A\left(T_{(k-1)}\right)}(\mathrm{d}a)\mathbb{1}\{\delta = \ell\} + \delta_{A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)}(\mathrm{d}(a,l))\mathbb{1}\{\delta = y\}\right\}
\end{aligned}$$

The last equality comes from considering the various cases with a discrete mark space. We see that again $\pi'(t,\delta,\mathrm{d}a)$ satisfies Equation 1 again. However, by explicitly integrating out various sets of Equation 3, we get that $\pi'(t,\delta,\mathrm{d}a) = P\left(A\left(T_{(k)}\right) \in \mathrm{d}a \mid T_{(k)}, \Delta_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right)$. □

3 Identification of the counterfactual mean outcome

Now, we let Now let $N^a(\mathrm{d}t \times \mathrm{d}x) = \sum_{x \in \mathcal{A}} \delta_x(\mathrm{d}x) N^{a,a_j}(\mathrm{d}t)$. We are interested in the outcome process Y under an intervention g^* which we denote by $\tilde{Y}$. Importantly, the intervention is defined as a static/dynamic intervention

$$N^{g^*}(\mathrm{d}t \times \mathrm{d}x) = \pi_t^*(\mathrm{d}x) N^a(\mathrm{d}t \times \mathcal{A})$$

where $\pi_t^*(\mathrm{d}x)$ is some kernel that specifies the treatment decision deterministically at time t in the sense that there are $\mathcal{F}_{T_{(k-1)}} \otimes \mathcal{B}([0,T])$-measurable functions g_k^* which return a treatment decision in $\mathcal{A}$ such that

$$\pi_t^*(\mathrm{d}x) = \sum_k \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_{g_k^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)}(\mathrm{d}x),$$

This means that, critically, $N^{g^*}(\mathrm{d}t \times \mathrm{d}x)$ is also a random measure. Note that N^{g^*} has the compensator

$$\mathcal{L}(N)(\mathrm{d}t \times \mathrm{d}x) = \pi_t^*(\mathrm{d}x)\underbrace{\Lambda^a(\mathrm{d}t \times \mathcal{A})}_{=:\Lambda^a(\mathrm{d}t)},$$

where $\Lambda^a(\mathrm{d}t)$ is the P-$\mathcal{F}_t$-compensator of $N^a(\mathrm{d}t \times \mathcal{A})$ – also deemed the total P-$\mathcal{F}_t$-compensator of N^a. We shall write similar notations for the other components of N. Let $\mathcal{L}$ denote the P-$\mathcal{F}_t$-canonical compensator of N^{g^*}. However, N^a generally has the compensator $\Lambda^a(\mathrm{d}t \times \mathrm{d}x) = \pi_t(\mathrm{d}x)\Lambda^a(\mathrm{d}t)$. Now define the time to deviation from the treatment regime as

$$\tau^{g^*} = \inf\{t \geq 0 \mid N^a((0,t] \times \{x\}) \neq N^{g^*}((0,t] \times \{x\}) \text{ for some } x \in \mathcal{A}\}.$$

Definition 1: Let $\widetilde{\mathcal{F}}_t \coloneqq \sigma\left(\widetilde{N}^d((0,s]), \widetilde{N}^y((0,s]), \widetilde{N}^a((0,s] \times \{x\}), \widetilde{N}^\ell((0,s] \times \{y\}) \mid s \in (0,t], x \in \mathcal{A}, y \in \mathcal{L}\right)$. Let $\dot{\Lambda}$ denote the canonical P-$\mathcal{F}_t$-compensator of N.

A multivariate random measure $\widetilde{N} = \left(\widetilde{N}^y, \widetilde{N}^a, \widetilde{N}^\ell\right)$ is a **counterfactual random measure** under the intervention g^* if it satisfies the following conditions.

1. $\widetilde{N}^a$ has compensator $\mathcal{L}\left(\widetilde{N}\right)$ with respect to $\widetilde{\mathcal{F}}_t$.
2. $\widetilde{N}^x$, has the same compensator $\dot{\Lambda}^x\left(\widetilde{N}\right)$ with respect to $\widetilde{\mathcal{F}}_t$ for $x \in \{y, \ell, d\}$.

Now let $\left(T_{(k)}\right)_k$ denote the ordered event times of N. The main outcome of interest is the counterfactual outcome process $\tilde{Y} \coloneqq \widetilde{N}^y$; and we wish to identify $\mathbb{E}_P\left[\tilde{Y}_t\right]$. Also note that this definition aligns closely with the definition of potential outcomes in discrete time based on structural equations (Richardson & Robins (2013)).

Now we come to a martingale result providing identification.

Theorem 1: If *all* of the following conditions hold:

- **Consistency**: $\tilde{Y}_{\cdot}\mathbb{1}\{\tau^{g^*} > \cdot\} = Y_{\cdot}\mathbb{1}\{\tau^{g^*} > \cdot\} \quad P-\text{a.s.}$
- **Exchangeability**: Define $\mathcal{H}_t := \mathcal{F}_t \vee \sigma\left(\tilde{Y}\right)$. Let $\Lambda^{a,a_j}(\mathrm{d}t) = \pi_t(\{a_j\})\Lambda^a(\mathrm{d}t)$ denote the P-$\mathcal{F}_t$-compensator of N^{a,a_j} and $\Lambda_H^{a,a_j}(\mathrm{d}t) = \pi_t^H(\{a_j\})\Lambda_H^a(\mathrm{d}t)$ denote the P-$\mathcal{H}_t$-compensator of N^{a,a_j}, and that $\frac{\pi_t^*(\{a_j\})}{\pi_t^H(\{a_j\})}$ can be chosen càglàd, and $\mathcal{H}_t$-predictable. We have for all $j \in \{1, ..., k\}$ and $m \in \mathbb{N}$ that $\pi^*_{T_{(m)}}(\{a_j\})\pi_{T_{(m)}}(\{a_j\}) = \pi^*_{T_{(m)}}(\{a_j\})\pi^H_{T_{(m)}}(\{a_j\})$ P-a.s.
- **Positivity**:

$$W(t) := \prod_{j=1}^{N_t} \left(\prod_{i=1}^{k} \left(\frac{\pi^*_{T_{(j)}}\left(\{a_i\}; \mathcal{F}_{T_{(j-1)}}\right)}{\pi_{T_{(j)}}\left(\{a_i\}; \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\{A(T_{(k)})=a_i\}} \right)^{\mathbb{1}\{\Delta_{(j)}=a\}} \tag{4}$$

  is uniformly integrable P-$\mathcal{F}_t$ martingale.

Further, let

$$K_t = \int_0^t \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) M^{a,a_j}(\mathrm{d}s)$$

Then,

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W(t)] \tag{5}$$

and $W(t) = \mathcal{E}(K)_t$ is a uniformly integrable P-$\mathcal{F}_t$-martingale, where $\mathcal{E}$ denotes the Doléans-Dade exponential (Protter (2005)).

Proof: We shall use that the likelihood ratio solves a specific stochastic differential equation. First, we have that

$$\begin{aligned}
K_t &= \int_0^t \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) M^{a,a_j}(\mathrm{d}s) \\
&= \int_0^t \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) N^{a,a_j}(\mathrm{d}s) - \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) \Lambda^{a,a_j}(\mathrm{d}s) \\
&= \int_0^t \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) N^{a,a_j}(\mathrm{d}s) - \sum_{j=1}^{k} (\pi^*_s(\{a_j\}) - \pi_s(\{a_j\}))\Lambda^a(\mathrm{d}s) \\
&= \int_0^t \sum_{j=1}^{k} \left(\frac{\pi^*_s(\{a_j\})}{\pi_s(\{a_j\})} - 1 \right) N^{a,a_j}(\mathrm{d}s).
\end{aligned}$$

This shows that $W(t) = \mathcal{E}(K)_t$. By exchangeability and the last equality, we also have that $K_t = \int_0^t \sum_{j=1}^k \left(\frac{\pi^*_s(\{a_j\})}{\pi^H_s(\{a_j\})} - 1 \right) M^{a,a_j}_{\mathcal{H}}(\mathrm{d}s)$, where $M^{a,a_j}_{\mathcal{H}} = N^{a,a_j} - \Lambda^{a,a_j}_{\mathcal{H}}$ and $\Lambda^{a,a_j}_{\mathcal{H}}$ denotes the P-$\mathcal{H}_t$ compensator of N^{a,a_j}. Thus, since we can choose $\frac{\pi^*_t(\{a_j\})}{\pi^H_t(\{a_j\})}$ to be $\mathcal{H}_t$-predictable and locally bounded, we see that generally K_t is a local P-$\mathcal{H}_t$-martingale. Since W is a uniformly integrable martingale on $[0, T]$, we have that

$$\mathbb{E}_P[W(T)] = \mathbb{E}_P[W(0)] = \mathbb{E}_P[1] = 1$$

by Theorem 10.3.2 of Last & Brandt (1995). Hence, by Theorem 15.3.2 of Cohen & Elliott (2015), we have that W is a P-$\mathcal{H}_t$ martingale.

This implies that

$$
\begin{aligned}
\mathbb{E}_P[Y_t W(t)] &= \mathbb{E}_P[Y_t \mathbb{1}\{\tau^{g^*} > t\} W(t)] \\
&\overset{(**)}{=} \mathbb{E}_P\left[\tilde{Y}_t \mathbb{1}\{\tau^{g^*} > t\} W(t)\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_t W(t)\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_t \mathbb{E}_P[W(t) \mid \mathcal{H}_0]\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_t W(0)\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_t\right]
\end{aligned}
$$

where in $(**)$ we used consistency. □

We note some things here.

- First that Equation 4 is the same likelihood ratio as in Rytgaard et al. (2022).
- Also note that the Radon-Nikodym derivative is uniquely given by the second part of Lemma 1, so that if π' and π are the corresponding Radon-Nikodym derivatives of Lemma 1, then Equation 4 obtained from either will be indistinguishable from each other.
- Also note that in the proof, it suffices that W is uniformly bounded because then it will also be a P-$\mathcal{H}_t$-martingale since it is a local, bounded P-$\mathcal{H}_t$-martingale.
- Note that a more general alternative to the above conditions is simply to assume that W is a uniformly martingale on both filtrations $\mathcal{F}_t$ and $\mathcal{H}_t$.
- Note that alternatively, we can also require strong consistency. What strong consistency dictates is that $\tilde{Y}_{\cdot \wedge \tau^{g^*}} = Y_{\cdot \wedge \tau^{g^*}}$ P-a.s. The subtle difference is that the strict inequality is replaced by a weak one. An illustration of strong consistency is presented in the figure below for a recurrent event outcome below.

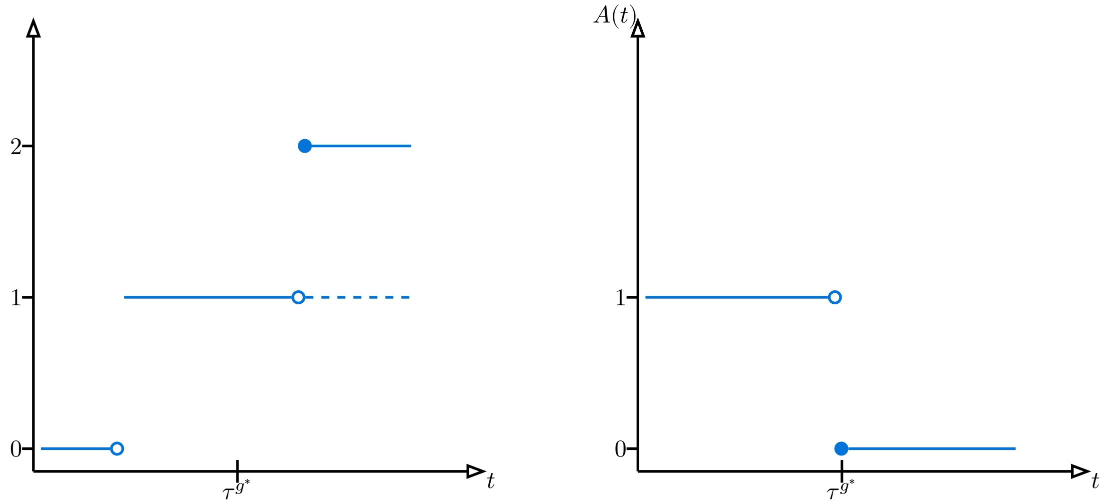

Figure 1: The figure illustrates the consistency condition for the potential outcome framework for single individual. The left panel shows the potential outcome process $\tilde{Y}(t)$ (dashed) and the observed process $Y(t)$ (solid). The right panel shows the treatment process $A(t)$. At time τ^{g^*}, the treatment is stopped and the processes may from some random future point diverge from each other. In this case, the treatment is beneficial for the user, as it would have prevented another recurrent event from happening.

In Rytgaard et al. (2022), both an exchangeability condition and a positivity condition are presented, but no proof is given that these conditions imply that their target parameter is identified. Our proposal shows that under the conditions of Theorem 1, the g-formula given in Rytgaard et al. (2022) causally identifies the counterfactual mean outcome under the assumption that the other martingales are orthogonal to the treatment martingale. Lemma 1 of Ryalen (2024) then gives the desired target parameter. Note that this is weaker than the assumptions in Rytgaard et al. (2022), as they implicitly require that *all* martingales are orthogonal due to their factorization of the likelihood. This is because $\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X+Y)$ if and only if $[X,Y]=0$. This can be seen by applying Theorem 38, p. 130 of Protter (2005) and using that the stochastic exponential solves a specific stochastic differential equation. We do not provide a more concise statement here.

Theorem 2 (g-formula): Let, further, $Q = W(T) \cdot P$ denote the probability measure defined by the likelihood ratio $W(T)$ given in Equation 4. Under positivity, then

1. The Q-$\mathcal{F}_t$ compensator of $N^a(\mathrm{d}t \times \mathrm{d}x)$ is $\pi_t^*(\mathrm{d}x)\Lambda^a(\mathrm{d}t)$.
2. The Q-$\mathcal{F}_t$ compensator of N^x is Λ^x for $x \in \{y, \ell\}$.

Proof: First note that for a local $\mathcal{F}_t$-martingale X in P, we have

$$\int_0^t \frac{1}{W_{s-}} \mathrm{d}\langle W, X\rangle_s^P = \langle K, X\rangle_t^P \tag{6}$$

since we have that $W_t = 1 + \int_0^t W_{s-} \, \mathrm{d}K_s$; whence

$$\langle W, X\rangle_t = \langle 1, X\rangle_t + \langle W_- \bullet K, X\rangle_t = W_{t-} \bullet \langle K, X\rangle_t$$

With $X = M^{a,x}$, we find

$$\begin{aligned}
\langle K, M^{a,x}\rangle_t^P &= \int_0^t \sum_j \left(\frac{\pi_s^*(a_j)}{\pi_s(a_j)} - 1\right) \mathrm{d}\langle M^{a,a_j}, M^{a,x}\rangle_s^P \\
&= \int_0^t \left(\frac{\pi_s^*(x)}{\pi_s(x)} - 1\right) \mathrm{d}\langle M^{a,x}\rangle_s^P + \sum_{j \neq x} \int_0^t \left(\frac{\pi_s^*(a_j)}{\pi_s(a_j)} - 1\right) \mathrm{d}\langle M^{a,a_j}, M^{a,x}\rangle_s^P \\
&= \int_0^t \left(\frac{\pi_s^*(x)}{\pi_s(x)} - 1\right) \pi_s(x)\Lambda^a(\mathrm{d}s) - \int_0^t \left(\frac{\pi_s^*(x)}{\pi_s(x)} - 1\right) \Delta(\pi(x)\Lambda^a)_s \pi_s(x)\Lambda^a(\mathrm{d}s) \\
&\quad - \sum_{j \neq x} \int_0^t \left(\frac{\pi_s^*(a_j)}{\pi_s(a_j)} - 1\right) \Delta(\pi(x)\Lambda^a)_s \pi_s(a_j)\Lambda^a(\mathrm{d}s) \\
&= \int_0^t (\pi_s^*(x) - \pi_s(x))\Lambda^a(\mathrm{d}s) - \sum_j \int_0^t (\pi_s^*(a_j) - \pi_s(a_j))\Delta(\pi(x)\Lambda^a)_s \Lambda^a(\mathrm{d}s) \\
&= \int_0^t (\pi_s^*(x) - \pi_s(x))\Lambda^a(\mathrm{d}s).
\end{aligned} \tag{7}$$

Girsanov's theorem (Theorem 41, p. 136 of Protter (2005)) together with Equation 6 and Equation 7 gives that

$$N^a(\mathrm{d}t \times \mathrm{d}x) - \pi_t(\mathrm{d}x)\Lambda^a(\mathrm{d}t) - (\pi_t^*(\mathrm{d}x) - \pi_t(\mathrm{d}x))\Lambda^a(\mathrm{d}t) = N^a(\mathrm{d}t \times \mathrm{d}x) - \pi_t^*(\mathrm{d}x)\Lambda^a(\mathrm{d}t)$$

is a Q-$\mathcal{F}_t$-local martingale. The second statement follows by noting that

$$[M^y, K]_t = \int_0^t \Delta N_t^y \sum_j \left(\frac{\pi_s^*(a_j)}{\pi_s(a_j)} - 1\right) N^{a,a_j}(\mathrm{d}s) - \int_0^t \Delta\Lambda^y(s) \sum_j \left(\frac{\pi_s^*(a_j)}{\pi_s(a_j)} - 1\right) M^{a,a_j}(\mathrm{d}s)$$

where we apply the trick with adding and subtracting the treatment compensators in the second term. The first term is zero because no two counting processes jump at the same time. The second term is a local martingale. This implies $\langle M^y, K\rangle_t^P = 0$. For $x = \ell$ the argument is the same. □

4 Comparison with Ryalen (2024)

It is natural to ask oneself: how does our conditions relate to the ones of Ryalen (2024)? The condition of consistency is the same. However, the exchangeability condition and the positivity condition are different in general. The overall point of the present subsection is to argue that the identification formula in Ryalen (2024) cannot generally identify the causal estimand, but that there is, in general, a different causal interpretation behind that identification formula. Moreover, we argue that under structural, albeit unrestrictive assumptions, the two causal estimands are the same.

First, let $\mathbb{N}_t^a = \mathbb{1}\{\tau^{g^*} \leq t\}$ and let $\mathbb{L}_t$ denote its P-$\mathcal{F}_t$-compensator. Then, exchangeability in Ryalen (2024) is that

- **Exchangeability**: Define $\mathcal{H}_t := \mathcal{F}_t \vee \sigma\left(\tilde{Y}\right)$. The P-$\mathcal{F}_t$ compensator for $\mathbb{N}^a$ is also the P-$\mathcal{H}_t$ compensator.

Let $\widetilde{W}(t) := \frac{\mathcal{E}(-\mathbb{N}^a)_t}{\mathcal{E}(-\mathbb{L}^a)_t} = \mathcal{E}(\mathbb{K}^a)_t$ and $\tilde{Q} := \widetilde{W}(T) \cdot P$, where $\mathbb{K}^a_t = -\int_0^t \frac{1}{1-\Delta \mathbb{L}^a_s}(\mathbb{N}^a(\mathrm{d}s) - \mathbb{L}^a(\mathrm{d}s))$. If additionally positivity as described in Ryalen (2024) holds, then

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P\left[\widetilde{W}(t) Y_t\right]. \tag{8}$$

4.1 A different intervention yielding the same identification formula as Ryalen (2024)

Suppose that we represent N as the multivariate counting process $N = \left(N^y, N^d, N^{\ell,l_1}, ..., N^{\ell,l_{d_l}}, N^{a,a_1}, N^{a,a_0}\right)$, where, for simplicity, we take $\mathcal{A} = \{a_0, a_1\}$. We are interested in the effect of staying on treatment a_1, which prevents N^{a,a_0}-events. From a philosophical point of view there is a difference in actually constitutes complete data can be different. The intervention we have discussed earlier would include *all* treatment visitation times as complete data. This intervention, however, would only include the treatment visitation times where treatment a_1 is assigned to the patient.

This is a predictable intervention; however, the intervention that is formulated previously is an optional intervention, in general. With canonical compensators, we see that for the potential outcome process for this intervention

$$\begin{aligned}
\tilde{\Lambda}^x(\mathrm{d}t) &= \dot{\Lambda}^x\left(\widetilde{N}'\right)(\mathrm{d}t), x \neq a, a_0, a, a_1 \\
\tilde{\Lambda}^{a,a_1}(\mathrm{d}t) &= \pi_t\left(\widetilde{N}', \{a_1\}\right) \dot{\Lambda}^a\left(\widetilde{N}'\right)(\mathrm{d}t) \\
\tilde{\Lambda}^{a,a_0}(\mathrm{d}t) &= 0
\end{aligned}$$

However, because $\pi_t\left(\widetilde{N}', \{a_1\}\right) \neq 1$, it is not guaranteed that $\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P\left[\tilde{Y}'_t\right]$. Ryalen (2024) gives conditions under which there exists such a potential outcome process; for instance if $\langle M^{a_0}, M^x \rangle^P_s = 0$ for $x \neq a, a_0$ in this situation. On the other hand, note that $\tau^{g^*} = \inf\{t > 0 \mid N^{a,a_1}(t) \neq N^{a,a_1}(t) + N^{a,a_0}(t) \vee N^{a,a_0}(t) \neq 0\} = \inf\{t > 0 \mid N^{a,a_1}(t) \neq 0\}$. Thus, the identification formulae for $\mathbb{E}_P\left[\tilde{Y}_t\right]$ and $\mathbb{E}_P\left[\tilde{Y}'_t\right]$ are actually the same for two *different* interventions; and the exchangeability conditions are the same – we replace $\tilde{Y}$ with $\tilde{Y}'$ here. We argue that there are situations for observed in which the exchangeability condition for $\tilde{Y}$ can fail, but as we have argued this cannot occur for $\tilde{Y}'$ under the assumption of orthogonal martingales.

4.2 Example with no time-varying confounding and one treatment event showing that the identification formulas may be different

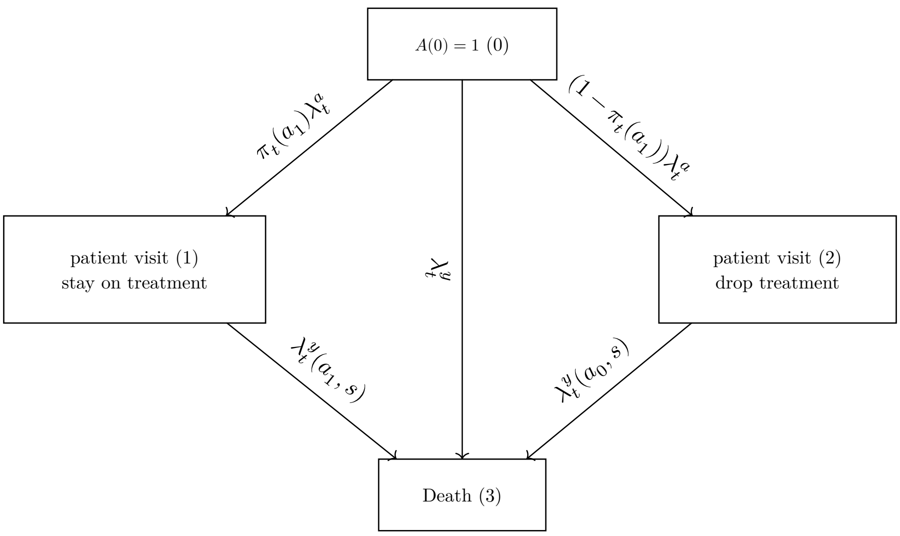

Figure 2: A multi-state model allowing one visitation time for the treatment with the possible treatment values 0/1.

Consider a simple example where $N^a(t) \leq 1$ for all t, we observe $N = (N^y, N^{a,a_0}, N^{a,a_1})$, and that people are assigned treatment at $t = 0$ as in a randomized trial. We consider the intervention $\pi^*_t(a_1) = 1$ for all t. Suppose that $(N^y, N^{a,a_0}, N^{a,a_1})$ has compensator

$$\Lambda^y(\mathrm{d}t)(P) = \mathbb{1}\left\{t \leq T_{(1)}\right\}\lambda^y_t \,\mathrm{d}t + \mathbb{1}\left\{T_{(1)} < t \leq T_{(2)}\right\}\lambda^y_t\left(A\left(T_{(1)}\right), T_{(1)}\right) \mathrm{d}t,$$

$$\Lambda^{a,a_j}(\mathrm{d}t)(P) = \mathbb{1}\left\{t \leq T_{(1)}\right\}\pi_t(a_j)\lambda^a_t \,\mathrm{d}t, j = 0, 1.$$

with respect to its natural filtration. Is it then possible to construct a potential outcome fulfilling consistency and exchangeability in the sense of Ryalen (2024) such that $\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P\left[\widetilde{W}(t)Y_t\right]$ and when is it possible?

First note that in in Q (defined in Theorem 2), we have

$$\begin{aligned}\Lambda^y(\mathrm{d}t)(Q) &= \Lambda^y(\mathrm{d}t)(P),\\ \Lambda^{a,a_0}(\mathrm{d}t)(Q) &= 0 \,\mathrm{d}t,\\ \Lambda^{a,a_1}(\mathrm{d}t)(Q) &= \Lambda^a(\mathrm{d}t)(P)\end{aligned}$$

However, in $\tilde{Q}$ (the change of measure in Ryalen (2024)), we have

$$\begin{aligned}\Lambda^y(\mathrm{d}t)\left(\tilde{Q}\right) &= \Lambda^y(\mathrm{d}t)(P)\\ \Lambda^{a,a_0}(\mathrm{d}t)\left(\tilde{Q}\right) &= 0 \,\mathrm{d}t,\\ \Lambda^{a,a_1}(\mathrm{d}t)\left(\tilde{Q}\right) &= \pi_t(\{a_1\}; P)\Lambda^a(\mathrm{d}t)(P)\end{aligned}$$

since $\tau^{g^*} = \infty$ almost surely in $\tilde{Q}$.

First, we identify when one can do this in this example.

4.2.1 Example showing the identification formulae are the same under a local independence condition

Suppose that $\lambda^y_t(1, s) = \lambda^y_t$ for $s < t$ (in P). Then, we show $\mathbb{E}_Q[Y_t] = \mathbb{E}_{\widetilde{Q}}[Y_t]$ by showing that $\mathbb{E}_{\widetilde{Q}}[Y_t]$ does not depend on the Radon-Nikodym derivative $\pi_t(\{a_1\})$. Thus,

$$
\begin{aligned}
\mathbb{E}_{\widetilde{Q}}[Y_t] &= \mathbb{E}_{\widetilde{Q}}[N^y(t)] \\
&= \mathbb{E}_{\widetilde{Q}}\Big[\mathbb{1}\Big\{T_{(1)} \leq t, \Delta_{(1)} = y\Big\} + \mathbb{1}\Big\{\Delta_{(1)} = a, A\Big(T_{(1)}\Big) = 1, T_{(2)} \leq t, \Delta_{(2)} = y\Big\}\Big] \\
&= \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\lambda_s^y\,\mathrm{d}s \\
&\qquad + \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a \int_s^t \exp\left(-\int_s^v \lambda_u^y(1,s)\,\mathrm{d}u\right)\lambda_v^y(1,s)\,\mathrm{d}v\,\mathrm{d}s \\
&= \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a\,\mathrm{d}u)\right)\lambda_s^y\,\mathrm{d}s \\
&\qquad + \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a\left(1 - \exp\left(-\int_s^t \lambda_u^y(1,s)\,\mathrm{d}u\right)\right)\mathrm{d}s \\
&= 1 - \exp\left(-\int_0^t (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right) - \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a\,\mathrm{d}s \\
&\qquad + \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a\left(1 - \exp\left(-\int_s^t \lambda_u^y(1,s)\,\mathrm{d}u\right)\right)\mathrm{d}s \\
&= 1 - \exp\left(-\int_0^t (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right) \\
&\qquad - \int_0^t \exp\left(-\int_0^s (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a \exp\left(-\int_s^t \lambda_u^y(1,s)\,\mathrm{d}u\right)\mathrm{d}s \\
&\overset{(!)}{=} 1 - \exp\left(-\int_0^t (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right) - \exp\left(-\int_0^t \lambda_u^y\,\mathrm{d}u\right)\int_0^t \exp\left(-\int_0^s (\pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right)\pi_s(a_1)\lambda_s^a\,\mathrm{d}s \\
&= 1 - \exp\left(-\int_0^t (\lambda_u^y + \pi_u(a_1)\lambda_u^a)\,\mathrm{d}u\right) - \exp\left(-\int_0^t \lambda_u^y\,\mathrm{d}u\right)\left(1 - \exp\left(-\int_0^t \lambda_u^a\pi_u(a_1)\,\mathrm{d}u\right)\right) \\
&= 1 - \exp\left(-\int_0^t \lambda_u^y\,\mathrm{d}u\right)
\end{aligned} \tag{9}
$$

In (!), we apply the stated local independence condition.

4.2.2 Example showing the identification formulae are different

In the observed data, take

$$
\begin{aligned}
\pi_t(\{a_1\}) &= \pi \in (0,1), \\
\lambda_t^y &= \lambda^y \\
\lambda_t^a &= \lambda^a \\
\lambda_t^y(1,s) &= \lambda^{y,2} \neq \lambda^y
\end{aligned}
$$

and further pick them so that $\lambda^a + \lambda^y - \lambda^{y,2} \neq 0$. Then, we have from the fourth last equality of Equation 9,

$$
\begin{aligned}
\mathbb{E}_Q[Y_t] &= 1 - \exp(-(\lambda^y + \lambda^a)t) - \int_0^t \exp(-(\lambda^y + \lambda^a)s)\lambda^a \exp(-\lambda^{y,2}(t-s))\,\mathrm{d}s \\
&= 1 - \exp(-(\lambda^y + \lambda^a)t) - \lambda^a \frac{\exp(-\lambda^{y,2}t)}{\lambda^a + \lambda^y - \lambda^{y,2}}(1 - \exp(-(\lambda^a + \lambda^y - \lambda^{y,2})t))
\end{aligned}
$$

Replacing λ^a with $\pi\lambda^a$ yields $\mathbb{E}_{\widetilde{Q}}[Y_t]$, i.e.,

$$
\mathbb{E}_{\widetilde{Q}}[Y_t] = 1 - \exp(-(\lambda^y + \pi\lambda^a)t) - \pi\lambda^a \frac{\exp(-\lambda^{y,2}t)}{\pi\lambda^a + \lambda^y - \lambda^{y,2}}(1 - \exp(-(\pi\lambda^a + \lambda^y - \lambda^{y,2})t))
$$

Taking

$$
\begin{aligned}
\pi = \lambda^a &= 0.5 \\
\lambda^y &= 0.2 \\
\lambda^{y,2} &= 0.8
\end{aligned}
$$

with $t = 3$, we have

$$\mathbb{E}_Q[Y_t] = 1 - \exp(-(0.2+0.5)3) - 0.5\frac{\exp(-(0.8)3)}{0.5+0.2-0.8}(1-\exp(-(0.5+0.2-0.8)3)) \approx 0.72$$

while

$$\mathbb{E}_{\tilde{Q}}[Y_t] = 1 - \exp(-(0.2+0.5\times 0.5)3) - 0.5\times 0.5\frac{\exp(-(0.8)3)}{0.5\times 0.5+0.2-0.8}(1-\exp(-(0.5\times 0.5+0.2-0.8)3)) \approx 0.62$$

However by the argument in Section 6, we can construct the potential outcome process $\tilde{Y}$ fulfilling consistency and exchangeability as in Theorem 1, so we actually have that $\mathbb{E}_Q[Y_t] = \mathbb{E}_P\left[\tilde{Y}_t\right]$, but $\mathbb{E}_{\tilde{Q}}[Y_t] \neq \mathbb{E}_P\left[\tilde{Y}_t\right]$.

4.3 More general argument (sketch)

Theorem 3: Suppose that $(N^{a,a_1}, ..., N^{a,a_j}) \not\to N^{-a}(\cdot \wedge \tau^{g^*})$ (local independence understood in terms of the compensator). Here, N^{-a} denotes is every other counting process that is not a treatment counting process.

Then,

$$\mathbb{E}_{\tilde{Q}}[Y_t] = \mathbb{E}_Q[Y_t]$$

Proof: One can get the canonical compensators from the two measures Q and $\tilde{Q}$ from P via Theorem 2 and Lemma ? in Ryalen (2024). Here, we see that Q and $\tilde{Q}$ agree on the canonical compensator for every component that is not a treatment component. We can now induce measures Q' and $\tilde{Q}'$ restricting the measures to every other component than the treatment component (can be done explicitly) and consider its smaller generated natural filtration of the new marked point process. Under a local independence statement, the compensators under smaller filtration and the original filtration are the same. However, this means that the canonical compensators in Q' and $\tilde{Q}'$ will be the same and hence the distribution of $(Y_t)_{t\in[0,T]}$ will be the same in the two induced measures and hence in the original probability measures. □

Another example where they are the same is when the total treatment process is $\mathcal{F}_t$-predictable: Now, we calculate

$$
\begin{aligned}
\mathbb{K}_t^a &= -\int_0^t \frac{1}{1-\Delta\mathbb{L}_s^a}\,\mathrm{d}(\mathbb{N}_s^a - \mathbb{L}_s^a) \\
&= -\int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} (1-\pi_s^*(x)) \frac{1}{1-\sum_{y\in\mathcal{A}}(1-\pi_s^*(y))\pi_s(y)} \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&= -\int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} (1-\pi_s^*(x)) \frac{1}{\sum_{y\in\mathcal{A}}\pi_s^*(y)\pi_s(y)} \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&= \int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} \frac{\pi_s^*(x)}{\pi_s(x)} \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&\quad - \int_0^{t\wedge\tau^{g^*}} \frac{1}{\sum_{y\in\mathcal{A}}\pi_s^*(y)\pi_s(y)} \left(\sum_{x\in\mathcal{A}} N^x(\mathrm{d}s) - \sum_{x\in\mathcal{A}} \pi_s(x)\bar{N}^a(\mathrm{d}s)\right) \\
&= \int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} \frac{\pi_s^*(x)}{\pi_s(x)} \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&\quad - \int_0^{t\wedge\tau^{g^*}} \frac{1}{\sum_{y\in\mathcal{A}}\pi_s^*(y)\pi_s(y)} \big(\bar{N}^a(\mathrm{d}s) - \bar{N}^a(\mathrm{d}s)\big) \\
&= \int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} \frac{\pi_s^*(x)}{\pi_s(x)} \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&= \int_0^{t\wedge\tau^{g^*}} \sum_{x\in\mathcal{A}} \left(\frac{\pi_s^*(x)}{\pi_s(x)} - 1\right) \big(N^x(\mathrm{d}s) - \pi_s(x)\bar{N}^a(\mathrm{d}s)\big) \\
&= K_t^*
\end{aligned}
$$

5 Variations of the sequential exchangeability criteria yielding the same identification formula

We now provide a sequential representation of the exchangeability condition. It aligns closely with the postulated exchangeability condition in Rytgaard et al. (2022); however, notably on the conditioning set, we include the k'th event time, which is not included in Rytgaard et al. (2022). The sequential condition is a bit unusual in the sense that we do not condition on the strict history. We also provide an equivalent statement for the martingale condition.

Theorem 4: Suppose consistency and positivity holds as in Theorem 1. Then, we have

$$\mathbb{E}_P\big[\tilde{Y}_t\big] = \mathbb{E}_P[W_t Y_t],$$

for all $t \in [0, T]$, if for $k \in \mathbb{N}$ and $t \in [0, T]$ it holds that

$$\tilde{Y}_t \perp \mathbb{1}\Big\{A\big(T_{(k)}\big) = g^*\Big(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\Big)\Big\} \mid \mathcal{F}^{g^*}_{T_{(k-1)}}, T_{(k)} \le t, \Delta_{(k)} = a,$$

where

$$\mathcal{F}^{g^*}_{T_{(k)}} = \sigma\Big(L\big(T_{(k)}\big), \Delta_{(k)}, \mathbb{1}\Big\{A\big(T_{(k)}\big) = g_k^*\Big(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\Big)\Big\}, ..., \mathbb{1}\{A(0) = g_0^*(L(0))\}, L(0)\Big)$$

Proof: We see immediately that,

$$
\begin{aligned}
&\int W_{s-} \mathbb{1}\left\{T_{(m)} < s < T_{(m+1)} \wedge t\right\} \mathrm{d}K_s \\
&= W_{T_{(m)}} \int \mathbb{1}\left\{T_{(m)} < s < T_{(m+1)} \wedge t\right\} \mathrm{d}K_s \\
&= W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} \left(\sum_j \frac{\pi^*_{T_{(m)}}\left(\left\{a_j\right\}\right)}{\pi_{T_{(m)}}\left(\left\{a_j\right\}\right)} - 1 \right) N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \\
&= W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} \left(\frac{\mathbb{1}\left\{A\left(T_{(m)}\right) = g^*_m\left(\mathcal{F}_{T_{(m-1)}}, T_{(m)}\right)\right\}}{\pi_{T_{(m)}}\left(\left\{g^*_m\left(\mathcal{F}_{T_{(m-1)}}, T_{(m)}\right)\right\}\right)} - 1 \right) N^{a,a_j}\left(T_{(m+1)} \wedge t\right)
\end{aligned}
$$

By consistency and positivity, the desired result is equivalent to

$$
\sum_{m=0}^{\infty} \mathbb{E}_P\left[\int W_{s-} \mathbb{1}\left\{T_{(m)} < s < T_{(m+1)} \wedge t\right\} \mathrm{d}K_s \tilde{Y}_t\right] = 0
$$

by Lemma 4 of Ryalen (2024), so

$$
\begin{aligned}
&\sum_{m=0}^{\infty} \mathbb{E}_P\left[\int W_{s-} \mathbb{1}\left\{T_{(m)} < s < T_{(m+1)} \wedge t\right\} \mathrm{d}K_s \tilde{Y}_t\right] \\
&= \sum_{m=0}^{\infty} \mathbb{E}_P\left[W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} \left(\frac{\mathbb{1}\left\{A\left(T_{(m)}\right) = g^*_m\left(\mathcal{F}_{T_{(m)}}, T_{(m+1)}\right)\right\}}{\pi_{T_{(m)}}\left(\left\{g^*_m\left(\mathcal{F}_{T_{(m)}}, T_{(m+1)}\right)\right\}\right)} - 1 \right) N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \tilde{Y}_t \right] \\
&= \sum_{m=0}^{\infty} \mathbb{E}_P\Bigg[W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \\
&\quad \times \mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(m)}\right) = g^*_m\left(\mathcal{F}_{T_{(m)}}, T_{(m+1)}\right)\right\}}{\pi_{T_{(m)}}\left(\left\{g^*_m\left(\mathcal{F}_{T_{(m-1)}}, T_{(m)}\right)\right\}\right)} - 1 \right) \tilde{Y}_t \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq t, \Delta_{(k)} = a \right] \Bigg] \\
&= \sum_{m=0}^{\infty} \mathbb{E}_P\Bigg[W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \\
&\quad \times \mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(m)}\right) = g^*_m\left(\mathcal{F}_{T_{(m)}}, T_{(m+1)}\right)\right\}}{\pi_{T_{(m)}}\left(\left\{g^*_m\left(\mathcal{F}_{T_{(m-1)}}, T_{(m)}\right)\right\}\right)} - 1 \right) \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq, \Delta_{(k)} = a \right] \\
&\quad \times \mathbb{E}_P\left[\tilde{Y}_t \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq t, \Delta_{(k)} = a\right] \Bigg] \\
&= \sum_{m=0}^{\infty} \mathbb{E}_P\Bigg[W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \\
&\quad \times \mathbb{E}_P\Bigg[\mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(m)}\right) = g^*_m\left(\mathcal{F}_{T_{(m)}}, T_{(m+1)}\right)\right\}}{\pi_{T_{(m)}}\left(\left\{g^*_m\left(\mathcal{F}_{T_{(m-1)}}, T_{(m)}\right)\right\}\right)} - 1 \right) \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)}, \Delta_{(m+1)} = a \right] \\
&\quad \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq t, \Delta_{(k)} = a \Bigg] \\
&\quad \times \mathbb{E}_P\left[\tilde{Y}_t \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq t, \Delta_{(k)} = a\right] \Bigg] \\
&= \sum_{m=0}^{\infty} \mathbb{E}_P\left[W_{T_{(m)}} \mathbb{1}\left\{T_{(m)} < t\right\} N^{a,a_j}\left(T_{(m+1)} \wedge t\right) \times (1-1) \mathbb{E}_P\left[\tilde{Y}_t \mid \mathcal{F}_{T_{(m)}}, T_{(m+1)} \leq t, \Delta_{(k)} = a\right]\right] \\
&= 0.
\end{aligned}
$$

□

Now we compare a sequential criterion to the full exchangebility statement.

Lemma 2: Note that

$$\left(\tilde{Y}_t\right)_{t\in[0,T]} \perp \mathbb{1}\left\{A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)\right\} \mid \mathcal{F}^{g^*}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a,$$

for all $k \in \mathbb{N}$ if and only if $W(t)$ is $\mathcal{F}_t$-measurable.

Proof: "If": Follows from the previous result.

Conversely, suppose that $W(t)$ is $\mathcal{F}_t$-measurable. Then, we have

$$\frac{W\left(T_{(k)}\right)}{W\left(T_{(k-1)}\right)}\mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty\right\}$$

$$= \left(\frac{\mathbb{1}\left\{A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)\right\}}{P\left(A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \left(\tilde{Y}_t\right)_{t\in[0,T]}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k+1)}=a\right\}} \mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty\right\}$$

is $\mathcal{F}_{T_{(k)}}$-measurable. Hence

$$\begin{aligned}
&\mathbb{1}\left\{A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right), \Delta_{(k)} = a\right\} \\
&\quad\times \frac{1}{P\left(A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \left(\tilde{Y}_t\right)_{t\in[0,T]}\right)}\mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty\right\} \\
&= \mathbb{1}\left\{A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right), \Delta_{(k)} = a\right\}\frac{W\left(T_{(k)}\right)}{W\left(T_{(k-1)}\right)}\mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty\right\} \\
&\overset{(*)}{=} f_k\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right), ..., A(0), L(0)\right) \\
&= \mathbb{1}\left\{A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right), \Delta_{(k)} = a\right\} f_k\left(T_{(k)}, a, g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right), L\left(T_{(k)}\right), ..., A(0), L(0)\right)
\end{aligned}$$

In $(*)$, we use that that the previous line is $\mathcal{F}_{T_{(k)}}$-measurable. Take the conditional expectation on both sides given $\mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \left(\tilde{Y}_t\right)_{t\in[0,T]}$ to conclude that

$$\begin{aligned}
&P\left(A\left(T_{(k)}\right) = g^*\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \left(\tilde{Y}_t\right)_{t\in[0,T]}\right)\mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty, \Delta_{(k)} = a\right\} \\
&= \frac{1}{f_k\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right), ..., A(0), L(0)\right)}\mathbb{1}\left\{\tau^{g^*} > T_{(k-1)}, T_{(k-1)} < \infty, \Delta_{(k)} = a\right\}
\end{aligned}$$

is $\sigma\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)$-measurable whenever the probability is non-zero. This suffices for the sequential condition. □

6 On the existence of counterfactual processes fulfilling consistency and exchangeability

It is natural to ask oneself whether there exist counterfactual processes for any given law of N such that consistency and exchangeability holds. This question was already posed by Gill & Robins (2001) in the discrete time setting. If this does not hold, then, certainly, we would implicitly be imposing restrictions on the observed data law of N. As the theorem below shows, we can assume that such counterfactual processes do in fact exist, as they cannot be ruled out by the distribution of observed data.

Theorem 5: For any law of N, we can construct a probability space, wherein a counterfactual process $\widetilde{N}$ and N exists such that (strong) consistency and exchangeability (in the sense of Theorem 4) holds. Here (strong) consistency means that

$$\widetilde{N}_t \mathbb{1}\{\tau^{g^*} \geq t\} = N_t \mathbb{1}\{\tau^{g^*} \geq t\} \quad P-\text{a.s.}$$

which implies that

$$\int_0^t \mathbb{1}\{s \leq \tau^{g^*}\} \, \mathrm{d}\widetilde{N}_t = \int_0^t \mathbb{1}\{s \leq \tau^{g^*}\} \, \mathrm{d}N_t,$$

or

$$\widetilde{N}_{\cdot \wedge \tau^{g^*}} = N_{\cdot \wedge \tau^{g^*}}$$

Proof: We provide an argument somewhat analogous to the one given in Section 6 of Gill & Robins (2001).

Construction of counterfactual process

First, we let $\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right)_{k \in \mathbb{N}}$ denote the marked point process corresponding to N and $(A(0), L(0))$ be the initial values of the treatment and covariate process. First, we generate $L(0)$ from its marginal distribution. Next, for $a_0 \in \mathcal{A}$, we generate $\left(L^{a_0}\left(T_{(1)}^{a_0}\right), T_{(1)}^{a_0}, \Delta_{(1)}^{a_0}\right) \sim L\left(T_{(1)}\right), T_{(1)}, \Delta_{(1)} \mid A(0) = a_j, L(0)$ (for each value of a_0, these can be generated independently). Next, for each combination of $a_0 \in \mathcal{A}$ and $a_1 \in \mathcal{A}$ where $\Delta_{(1)}^{a_0} = a$, we generate

$$\begin{aligned} &\left(L^{a_0,a_1}\left(T_{(2)}^{a_0,a_1}\right), T_{(2)}^{a_0,a_1}, \Delta_{(2)}^{a_0,a_1}\right) \\ &\sim L\left(T_{(2)}\right), T_{(2)}, \Delta_{(2)} \mid L\left(T_{(1)}\right) = l_1, A\left(T_{(1)}\right) = a_1, T_{(1)} = t_1, \Delta_{(1)} = s_1, A(0) = a_0, L(0) \end{aligned}$$

where $(l_1, t_1, s_1) = \left(L^{a_0}\left(T_{(1)}^{a_0}\right), T_{(1)}^{a_0}, \Delta_{(1)}^{a_0}\right)$ for $\Delta_{(1)}^{a_0} \neq y$ and $T_{(1)}^{a_0} < T$. If $\Delta_{(1)}^{a_0} = y$, put $\left(L^{a_0,a_1}\left(T_{(2)}^{a_0,a_1}\right), T_{(2)}^{a_0,a_1}, \Delta_{(2)}^{a_0,a_1}\right) = (\emptyset, \infty, \emptyset)$. Continue in this manner to define $\left(L^{a_0,\dots,a_k}\left(T_{(k)}^{a_0,\dots,a_1}\right), T_{(k)}^{a_0,\dots,a_k}, \Delta_{(k)}^{a_0,\dots,a_k}\right)$ for all $k \in \mathbb{N}$.

Then, we define $\widetilde{N}$ by the marked point process. Let as a shorthand $g_0^*(L(0)) = g_0^+$, $g_1^* = g_1^*\left(L(0), g_0^*, T_{(1)}^{g_0^*}\right)$ and so on. Let

$$\left(\tilde{L}\left(\tilde{T}_1\right), \tilde{A}\left(\tilde{T}_1\right), \tilde{T}_1, \tilde{\Delta}_1\right) = \left(L^{g_0^*}\left(T_{(1)}^{g_0^*}\right), g_1^* \mathbb{1}\left\{\Delta^{(g_0^*=a)}\right\} + g_0^* \mathbb{1}\left\{\Delta^{g_0^*} \neq a\right\}, T_{(1)}^{g_0^*}, \Delta^{g_0^*}\right)$$

and continue in this manner for all $k \in \mathbb{N}$.

Construction of factual process

Next, we construct the observed data process N. We can generate the A's independently from all other considered random variables. Generate $A(0)$ from its conditional distribution given $L(0)$. Then, let

$$\left(L\left(T_{(1)}\right), T_{(1)}, \Delta_{(1)}\right) = \left(L^{A(0)}\left(T_{(1)}^{A(0)}\right), T_{(1)}^{A(0)}, \Delta_{(1)}^{A(0)}\right).$$

Then, again, generate $A\left(T_{(1)}\right)$ from its conditional distribution given $\mathcal{F}_0, T_{(1)}, \Delta_{(1)} = a$. Afterwards, let

$$\left(L\left(T_{(2)}\right), T_{(2)}, \Delta_{(2)}\right) = \left(L^{A(0),A\left(T_{(1)}\right)}\left(T_{(2)}^{A(0),A\left(T_{(1)}\right)}\right), T_{(2)}^{A(0),A\left(T_{(1)}\right)}, \Delta_{(2)}^{A(0),A\left(T_{(1)}\right)}\right).$$

Continue in this manner for all $k \in \mathbb{N}$.

Consistency

Next, we show consistency. Define

$$\widetilde{N}_t^x = \sum_k \mathbb{1}\left\{T_{(k)}^{g_0^*,\dots,g_{k-1}^*} \leq t, \Delta_{(k)}^{g_0^*,\dots,g_{k-1}'} = x\right\}$$

and

$$N_t^x = \sum_k \mathbb{1}\left\{T_{(k)} \leq t, \Delta_{(k)} = x\right\}$$

Now note that

$$
\begin{aligned}
N_t^x \mathbb{1}\{\tau^{g^*} \geq t\} &= \sum_k \mathbb{1}\Big\{T_{(k)} \leq t, \Delta_{(k)} = x\Big\}\mathbb{1}\{\tau^{g^*} > t\} \\
&= \sum_k \mathbb{1}\Big\{T_{(k)} \leq t, \Delta_{(k)} = y, A\Big(T_{(j)}\Big) = g_j^*, \forall j < k\Big\} \\
&= \sum_k \mathbb{1}\Big\{T_{(k)}^{g_0^*,\dots,g_{k-1}^*} \leq t, \Delta_{(k)}^{g_0^*,\dots,g_{k-1}'} = x, A\Big(T_{(j)}\Big) = g_j^*, \forall j < k\Big\} \\
&= \sum_k \mathbb{1}\Big\{T_{(k)}^{g_0^*,\dots,g_{k-1}^*} \leq t, \Delta_{(k)}^{g_0^*,\dots,g_{k-1}'} = x\Big\}\mathbb{1}\{\tau^{g^*} > t\} \\
&= \widetilde{N}_t^x \mathbb{1}\{\tau^{g^*} > t\}.
\end{aligned}
$$

as desired.

Exchangeability

By construction, we then have

$$
\Big(\widetilde{N}_t\Big)_{t\in[0,T]} \perp A\Big(T_{(k)}\Big) \mid \mathcal{F}^g_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a
$$

which suffices for exchangeability by Theorem 4.

Distribution

It is immediate the $(N_t)_{t\in[0,T]}$ has the right distribution since the described procedure simply generates the desired distribution.

Now we check the counterfactual outcome process. By Theorem 4.1.11 (ii) of Last & Brandt (1995), we get the compensator

$$
\mu(\mathrm{d}(l,a,\delta,t)) := \sum_k \mathbb{1}\Big\{\tilde{T}_{k-1} < t \leq \tilde{T}_k\Big\} \frac{P\Big(\Big(\tilde{L}\Big(\tilde{T}_k\Big), \tilde{A}\Big(\tilde{T}_k\Big), \tilde{\Delta}_k, \tilde{T}_k\Big) \in \mathrm{d}(l,a,\delta,t) \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)}{P\Big(\tilde{T}_k \geq t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)}
$$

with respect to the filtration $\widetilde{\mathcal{F}}_t$; the natural filtration of the counterfactual process. By integrating over the respective sets, we find that for every non-treatment component that the compensator is the same as in the observed data. We find that

$$
\begin{aligned}
\mu\big(\mathcal{L} \times \{a_j\} \times \{a\} \times (0,t]\big) &= \int_{(0,t]} \sum_k \mathbb{1}\Big\{\tilde{T}_{k-1} < t \leq \tilde{T}_k\Big\} \frac{P\Big(\tilde{A}\Big(\tilde{T}_k\Big) = a_j, \tilde{\Delta}_k = a, \tilde{T}_k \in \mathrm{d}t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)}{P\Big(\tilde{T}_k \geq t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)} \\
&= \int_{(0,t]} \sum_k \mathbb{1}\Big\{\tilde{T}_{k-1} < t \leq \tilde{T}_k\Big\} \\
&\qquad \times P\Big(\tilde{A}\Big(\tilde{T}_k\Big) = a_j \mid \tilde{\Delta}_k = a, \tilde{T}_k = t, \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big) \frac{P\Big(\tilde{\Delta}_k = a, \tilde{T}_k \in \mathrm{d}t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)}{P\Big(\tilde{T}_k \geq t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)} \\
&= \int_{(0,t]} \sum_k \mathbb{1}\Big\{\tilde{T}_{k-1} < t \leq \tilde{T}_k\Big\}\mathbb{1}\big\{a_j = g_j^*\big\} \frac{P\Big(\tilde{\Delta}_k = a, \tilde{T}_k \in \mathrm{d}t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)}{P\Big(\tilde{T}_k \geq t \mid \widetilde{\mathcal{F}}_{\tilde{T}_{k-1}}\Big)} \\
&= \int_{(0,t]} \pi_s^*\big(\{a_j\}\big)\Lambda^a(\mathrm{d}s),
\end{aligned}
$$

and the proof is complete. □

6.1 Full exchangeability and standard exchangeability

Consider a basic example where we observe N^y (primary event), N^d (competing event), N^a, and $A(t)$, but not necessarily the baseline confounder L. We can let

$$
\begin{aligned}
O &= \Big(T_{(1)}, \Delta_{(1)}, A\Big(T_{(1)}\Big), T_{(2)}, \Delta_{(2)}\Big) \\
\tilde{O} &= \Big(T_{(1)}, \Delta_{(1)}, 1, \tilde{T}_2, \tilde{\Delta}_2\Big)
\end{aligned}
$$

Then $\widetilde{N}_t^x = \mathbb{1}\Big\{T_{(1)} \leq t, \Delta_{(1)} = x\Big\} + \mathbb{1}\Big\{\tilde{T}_2 \leq t, \tilde{\Delta}_2 = x\Big\}$. Then, we can have

$$
P\left(A\Big(T_{(1)}\Big) = 0 | T_{(1)}, \Delta_{(1)} = a, \Big(\widetilde{N}_t^y\Big)_{t\in[0,T]}\right) = P\Big(A\Big(T_{(1)}\Big) = 0 | T_{(1)}, \Delta_{(1)} = a\Big),
$$

but

$$P(A\big(T_{(1)}\big) = 0|T_{(1)}, \Delta_{(1)} = a, \big(\widetilde{N}^y_t, \widetilde{N}^d_t\big)_{t\in[0,T]} \neq P\big(A\big(T_{(1)}\big) = 0|T_{(1)}, \Delta_{(1)} = a\big).$$

We provide an example showing this when the event times have densities. First note

$$\begin{aligned}
&P\Big(A\big(T_{(1)}\big) = 0|T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}_2 = t_2, \tilde{\Delta}_2 = x, L = l\Big)\\
&= \frac{p_{(T_{(1)},\Delta_{(1)},\tilde{T}_2,\tilde{\Delta}_2) \mid A(T_{(1)}),L}(t, a, t_2, x \mid 0, l) P\big(A\big(T_{(1)}\big) = 0 \mid L = l\big)}{p_{T_{(1)},\Delta_{(1)},\tilde{T}_2,\tilde{\Delta}_2 \mid L}(t, a, t_2, x \mid l)}\\
&= \frac{p_{(\tilde{T}_2,\tilde{\Delta}_2) \mid T_{(1)},\Delta_{(1)},A(T_{(1)}),L}(t_2, x \mid t, a, 0, l) p_{T_{(1)},\Delta_{(1)} \mid A(T_{(1)}),L}(t, a \mid 0, l) P\big(A\big(T_{(1)}\big) = 0 \mid L = l\big)}{p_{\tilde{T}_2,\tilde{\Delta}_2 \mid T_{(1)},\Delta_{(1)},L}(t_2, x \mid t, a, l) p_{T_{(1)},\Delta_{(1)} \mid L}(t, a \mid l)}
\end{aligned}$$

Then, we have

$$\begin{aligned}
&P\Big(A\big(T_{(1)}\big) = 0|T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}_2 = t_2, \tilde{\Delta}_2 = x\Big)\\
&= \sum_{l=0,1} P\Big(A\big(T_{(1)}\big) = 0|T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}_2 = t_2, \tilde{\Delta}_2 = x, L = l\Big)\\
&\quad P\Big(L = l \mid T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}_2 = t_2, \tilde{\Delta}_2 = x\Big)\\
&= \sum_{l=0,1} P\Big(A\big(T_{(1)}\big) = 0|T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}_2 = t_2, \tilde{\Delta}_2 = x, L = l\Big)\\
&\quad \frac{p\big(\tilde{T}_2 = t_2, \tilde{\Delta}_2 = x \mid T_{(1)} = t, \Delta_{(1)} = a, L = l\big) p\big(T_{(1)} = t, \Delta_{(1)} = a \mid L = l\big) P(L = l)}{p\big(T_{(1)} = t, \Delta_{(1)} = a\big) p\big(\tilde{T}_2 = t_2, \tilde{\Delta}_2 = x \mid T_{(1)} = t, \Delta_{(1)} = a\big)}
\end{aligned}$$

This is not a function of t_2 and x if

$$p_{\tilde{T}_2,\tilde{\Delta}_2 \mid T_{(1)},\Delta_{(1)},A(T_{(1)}),L}(t_2, x \mid t, a, 0, l) = p_{\tilde{T}_2,\tilde{\Delta}_2 \mid T_{(1)},\Delta_{(1)}}(t_2, x \mid t, a)$$

but otherwise it is generally. Let $\tilde{T}^y$ denote the event time of $\widetilde{N}^y$. Then, likewise for the other statement,

$$P\Big(A\big(T_{(1)}\big) = 0|T_{(1)} = t, \Delta_{(1)} = a, \tilde{T}^y = t_2\Big)$$

is not a function of t_2 if

$$p_{\tilde{T}^y \mid T_{(1)},\Delta_{(1)},A(T_{(1)}),L}(t_2 \mid t, a, 0, l) = p_{\tilde{T}^y \mid T_{(1)},\Delta_{(1)}}(t_2 \mid t, a)$$

Now note that

$$P\Big(\tilde{T}^y \leq t \mid T_{(1)}, \Delta_{(1)} = a, A\big(T_{(1)}\big) = 0, L\Big) = P\Big(\tilde{T}_2 \leq t, \tilde{D}_2 = y \mid T_{(1)}, \Delta_{(1)} = a, A\big(T_{(1)}\big) = 0, L\Big)$$

and

$$P\Big(\tilde{T}^y = \infty \mid T_{(1)}, \Delta_{(1)} = a, A\big(T_{(1)}\big) = 0, L\Big) = P\Big(\tilde{D}_2 = d \mid T_{(1)}, \Delta_{(1)} = a, A\big(T_{(1)}\big) = 0, L\Big)$$

However, it is possible to construct the distribution of the second event times in this way. For example we can let $\tilde{\Delta}_2$ be independent of $A\big(T_{(1)}\big)$ and L, then let $\tilde{T}_2$ be dependent on A and L given $\tilde{\Delta}_2 = d$, but not given $\tilde{\Delta}_2 = y$. We could for instance use the procedure described in Theorem 5.

7 Right-censoring

Now suppose that in addition to the processes we observe, we also observe a component $N^c(t)$ which counts whether or not the subject has dropped out of the study at time t. τ^C denotes the first time at which this process is non-zero. Now consider the weight process,

$$W(t) = W^{g^*}(t) W^c(t) \tag{10}$$

where W^{g^*} denotes the previously studied weight process given in Equation 4 and

$$W^c(t) = \frac{\mathbb{1}\{\tau^C > t\}}{\mathcal{E}(-\Lambda^c)_t} = \mathcal{E}\left(-\int_0^\cdot \frac{1}{1 - \Delta\Lambda^c(s)} \mathrm{d}(N^c - \Lambda^c)\right)_t := \mathcal{E}(\mathbb{K}^c)_t$$

where Λ^c denotes the $\mathcal{F}_t$-compensator of N^c. We will assume that $\langle M^c, M^x\rangle^P_s = 0$ for all $x \neq c$. This impies that

$$W(t) = \mathcal{E}(K^* + \mathbb{K}^c)_t$$

This now yields the following g-formula.

Theorem 6 (g-formula): Let, further, $Q = W(T) \cdot P$ denote the probability measure defined by the likelihood ratio $W(T)$ given in Equation 10. Under positivity, then
1. The Q-$\mathcal{F}_t$ compensator of $N^a(\mathrm{d}t \times \mathrm{d}x)$ is $\pi_t^*(\mathrm{d}x)\Lambda^a(\mathrm{d}t)$.
2. The Q-$\mathcal{F}_t$ compensator of N^x is Λ^x for $x \in \{y, \ell_1, ..., \ell_k\}$.
3. The Q-$\mathcal{F}_t$ compensator of N^c is 0.

Proof: By the same logic as the original theorem, we find

$$\int_0^t \frac{1}{W(s-)} \mathrm{d}\langle W, X\rangle_s^P = \langle K^*, X\rangle_t + \langle \mathbb{K}^c, X\rangle_t$$

Since $\langle \mathbb{K}^c, X\rangle_t = -\int_0^t \frac{1}{1-\Delta\Lambda^c(t)} \mathrm{d}\langle M^c, X\rangle_s$, we find for $X \neq M^c$, the same result as in the previous theorem since $\langle M^c, M^x\rangle_s^P = 0$ for all $x \neq c$. On the other hand, for $X = M^c$, we find

$$\int_0^t \frac{1}{W(s-)} \mathrm{d}\langle W, X\rangle_s^P = \langle \mathbb{K}^c, X\rangle_t = -\int_0^t \frac{1}{1-\Delta\Lambda^c(t)} \mathrm{d}\langle M^c\rangle_s = -\Lambda^c(t)$$

which shows the N^c's compensator under Q is zero. □

8 More general exchangeability conditions

We now consider more general exchangeability conditions.

Theorem 7: Let $Q_\kappa = \mathcal{E}(\kappa)_T \cdot P$ where κ is a local P-$\mathcal{F}_t$ martingale with $\Delta\kappa_t \geq -1$. If
1. Consistency holds as in Theorem 1.
2. $\mathcal{E}(\kappa)_t \mathcal{E}(-\mathbb{N}^a)_t = \mathcal{E}(\kappa)_t$ for all $t \in [0, T]$ P-a.s.
3. Q_κ is a uniformly integrable P-$\mathcal{F}_t$-martingale and P-$\mathcal{H}_t$-martingale, where $\mathcal{H}_t = \mathcal{F}_t \vee \sigma(\tilde{Y})$.

Then,

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t \mathcal{E}(\kappa)_t]$$

Proof: The proof is the same as in Theorem 1 – mutatis mutandis. □

We provide an equivalent characterization of condition 2 in the above theorem which gives direct interpretability of that condition in the sense that it should induce a probability measure Q_κ under which the time to deviation from the treatment regime is infinite almost surely.

Lemma 3: $Q_\kappa(\tau^{g^*} = \infty) = 1$ if and only if $\mathcal{E}(\kappa)_t \mathcal{E}(-\mathbb{N}^a)_t = \mathcal{E}(\kappa)_t$ for all $t \in [0, T]$ P-a.s.

Proof: "If" part:

$$\begin{aligned}
Q_\kappa(\tau^{g^*} = \infty) &= \mathbb{E}_P[\mathcal{E}(\kappa)_T \mathbb{1}\{\tau^{g^*} = \infty\}] \\
&= \mathbb{E}_P\left[\lim_{t\to\infty} \mathcal{E}(\kappa)_t \mathbb{1}\{\tau^{g^*} > t\}\right] \\
&= \mathbb{E}_P\left[\lim_{t\to\infty} \mathcal{E}(\kappa)_t\right] \\
&= \mathbb{E}_P[\mathcal{E}(\kappa)_T] = 1.
\end{aligned}$$

"Only if" part: Suppose that $Q_\kappa(\tau^{g^*} = \infty) = 1$. Then for every $t \in [0, T]$,

$$\begin{aligned}
\mathbb{E}_P[\mathcal{E}(\kappa)_t \mathbb{1}\{\tau^{g^*} > t\}] &= \mathbb{E}_P[\mathbb{E}_P[\mathcal{E}(\kappa)_T \mid \mathcal{F}_t] \mathbb{1}\{\tau^{g^*} > t\}] \\
&= \mathbb{E}_P[\mathcal{E}(\kappa)_T \mathbb{1}\{\tau^{g^*} > t\}] \\
&= Q(\tau^{g^*} > t) \\
&= 1.
\end{aligned}$$

On the other hand, $\mathbb{E}_P[\mathcal{E}(\kappa)_t] = \mathbb{E}_P[\mathbb{E}_P[\mathcal{E}(\kappa)_T \mid \mathcal{F}_t]] = \mathbb{E}_P[\mathcal{E}(\kappa)_T] = 1$. Conclude that

$$\mathbb{E}_P\left[\mathcal{E}(\kappa)_t\left(1 - \mathbb{1}\{\tau^{g^*} > t\}\right)\right] = 0.$$

The integrand on the left hand side is non-negative and so it must be zero P-a.s. □

Lemma 4: Let κ_t be a (local) martingale with $\Delta\kappa_t \geq -1$ and $\Delta\kappa_t^* > -1$ if $t < \tau^{g^*}$. Then,

$$W_t^* := \mathcal{E}(\kappa)_t = \mathcal{E}(\kappa)_t \mathcal{E}(-\mathbb{N}^a)_t$$

if and only if $\Delta\kappa_{\tau^{g^*}} = -1$ whenever $\tau^{g^*} < \infty$.

Proof: Using the well-known formula $\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X, Y])$, we have

$$\mathcal{E}(\kappa) = \mathcal{E}(\kappa - \mathbb{N}^a - [\kappa, \mathbb{N}^a])$$

This holds if and only if

$$1 + \int_0^t W_{t-} \, \mathrm{d}\kappa_s = 1 + \int_0^t W_{t-} \, \mathrm{d}\left(\kappa_s - \mathbb{N}_s^a - [\kappa, \mathbb{N}^a]_s\right)$$

if and only if

$$\int_0^t W_{t-} \Delta\kappa_s \, \mathrm{d}\mathbb{N}_s^a = -\int_0^t W_{t-} \, \mathrm{d}\mathbb{N}_s^a$$

and this is

$$\mathbb{1}\{\tau^{g^*} \leq t\} W_{\tau^{g^*}-} \Delta\kappa_{\tau^{g^*}} = -\mathbb{1}\{\tau^{g^*} \leq t\} W_{\tau^{g^*}-}$$

By assumption, $W_{\tau^{g^*}-} > 0$ (looking at the explicit solution of the SDE) and so the above holds if and only if

$$\mathbb{1}\{\tau^{g^*} \leq t\} \Delta\kappa_{\tau^{g^*}} = -\mathbb{1}\{\tau^{g^*} \leq t\}$$

Taking $t \to \infty$ gives the desired result. On the other hand, if the result holds then,

$$\begin{aligned}
\mathbb{1}\{\tau^{g^*} \leq t\} \Delta\kappa_{\tau^{g^*}} &= \mathbb{1}\{\tau^{g^*} \leq t\} \mathbb{1}\{\tau^{g^*} < \infty\} \Delta\kappa_{\tau^{g^*}} \\
&= \mathbb{1}\{\tau^{g^*} \leq t\} \mathbb{1}\{\tau^{g^*} < \infty\}(-1) = -\mathbb{1}\{\tau^{g^*} \leq t\}
\end{aligned}$$

□

9 Uniqueness

Now, we consider only κ's of the form

$$\kappa(t) = \int \sum_{x \in \mathcal{A}} \mathbb{1}\{s \leq t\} \tilde{h}(s, x) M^{a,x}(\mathrm{d}s)$$

with $\tilde{h}(s, x)$ P-$\mathcal{H}_t$ predictable with the restriction stated in the above theorem and $M^{a,x}$ are P-$\mathcal{H}_t$ local martingales. We make this restriction as any reasonable exchangeability conditions should be placed on treatment and not anything else.

Now, we consider the more interesting question is there a different probability measure $Q_\kappa \neq Q$ such that

$$\mathbb{E}_{Q_\kappa}[Y_t] = \mathbb{E}_Q[Y_t] \tag{11}$$

The answer is no if, additionally, we assume that there is an intensity for the total treatment process $\bar{N}$ in the filtration $\mathcal{F}_t$ and, that, for every other component there, likewise, exists an intensity, and further that

$$\mathbb{E}_{Q_\kappa}[Y_{t \wedge S}] = \mathbb{E}_Q[Y_{t \wedge S}]$$

for every $\mathcal{F}_t$-stopping time S. **NOTE:** We can assume this to hold for slightly different exchangeability conditions where we add $\tilde{Y}_{\cdot \wedge S}$ to the filtration at time zero. Also simpler to just look at the filtrations with $\tilde{Y}_{\cdot \wedge T_{(k)}}$. By the earlier constructions, we can actually get this independence statement.

NOTE: We do not consider baseline elements. We will further assume that $\mathcal{A} = \{a_0, a_1\}$ and that $\pi_s^*(a_1) = 1$. This can be proven in the following way. First, note that κ only contains treatment martingales and by orthogonality that the compensator in Q_κ of every other component than treatment will be the same as in P. Likewise, assuming that $Q_\kappa(\tau^{g^*} = \infty) = 1$, we obtain that the compensator of N^{a,a_0} in Q_κ must be zero. That leaves the compensator N^{a,a_1} in Q_κ to be specified. First take $S = T_{(1)}$. Then, obtain that the corresponding component for the compensator N^{a,a_1} must be equal to that of Q differentiating and using properties of Lebesgue integrals. Continue in this manner for $S = T_{(2)}, T_{(3)}$, etc.

NOTE: Can we make do with orthogonality?

10 Comparison with Coarsening at Random (CAR) conditions of van der Vaart (2004)

NOTE: Need to add to explicitly add likelihood factorization to compare with the factorization of rytgaard.

Let us define the process by $Z(t) = (N^y(t), N^\ell(t), L(t), N^a(t))$. Consider also its potential outcome process $\tilde{Z} = (\widetilde{N}^y, \widetilde{N}^\ell, \tilde{L}, \widetilde{N}^a)$. These are both multivariate cadlag processes. Critically, we take $\mathcal{F}_t = \sigma(Z(s), s \le t)$ – the natural filtration of the observed data process and $\mathcal{H}_t = \mathcal{F}_t \vee \sigma(\tilde{Z}(\cdot))$.

Let R denote the conditional distribution of τ^{g^*} given $\tilde{Z}$ (the conditional distributions exist since the sample space is Polish). Before continuing, we discuss how this distribution is defined and how it relates to our intervention. First given the full process and hence the visitation times, the subject can only deviate at the visitation times. We now generate the process A. Let $A(0) = a_0$ and let $A(t) = a_0$ for $t < T_1$. For each visitation time, i.e., T_k with $\Delta_k = a$, a new random variable $A(T_k)$ is drawn based on the history up to that point, conditional on $\sigma(A(\cdot \wedge T_{k-1}), \tilde{Z})$. Then, we put $A(t) = A(T_k)$ for $T_{(k)} \le t < T_{(k+1)}$. Now let, $k^* := \inf\left\{k \mid \Delta_{(k)} = a, A(T_{(k)}) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)\right\}$. Then, $\tau^{g^*} = T_{(k^*)}$. The coarsened data consists of $X = \left(\tau^{g^*}, \tilde{Z}_{\cdot \wedge \tau^{g^*}}\right)$. For any finite $t > 0$, this means that

$$\begin{aligned}
P\left(\tau^{g^*} \in \mathrm{d}t \mid \tilde{Z} = \tilde{z}\right) &= \sum_k P\left(T_{(k)} \in \mathrm{d}t, \Delta_{(k)} = a, A\left(T_{(k)}\right) = 0, A\left(T_{(k-1)}\right) = \dots = A(0) = 1 \mid \tilde{Z} = \tilde{z}\right) \\
&= \sum_k P\left(A\left(T_{(k)}\right) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = t, \Delta_{(k)} = a, A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right) \\
&\quad \times \prod_{j=1}^{k-1} \left(1 - P\left(A\left(T_{(j)}\right) = 0 \mid \Delta_{(j)} = a, A\left(T_{(j-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right)\right) \\
&\quad \times P\left(T_{(k)} \in \mathrm{d}t, \Delta_{(k)} = a \mid A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right) \qquad (12) \\
&= \sum_k P\left(A\left(T_{(k)}\right) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = t, \Delta_{(k)} = a, A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right) \\
&\quad \times \prod_{j=1}^{k-1} \left(1 - P\left(A\left(T_{(j)}\right) = 0 \mid T_{(j)}, \Delta_{(j)} = a, A\left(T_{(j-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right)\right) \\
&\quad \times \mathbb{1}\left(\tilde{t}_k \in \mathrm{d}t, \tilde{\delta}_k = a\right).
\end{aligned}$$

(We allow the product to be empty, if the person dies before getting a visitation time). Here, we use $\tilde{t}_k$ and $\tilde{\delta}$ to denote the observed event times. Let $k(\tilde{z})$ denote the number of treatment event times in the observed sample. Then, we also have that

$$P\left(\tau^{g^*} = \infty \mid \tilde{Z} = \tilde{z}\right) = 1 - \sum_{j=1}^{k(\tilde{z})} P\left(\tau^{g^*} = t \mid \tilde{Z} = \tilde{z}\right)$$

In van der Vaart (2004), it is assumed that

$$P\left(\tau^{g^*} \in \mathrm{d}t \mid \tilde{Z} = \tilde{z}\right) = p_{\tau^{g^*}}\left(t \mid \tilde{Z} = \tilde{z}\right)\mu(\mathrm{d}t), \qquad (13)$$

for some σ-finite measure μ on $[0, \infty]$. (Note that we allow infinite values of τ^{g^*}, corresponding to no treatment deviation). They introduce a Coarsening at Random (CAR) condition which in our setting can be stated as follows

$$p_{\tau^{g^*}}\left(t \mid \tilde{Z} = \tilde{z}\right) = h(t, \tilde{z}_{\cdot \wedge t}) \qquad (14)$$

for some measurable function $h : [0, \infty] \times D_{[0,\infty)}(\mathbb{R}^d) \to [0, 1]$, where $D_{[0,\infty)}(\mathbb{R}^d)$ denotes the space of càdlàg functions from $[0, \infty)$ to $\mathbb{R}^d$.

Notably, this choice of μ may not depend on $\tilde{z}$. This is a problem as can be seen from the following proposition.

Theorem 8: Suppose that the treatment event times have discrete support, that is there is countable set A, including infinity, such that $P\left(T_{(k)} \in A, \Delta_{(k)} = a\right) = 1$. Then μ can be taken to be the counting measure on A. Suppose that treatment event times are totally inacessible. Then, there exists no such μ.

Proof: The first statement is obvious. On the other hand, note that $\nu := P\left(\tau^{g^*} \in \mathrm{d}t \mid \tilde{Z}\right)$ is a random measure. Its distribution is therefore determined by the Campbell measure C_ν. This means that for all measurable functions h that

$$
\begin{aligned}
&\mathbb{E}_P\left[\int h(s)\nu(\mathrm{d}s)\right] \\
&= \mathbb{E}_P\left[\sum_k \mathbb{1}\left\{\widetilde{\Delta_{(k)}} = a\right\} P\left(A\left(T_{(k)}\right) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)}, \Delta_{(k)} = a, A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right)\right. \\
&\left.\times\, h\left(\widetilde{T_{(k)}}\right)\right] \\
&= \mathbb{E}_P\left[\int h(s) p_{\tau^{g^*}}(s \mid \tilde{z})\mu(\mathrm{d}s)\right] \\
&= \int \mathbb{E}_P[h(s) p_{\tau^{g^*}}(s \mid \tilde{z})]\mu(\mathrm{d}s)
\end{aligned}
$$

Take $h(s) = \mathbb{1}\left\{\widetilde{\Delta_{(k)}} = a, \widetilde{T_{(k)}} = s\right\}$. By total inacesibility, we have that $\mathbb{E}_P[h(s) p_{\tau^{g^*}}(s \mid \tilde{z})] = 0$. On the other hand, this is equal to

$$\mathbb{E}_P\left[\mathbb{1}\left\{\widetilde{\Delta_{(k)}} = a\right\} P\left(A\left(T_{(k)}\right) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)}, \Delta_{(k)} = a, A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z} = \tilde{z}\right)\right]$$

, which would imply that $\mathbb{1}\left\{\widetilde{\Delta_{(k)}} = a\right\} P\left(A\left(T_{(k)}\right) \neq g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)}, \Delta_{(k)} = a, A\left(T_{(k-1)}\right) = \dots = A(0) = 1, \tilde{Z}\right)$ is almost surely equal to zero. This is generally almost equal to zero if there is probability > 0 for an event of type a for the k'th event and

2. non-degenerate treatment probabilities (the second is i.e., a positivity condition).

We can ignore the first requirement suppose that there is a least one such k. □

What Equation 12 suggests is that we work with the following sequential condition:

$$\tilde{Z} \perp \mathbb{1}\{A\left(T_{(k)}\right) = g^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)}, \Delta_{(k)} = a, \mathcal{F}^{g^*}_{T_{(k-1)}}, \tag{15}$$

so that Equation 12 depends on observed data only.

One may try to relax Equation 14 to the Markov kernel

$$P\left(\tau^{g^*} = t \mid \tilde{Z} = \tilde{z}\right) = h(t, \tilde{z}_{\cdot \wedge t})$$

However, we are no longer guaranteed a result like Theorem 2.1 of van der Vaart (2004) or Theorem 25.40 of van der Vaart (1998), without formally having to rederive it. Let us just do this as it is not too difficult.

Theorem 9: Suppose that the distribution of R is restricted by CAR and only CAR. Then $\mathcal{P}^{\cdot}$ is dense in $L_2(P)$.

Proof: Exactly the same as Theorem 25.40 in van der Vaart (1998), but bound everything and consider specific submodels $Q^t = (1 + th)Q^0$, with h bounded. First consider R, Submodels R_t of R are given by

$$\sum_k \mathbb{1}\left\{\Delta_{(k)} = a\right\} r_{t,k}(\dots)\delta_{T_{(k)}}$$

Then,

$$\int \sum_{k} \left(\frac{\left((1+th(x))r_{t,k}\right)^{\frac{1}{2}} - (r)^{\frac{1}{2}}}{t} - \frac{1}{2}(h(x)+a(y))r \right)^2 P ... \mathbb{1}\left\{\Delta_{(k)} = a\right\} \delta_{T_{(k)}}(\mathrm{d}t^a)\, \mathrm{d}Q^0 \to 0$$

Add and subtract $\frac{(r_{t,k})^{\frac{1}{2}}}{t}$. NOTE: Subtle details about dominated convergence here? If we restricted to uniformly bounded "densities", then we would probably also get the result about every score being on a certain form. Use also the argument to see that the scores for R are functions of X only… □

11 Discussion

When applying data analysis in practical scenarios, a key question remains: how best to analyze the data at hand. We explore several potential interventions that could be relevant to those discussed in this article.

Alternatively, a stochastic intervention could be considered, where both the timing of visits and the decisions surrounding them are intervened upon, so that the timing of visitation events is the same as in the observational data. However, such interventions may be difficult to incorporate into a potential outcomes framework. Finally, a simpler approach would be to entirely prevent patients from visiting the doctor, effectively eliminating any possibility of deviation from protocol.

In addition, a significant advantage of this approach, compared to preventative interventions, is the potential to model dynamic treatment regimes, providing alternative means of analysis to the general ones in Ryalen (2024).

Bibliography

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer US. https://doi.org/10.1007/978-1-4612-4348-9

Cohen, S. N., & Elliott, R. J. (2015). *Stochastic calculus and applications* (Vol. 2). Springer.

Gill, R. D., & Robins, J. M. (2001). Causal Inference for Complex Longitudinal Data: The Continuous Case. *The Annals of Statistics*, *29*(6), 1785–1811. https://doi.org/10.1214/aos/1015345962

Gill, R. D., & Robins, J. M. (2023,). *Causal Inference for Complex Longitudinal Data: The Continuous Time g-Computation Formula*. https://arxiv.org/abs/math/0409436

Last, G., & Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamical Approach*. Springer. https://link.springer.com/book/9780387945477

Lok, J. J. (2008). Statistical modeling of causal effects in continuous time. *The Annals of Statistics*, *36*(3), 1464–1507. https://doi.org/10.1214/009053607000000820

Protter, P. E. (2005). *Stochastic differential equations*. Springer.

Richardson, T. S., & Robins, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, *128*(30), 2013.

Robins, J. (1986). 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), 1393–1512. https://doi.org/https://doi.org/10.1016/0270-0255(86)90088-6

Robins, J. M., Hernan, M. A., & Brumback, B. (2000,). *Marginal structural models and causal inference in epidemiology* (Vol. 11, Issue 5, pp. 550–560). Lww.

Ryalen, P. (2024). *On the role of martingales in continuous-time causal inference*.

Rytgaard, H. C., Gerds, T. A., & van der Laan, M. J. (2022). Continuous-Time Targeted Minimum Loss-Based Estimation of Intervention-Specific Mean Outcomes. *The Annals of Stat Istics*, *50*(5), 2469–2491. https://doi.org/10.1214/21-AOS2114

van der Vaart, A. (2004). On Robins' formula. *Statistics & Decisions*, *22*(3), 171–200. https://doi.org/doi:10.1524/stnd.22.3.171.57065

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.