A note on the potential outcomes framework in continuous time

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

Abstract

In this brief note, we consider the seminal work by Ryalen (2024) and compare it with the approach given in Rytgaard et al. (2022), corresponding to their marked point process settings. We study these works in simple multi-state models.

We consider a multi-state model with at most one visitation time for the treatment (that is at most one point where treatment may change), no time-varying covariates, and no baseline covariates. In the initial state everyone starts as treated (0). We consider the setting with no censoring. The multi-state model is shown in Figure 1.

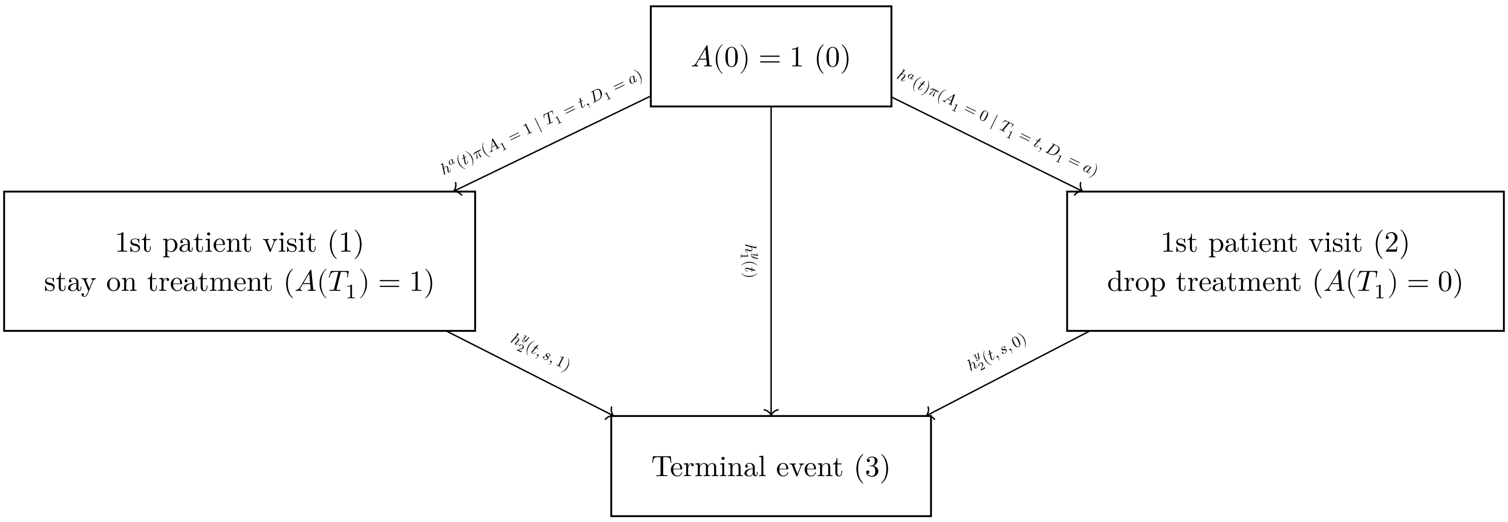

Figure 1: A multi-state model allowing one visitation time for the treatment with the possible treatment values 0/1.

We will represent the observations from such a multi-state model as a marked point process. This means that we can represent the observed data as $O = (T_1, D_1, A_1, T_2, D_2, A_2) \in ((0,\infty] \times (\{a, y\} \times \{0, 1, \emptyset\} \cup \{\nabla\}))^2 := \mathcal{N}^X$. The mark space is given as $X = \{a, y\} \times \{0, 1, \emptyset\} \cup \{\nabla\}$. Here ∇ denotes the empty mark, representing an event that never occurs (which may be the case for the second event) and $\emptyset$ is used for the outcome, T_i denotes the time of the i'th event, D_i is what kind of jump that occurs (so $D_1 \in \{a, y\}$), and A_i is the treatment value at the i'th event (the second event is always an outcome so $A_2 = \nabla$).

Consider a given probability measure P. Let $\mathcal{F}$ be the smallest σ-algebra making the mappings for the sequence

$$
\begin{aligned}
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto t_1, \\
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto d_1, \\
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto a_1, \\
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto t_2, \\
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto d_2, \\
\mathcal{N}^X &\ni (t_1, d_1, a_1, t_2, d_2, a_2) \mapsto a_2
\end{aligned} \tag{1}
$$

measurable (for the real numbers $\mathbb{R}_+ = (0, \infty)$ we use the Borel σ-algebra $\mathcal{B}(0, \infty)$). Next, we consider the associated measure given by

$$
N((0,t] \times \cdot \times \cdot) = \sum_{j=1}^{2} \mathbb{1}\{t_j < \infty\} \delta_{(t_j, d_j, a_j)}((0,t] \times \cdot \times \cdot) = \sum_{j=1}^{2} \mathbb{1}\{t_j \leq t\} \delta_{(d_j, a_j)}(\cdot \times \cdot). \tag{2}
$$

on $(\mathbb{R}_+ \times X, \mathcal{B}(\mathbb{R}_+) \times \mathcal{X}))$. Putting in our observations to the above, we get a random measure that generates a filtration $\mathcal{F}_t = \sigma(N((0,s] \times D \times A) \mid s \leq t, D \subseteq \{a, y\}, A \subseteq \{0, 1, \emptyset\})$. Then, we are in the canonical setting if we use the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P\right)$. To follow along Ryalen (2024), we restrict the observations to the interval $[0, \tau]$ for $\tau > 0$. This means that the observations beyond the point τ have an event time set to ∞ and the mark set to the empty mark ∇. Throughout, we assume that the observations are observed in the interval $[0, \tau]$. We make the intervention definition specific to our marked setting. The important details will be the same.

Consider a mapping $g : \mathbb{R}_+ \times \{a, y\} \to \{0, 1\}$, representing a treatment protocol at the visitation time, e.g. $g(t_1, d_1) = 1$ if always treat at visitation time. Then, we can consider a new point process given by

$$
O^g = (T_1, D_1, g(T_1, D_1), T_2, D_2, A_2). \tag{3}
$$

and $N^g = \mathbb{1}\{T_1 < \infty\} \delta_{(T_1, D_1, g(T_1, D_1))} + \mathbb{1}\{T_2 < \infty\} \delta_{(T_2, D_2, A_2)}$ as the corresponding random measure. Let T^a be the possible time to deviation from the protocol g. That is,

$$
T^a = \begin{cases} T_1 \text{ if } D_1 = a \text{ and } g(T_1, D_1) \neq A_1 \\ \\ \infty \text{ otherwise} \end{cases} \tag{4}
$$

Next, we define the potential outcome process $\tilde{O}$. For this, we need to define the canonical compensator. For N, the counting process measure, the compensating measure $\Lambda = \Lambda(O, dt, dm, da)$ can be chosen to be a kernel from $\mathcal{N}_\tau^X$ to $\mathbb{R}_+ \times X$. This means that $\Lambda(o, dt, dm, da)$ is a measure on $\mathbb{R}_+ \times \{a, y\} \times \{0, 1, \emptyset\}$ such that for each $o \in \mathcal{N}_\tau^X$, $\Lambda(o, \cdot \times \cdot \times \cdot)$ is a measure on $\mathbb{R}_+ \times \{a, y\} \times \{0, 1, \emptyset\}$ and $o \mapsto \Lambda(o, S)$ is measurable for each $S \in \mathcal{B}(\mathbb{R}_+) \times \{a, y\} \times \{0, 1, \emptyset\}$. In our presentation we will make explicit use of this compensator. The intuition behind this definition is that the canonical compensators uniquely determine the distribution of a marked point process, that is the distribution of the jump times and marks. The canonical compensator also makes the marked point processes into a sort-of structural equation system, where we replace the "structural" equations with the compensators. First, we look at the case

$$
\begin{aligned}
\Lambda(O, dt, dm, da) = \delta_a(dm) h_a(t) \mathbb{1}\{t \le T_1\} (\delta_1(da) \pi(A_1 = 1 \mid T_1 = t, D_1 = a) \\
+ \delta_0(da)(1 - \pi(A_1 = 1 \mid T_1 = t, D_1 = a))) dt \\
+ \delta_y(dm) h_1^y(t) \mathbb{1}\{t \le T_1\} dt \\
+ \delta_y(dm) h_2^y(t, s, 1) \mathbb{1}\{D_1 = a, A_1 = 1, T_1 < t \le T_2\} dt \\
+ \delta_y(dm) h_2^y(t, s, 0) \mathbb{1}\{D_1 = a, A_1 = 0, T_1 < t \le T_2\} dt
\end{aligned}
\tag{5}
$$

Here we take $h^{a,1} : \mathbb{R}_+ \to \mathbb{R}_+, h^{a,0} : \mathbb{R}_+ \to \mathbb{R}_+, h_1^y : \mathbb{R}_+ \to \mathbb{R}_+$ and $h_2^y : \mathbb{R}_+ \times \mathbb{R}_+ \times \{0,1\} \to \mathbb{R}_+$ to be fixed and non-random functions, $\pi : \{0,1\} \times \mathbb{R}_+ \times \{a, y\} \to [0,1]$ to be a fixed and non-random function, such that $\pi(A_1 = 1 \mid T_1 = t, D_1 = a) + \pi(A_1 = 0 \mid T_1 = t, D_1 = a) = 1$. Afterwards, we look at the case that the total compensator is not absolutely continuous. This turns out to be important. Notice the sums between different types of events. This corresponds to assuming that the events cannot happen at the same time. Let us consider the static intervention $g(t, d) = 1$ and find the canonical compensator for O^g. This can be found by computing N^g at various points. For example,

$$
\begin{aligned}
N^g((0,t] \times \{a\} \times \{0\}) &= 0 \\
N^g((0,t] \times \{a\} \times \{0,1\}) &= N((0,t] \times \{a\} \times \{0,1\}) \\
N^g((0,t] \times \{y\} \times \{\emptyset\}) &= N((0,t] \times \{y\} \times \{\emptyset\})
\end{aligned}
\tag{6}
$$

from which the canonical compensator is easily derived to be,

$$
\begin{aligned}
\Lambda^g(O^g, dt, dm, da) = \delta_a(dm) h_a(t) \mathbb{1}\{t \le T_1\} \delta_1(da) dt \\
+ \delta_y(dm) h_1^y(t) \mathbb{1}\{t \le T_1\} dt \\
+ \delta_y(dm) h_2^y(t, s, 1) \mathbb{1}\{D_1 = a, A_1 = 1, T_1 < t \le T_2\} dt \\
+ \delta_y(dm) h_2^y(t, s, 0) \mathbb{1}\{D_1 = a, A_1 = 0, T_1 < t \le T_2\} dt
\end{aligned}
\tag{7}
$$

A process is a potential outcome $\tilde{O}$ if its compensator is given by $\Lambda^g\left(\tilde{O}, dt, dm, da\right)$. The outcome of interest is defined as a function of the data O, that is $Y = Y(O)$. The associated counterfactual filtration is $\tilde{\mathcal{F}}_t = \sigma\left(\tilde{N}((0,s] \times D \times A) \mid s \le t, D \subseteq \{a, y\}, A \subseteq \{0, 1, \emptyset\}\right)$. Conceptually, we should think of consistency in this setting as $O = \tilde{O}$ if $T^A = \infty$ and $O = \left(\tilde{T}_1, \tilde{D}_1, A_1, T_2, D_2, A_2\right)$ if $T^A < \infty$. In this sense, $\tilde{Y} = Y\left(\tilde{O}\right)$ is the potential outcome process of interest. In our setting, we put $Y_t(O) = \mathbb{1}\{T_1 \le t, D_1 = y\} + \mathbb{1}\{T_2 \le t, D_2 = y\}$, corresponding to a terminal event before time t.

Consider the intervention $g(t, d) = 1$. First, note that $(T^a \le t) = (D_1 = a, A_1 = 0, T_1 \le t) \in \mathcal{F}_t$ for each $t > 0$ which means that T^a is a stopping time with respect to the filtration $\mathcal{F}_t$. Then three basic conditions for identifiability are

1. Consistency: $\tilde{Y}_t \mathbb{1}\{T^A > t\} = Y_t \mathbb{1}\{T^A > t\}$ for all $t > 0$ P-a.s.
2. Exchangeability: The compensator $\Lambda(dt, \{a\}, da)$ for the filtration $\mathcal{F}_t$ is the same as it is for $\mathcal{G}_t = \mathcal{F}_t \vee \sigma\left(\tilde{Y}_s, \tau \ge s \ge 0\right)$.
3. Positivity: $W_t = \frac{\mathbb{1}\{T^A > t\}}{\exp(-\Lambda((0,t] \times \{a\} \times \{0\}))} = \frac{1 - \mathbb{1}\{T_1 \le t, D = a, A = 0\}}{\exp\left(-\int_0^t \mathbb{1}\{s \le T_1\} h^a(s) \pi(A_1 = 0 \mid T_1 = s, D_1 = a) ds\right)}$ a uniformly integrable martingale, where $\mathscr{P}$ denotes the product integral. This is the same as stating that the measure Q given by $dQ = V_\tau dP$ is a probability measure.

Then, $\mathbb{E}\left[\tilde{Y}_t\right] = \mathbb{E}[W_t Y_t]$ (Theorem 1 in Ryalen (2024)). This construction is different from the usual approach in the sense that inverse probability weighting is used to derive the g-formula and not the other way around. Note that W_t cannot be directly interpreted as a likelihood ratio of interest, for in this case the likelihood ratio that Rytgaard et al. (2022) dictates is

$$\bar{W}_t = \left(\frac{\mathbb{1}\{A(T_k) = 1\}}{\pi_{T_k}\left(A(T_k) = 1 \mid D(T_k) = a, \mathcal{F}_{T_K}\right)} \right)^{\mathbb{1}\{T_k \le t, D_k = a\}} \tag{8}$$

The difference is that $\bar{W}_t$ corresponds to the likelihood ratio in which the difference between the two probability measures is that we replace the intensity from state 0 to 2 with 0 and the intensity from state 0 to 1 with $h^a(t)$. On the other hand, W_t corresponds to likelihood ratio in which the difference between the two probability measures is that we replace the intensity from state 0 to 2 with 0. However, this is not a formal proof that the g-formula in Rytgaard et al. (2022) is not the same as the one in Ryalen (2024). There may be multiple inverse probability weightings that are valid.

We are now ready to compare the two approaches. The approach in Rytgaard et al. (2022) stipulates that the Q-$\mathcal{F}_t$ compensator of $N((0,t] \times \{a\} \times \cdot)$ is given by $\Lambda^Q((0,t], \{a\}, \cdot) = \Lambda^g((0,t], \{a\}, \cdot)$ and that the Q-$\mathcal{F}_t$ compensator of $N((0,t] \times \{y\} \times \cdot)$ is given by $\Lambda^Q((0,t], \{y\}, \cdot) = \Lambda((0,t], \{y\}, \cdot)$. We will show that this is indeed the case if $\Lambda(\{t\} \times \{a\} \times \{0\}) = 0$ which happens if $N((0,t] \times \{a\} \times \{0\})$ has a compensator which is absolutely continuous with respect to the Lebesgue measure. To this end, we shall calculate $N^{g \times a}(t) = \Big[N^g((0,\cdot] \times \{a\} \times A), \mathbb{1}\{T^a \le \cdot\Big]_t = \sum_{s \le t} \mathbb{1}\{1 \in A\} N^g(\{s\} \times \{a\} \times A)\Big)\mathbb{1}\{T^a = s\} = 0$ - the optional covariation of the two processes. This is zero because the two processes have no jumps in common, which is due to the fact that you cannot have a visitation time in which the doctor both treats and does not treat, i.e. you cannot jump to state 1 and 2 at the same time. Since we have that $N((0,t] \times \{y\} \times \cdot)$ and $N^g((0,t] \times \{a\} \times \cdot)$ cannot jump at the same time (follows from the decomposition), then the same argument applies to $N^{g \times y}(t) = [N((0,\cdot] \times \{y\} \times \cdot), \mathbb{1}\{T^a \le \cdot\}]_t$ Thus their respective compensators are $\Lambda^{g \times a}((0,t] \times \cdot) = \Lambda^{g \times y}((0,t] \times \cdot) = 0$. Using Pål's g-formula (Theorem 2 in Ryalen (2024)) gives

$$\begin{aligned}
\Lambda^Q((0,t] \times \{a\}, A) &= \int_{(0,t]} \frac{d\Lambda^g(\cdot \times \{a\} \times A) - d\Lambda^{g \times a}(\cdot \times A)}{1 - \Delta\Lambda^a_s(\{0\})} = \int_{(0,t]} \frac{\Lambda^g(\cdot \times \{a\} \times A) - 0}{1 - 0} \\
&= \Lambda^g((0,t] \times \{a\} \times A) \\
\Lambda^Q((0,t] \times \{y\}, \{\emptyset\}) &= \int_{(0,t]} \frac{d\Lambda^g(\cdot \times \{y\} \times \{\emptyset\}) - \Lambda^{g \times y}(\cdot \times \{\emptyset\})}{1 - \Delta\Lambda^a_s(\{0\})} = \int_{(0,t]} \frac{\Lambda(\cdot \times \{y\}, \{\emptyset\}) - 0}{1 - 0} \\
&= \Lambda((0,t] \times \{y\} \times \{\emptyset\}).
\end{aligned} \tag{9}$$

Critically, we used that the compensator is absolutely continuous with respect to the Lebesgue measure.

Now we assume in contrast that the total compensator is

$$\Lambda((0,t] \times \{a\} \times \{0,1\}) = \int_{(0,t]} \mathbb{1}\{s \le T_1\} dK(s) = c\mathbb{1}\{1 \le T_1 \wedge t\}, c \in (0,1) \tag{10}$$

where

$$K(s) = c\mathbb{1}\{1 \le s\} \tag{11}$$

Evidently, $\Lambda((0,t] \times \{a\} \times da) = \sum_{m=0,1} \delta_m(da) \pi(A(T_1) = m \mid D(T_1) = a, T_1 = s) \Lambda((0,t] \times \{a\} \times \{0,1\})$ is predictable, increasing, cadlag, but not absolutely continuous with respect to the Lebesgue measure. We find that with the same calculations as previously,

$$
\begin{aligned}
\Lambda^{Q}(\{s\}\times\{a\}\times\{1\}) &= \frac{1}{1-\Lambda(\{s\}\times\{a\}\times\{0\})}(\Lambda^{g}(\{s\}\times\{a\}\times\{1\})-\Lambda^{g\times a}(\{s\}\times\{a\}\times\{1\})) \\
&= \frac{1}{1-\pi(A(T_1)=0 \mid D(T_1)=a, T_1=s)c\mathbb{1}\{s\leq T_1, s=1)}\Lambda^{g}(\{s\}\times\{a\}\times\{1\})
\end{aligned} \tag{12}
$$

If $s \neq 1$, then $\Lambda^{Q}(\{s\}\times\{a\}\times\{1\}) = \Lambda^{g}(\{s\}\times\{a\}\times\{1\})$. However, with positive probability, $\Lambda^{Q}(\{s\}\times\{a\}\times\{1\}) = \frac{1}{1-\pi(A(T_1)=0 \mid D(T_1)=a, T_1=1)c\mathbb{1}\{1\leq T_1)}\Lambda^{g}(\{s\}\times\{a\}\times\{1\}) \neq \Lambda^{g}(\{s\}\times\{a\}\times\{1\})$ for $s=1$. A similar conclusion be drawn for the other compensators.

0.1 Notes to self (do not read!)

All considerations imply that interventional probability of dying before time t is given by

$$
\begin{aligned}
\mathbb{E}\left[\tilde{Y}_t\right] &= \int_0^t \exp\left(-\int_0^s h^{a,1}(u)+h_1^y(u)du\right)h_1^y(s)ds \\
&+ \int_0^t \exp\left(-\int_0^s h^{a,1}(u)+h_1^y(u)du\right)h^{a,1}(s)\int_s^t \exp\left(-\int_s^w h_2(u,s,1)du\right)h_2(w,s,1)dwds
\end{aligned} \tag{13}
$$

Now compare this with

$$
\begin{aligned}
\mathbb{E}[Y_t] = &\int_0^t \exp\left(-\int_0^s h^{a,1}(u)+h^{a,0}(u)+h_1^y(u)du\right)h_1^y(s)ds \\
&+ \int_0^t \exp\left(-\int_s^w h^{a,1}(u)+h^{a,0}(u)+h_1^y(u)du\right)h^{a,1}(s)\int_s^t \exp\left(-\int_s^w h_2(u,s,1)du\right)h_2(w,s,1)dwds \\
&+ \int_0^t \exp\left(-\int_s^w h^{a,1}(u)+h^{a,0}(u)+h_1^y(u)du\right)h^{a,0}(s)\int_s^t \exp\left(-\int_s^w h_2(u,s,0)du\right)h_2(w,s,0)dwds
\end{aligned} \tag{14}
$$

which is the observed probability of dying before time t.

Note to self: According to Pål's product integral formula (Lemma 1 in Ryalen (2024)), the previous equation can be written as

$$
\begin{aligned}
\mathbb{E}[Y_t] = \int_0^\infty &\exp\left(-\int_0^s h^{a,1}(u)\mathbb{1}\{u \leq \infty\} + h^{a,0}(u)\mathbb{1}\{u \leq \infty\} + h_1^y(u)\mathbb{1}\{u \leq \infty\}du\right) h_1^y(s)\mathbb{1}\{s \leq \infty\} \cdot \mathbb{1}\{s \leq t\}ds \\
+ \int_0^\infty &\exp\left(-\int_0^s h^{a,1}(u)\mathbb{1}\{u \leq \infty\} + h^{a,0}(u)\mathbb{1}\{u \leq \infty\} + h_1^y(u)\mathbb{1}\{u \leq \infty\}du\right) h^{a,1}(s)\mathbb{1}\{s \leq \infty\} \cdot 0ds \\
+ \int_0^\infty &\exp\left(-\int_0^s h^{a,1}(u)\mathbb{1}\{u \leq \infty\} + h^{a,0}(u)\mathbb{1}\{u \leq \infty\} + h_1^y(u)\mathbb{1}\{u \leq \infty\}du\right) h^{a,0}(s)\mathbb{1}\{s \leq \infty\} \cdot 0ds \\
+ \int_0^\infty &\exp\left(-\int_0^s h^{a,1}(u)\mathbb{1}\{u \leq \infty\} + h^{a,0}(u)\mathbb{1}\{u \leq \infty\} + h_1^y(u)\mathbb{1}\{u \leq \infty\}du\right) h^{a,1}(s)\mathbb{1}\{s \leq \infty\} \qquad (15) \\
&\int_0^\infty \exp\left(-\int_s^w h_2(u,s,1)\mathbb{1}\{s < u \leq \infty\}du\right) h_2(w,s,1)\mathbb{1}\{s < w \leq \infty\}\mathbb{1}\{s \leq t\}dwds \\
+ \int_0^\infty &\exp\left(-\int_0^s h^{a,1}(u)\mathbb{1}\{u \leq \infty\} + h^{a,0}(u)\mathbb{1}\{u \leq \infty\} + h_1^y(u)\mathbb{1}\{u \leq \infty\}du\right) h^{a,0}(s)\mathbb{1}\{s \leq \infty\} \\
&\int_0^\infty \exp\left(-\int_s^w h_2(u,s,0)\mathbb{1}\{s < u \leq \infty\}du\right) h_2(w,s,0)\mathbb{1}\{s < w \leq \infty\}\mathbb{1}\{s \leq t\}dwds
\end{aligned}
$$

The at-risk indicator's are present in the above formulation can be seen by the following: The zero process does not have any jump points. Hence, the first jump point is at infinity in the outer integral. Similarly, in the inner integral, the random measure $\delta_{(t_1,x_1)}$ does not have a second jump point.

If we do the calculation corresponding to a multi-state model where one of the intermediate states is treatment discontinuation, correspondin to the deletion of state 1, we will get the ordinary g-formula too.

The g-formula can fail in other situations, even if $\Delta\Psi_t^a = 0$. In this case, it has to be possible for $\mathscr{N}_t^a$ to jump at other times than the jump times of N_t^a. Consider the example multi-state model in Figure 2.

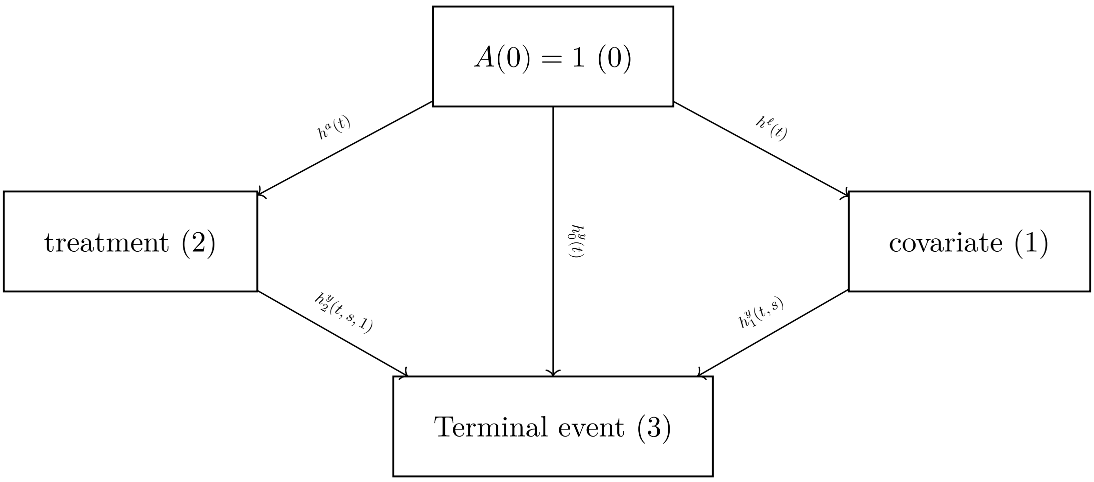

Figure 2: A multi-state model allowing a visitation time for the treatment and a different visitation time for the covariate (P).

We use instead the intervention of treatment given by $\boldsymbol{n}_t^a(\varphi) = \varphi^{01}$, that is treat if a covariate/comorbidity changed. In this case $\mathbb{1}\{\tau^A \le t\} = N^{01}(t) + N^{02}(t)$. However,

$$\mathbb{N}^{a,\psi} = \mathbb{1}\{\tau^A \le t, \mathcal{N}^a_\tau - \mathcal{N}^a_{\tau-} = 1\} = N^{01}_t. \tag{16}$$

So that the compensator is $\mathcal{L}_t^a = 0 \neq \int_0^t \lambda^{01}(s)ds$. If we calculate the Q-compensator for all the other states we get a multi-state model as illustrated in Figure 3.

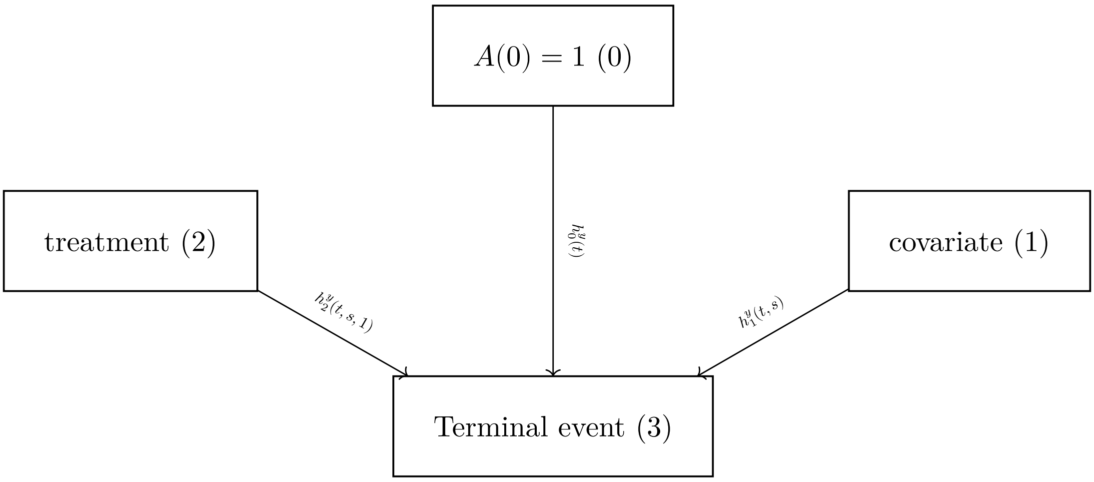

Figure 3: A multi-state model allowing a visitation time for the treatment and a different visitation time for the covariate (Q).

1 Multiple event points

Note that

$$\mathbb{1}\{\tau^A \le t\} = \sum_k \mathbb{1}\Big\{T_k \le t, D_k = a, A(T_k) = 0, A(T_{k-1}) = \ldots = A_{T_1} = 1\Big\} \tag{17}$$

In my notes I found that $\mathbb{1}\{T_k \le t, D_k = a, A(T_k) = 0\}$ is a martingale with compensator $\int_0^t \lambda_s(t, a, 1, \mathcal{X})\mathbb{1}\{T_{k-1} < s \le T_k\}ds$.

Consider for $s < t$

$$\begin{aligned}
&\mathbb{E}\Bigg[\mathbb{1}\{T_k \le t, D_k = a, A(T_k) = 0, A(T_{k-1}) = \ldots = A(T_1) = 1\} \\
&\qquad - \int_0^t \lambda_u(a, 1, \mathcal{X})\mathbb{1}\{T_{k-1} < u \le T_k)du\mathbb{1}\{A(T_{k-1}) = \ldots = A(T_1) = 1\} \mid \mathcal{F}_s\Bigg] \\
&= \mathbb{E}[\mathbb{1}\{T_{k-1} \le s\}\Bigg(\mathbb{1}\{T_k \le t, D_k = a, A(T_k) = 0, A(T_{k-1}) = \ldots = A(T_1) = 1\} \\
&\qquad - \int_0^t \lambda_u(a, 1, \mathcal{X})\mathbb{1}\{T_{k-1} < u \le T_k\}du\mathbb{1}\{A(T_{k-1}) = \ldots = A(T_1) = 1\} \mid \mathcal{F}_s\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{1}\{T_k \le t, D_k = a, A(T_k) = 0\} \\
&\qquad - \int_0^t \lambda_s(t, a, 1, \mathcal{X})\mathbb{1}\{T_{k-1} < s \le T_k\}ds| \mathcal{F}_s\Bigg]\mathbb{1}\{A(T_{k-1}) = \ldots = A(T_1) = 1\}\mathbb{1}\{T_{k-1} \le s\} \\
&= 0
\end{aligned} \tag{18}$$

Since $\int_0^t \lambda_{s(t,a,1,\mathcal{X})} \mathbb{1}\{T_{k-1} < s \leq T_k, A(T_{k-1}) = ... = A(T_1) = 1\} ds$ is continuous and adapted, it is predictable. It is also increasing, so it is the compensator to the process $\mathbb{1}\{T_k \leq t, D_k = a, A(T_k) = 0, A(T_{k-1}) = ... = A(T_1) = 1\}$. Thus $\Delta \Psi_t^a = 0$ (identically zero).

Using this decomposition, we find

$$
\begin{aligned}
&\mathbb{1}(\tau^A \leq t) \mathbb{1}(\mathcal{N}^a(\{\tau^a\}, \{1\})) \\
&= \sum_k \mathbb{1}\{T_k \leq t, D_k = a, A(T_k) = 0, A(T_{k-1}) = ... = A(T_1) = 1\} \mathbb{1}(\mathcal{N}^a(\{\tau^a\}, \{1\}) = 1) \\
&= \sum_k \mathbb{1}\{T_k \leq t, D_k = a, A(T_k) = 0, A(T_{k-1}) = ... = A(T_1) = 1\} (N^a_{\tau^A}(\{1\}) - N^a_{\tau^A -}(\{1\})) \mathbb{1}(A(T_k) = 1) \\
&= 0
\end{aligned}
\tag{19}
$$

This means that the g-formulas stay the same as with the single event case.

In the situation with dynamic treatment based on the history, i.e., $\pi^*(l, \varphi, t, dx) = \delta_{\{g(l,\varphi,t)\}}(dx)$, we will likely get the same result. Importantly, we need that the time-varying covariates cannot jump at the same time as the treatment visitation times and moreover that $\Delta \Psi_t^a = 0$, which is the case if Ψ_t^a is absolutely continuous with respect to the Lebesgue measure.

Bibliography

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

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