Causal inference in continuous time

June 13, 2024

Introduction

- Randomised control trials (RCTs) are desired to evaluate treatment effect, but they are usually hard to implement.
- Causal inference methods like g-methods are needed.
- G-methods: inverse probability weighting with marginal structural models, parametric g-formula, and g-estimation for structural nested models. These are well-developed when changes occur in discrete time (e.g. Robins g-methods[7]), but often treatments are changed on no fixed schedule.

Motivation: toy examples of treatment processes

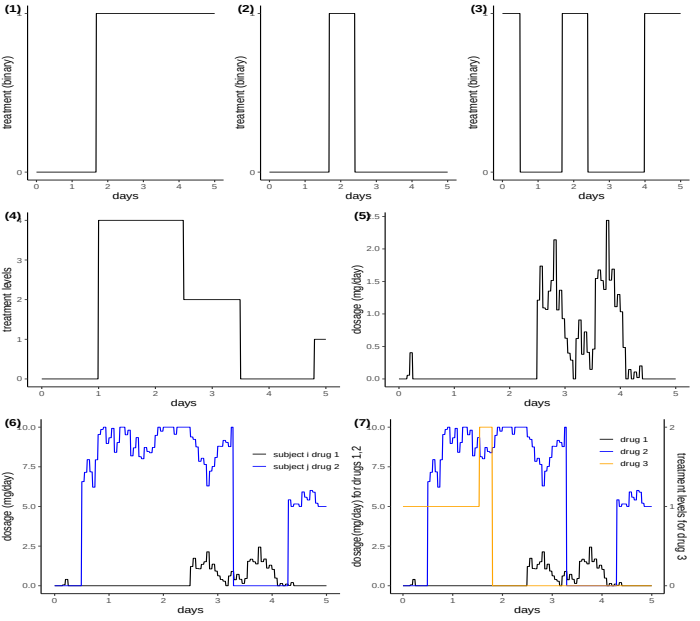


Figure 1: Toy examples of treatment processes.

The UK 'Towards A CurE for rheumatoid arthritis' (TACERA) cohort [2]

- The cohort consists of newly diagnosed seropositive rheumatoid arthritis (RA) patients who were followed up at approximately 3-month intervals for up to 18 months.
- Causal question: Does high dose ($\geq 15 mg/day$) of methotrexate affect time to disease remission?

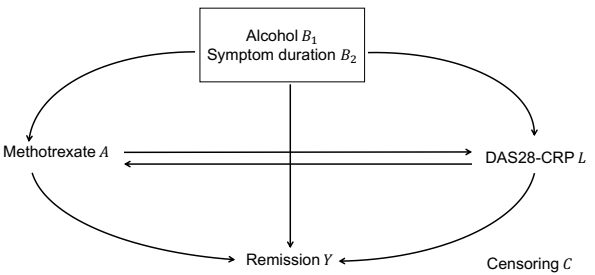


Figure 2: The local independence graph[3, 4].

Motivation cont.

- Challenges in time discretisation[6]:
  - too coarse: might invalidate the causal assumptions, e.g. no unmeasured confounding;
  - too fined: reduce bias but increase variance, practical violation of positivity assumption, computation burden;
  - different discretizations can lead to different target parameters.
- Several continuous-time g-methods have been proposed, but literature is dispersed and involves technical complexities, so have been little used.

Aim

To review of the existing methods to encourage adoption of their use by applied researchers.

Overview

1. Frame the causal problem for TACERA cohort.
2. Demonstrate
   - Continuous-time IPW-MSMs
   - Continuous-time g-computation formula
   - Continuous-time structural nested failure time models (SNFTMs)

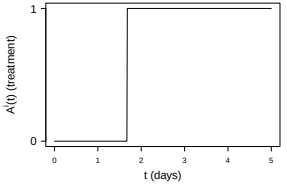
TACERA

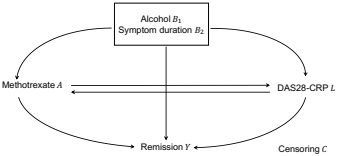
- Treatment: high-dose methotrexate $A(t) = 1$, otherwise $A(t) = 0$.
- Baseline confounder: alcohol consumption indicator (B_1), short symptom duration (< 5 months at baseline) indicator (B_2).
- Time-varying confounder: 28-joint Disease Activity Score with C-reactive protein (DAS28-CRP): $L(t) \in [0, 9.6]$.
- Outcome: disease remission ($Y(t) = 1$) happens when $L(t) < 2.6$, otherwise $Y(t) = 0$.
- Censoring[13]: $C(t) = 1$ if censored.

Exclude individuals who did not receive methotraxate. Exclude those with missing DAS28-CRP scores or whose baseline DAS28CRP scores < 2.6 at baseline. There are 172 individuals included in this study.

Other notation:

- The value of variable X just before time t: $X(t-)$. Let $\overline{X}(t) = (X(u) : u \in [0, t])$, $\underline{X}(t) = (X(u) : u \in [t, T])$.
- Time to event X: T^X.
- Counting process for process X that counts changes in $X(t)$: $N^X(t)$.
- At-risk process $G^X(t) = \mathbb{I}(t \leq T^X)$ and $X(t) = \int_0^t G^X(u) dN^X(u)$.
- Filtration of X till time t: $\mathcal{F}^X(t) = \sigma(\overline{X}(t))$.
- Hazard function: $h^X(t)$ and intensity process $\lambda^X(t)dt = E[dN^X(t) = 1|\mathcal{F}_{t-}] = P(dX(t) = 1|\mathcal{F}_{t-}) = G^X(t)h^X(t)$





Causal Assumptions[7]

Counterfactual had treatment strategy $\overline{a}$ been given: $Y^{\overline{a}}$. Let P and $\tilde{P}$ denote observational and hypothetical probability respectively.

- **No unmeasured confounding**: the confounder provides sufficient information such that at any time, the present treatment is independent of the future counterfactual given the history of confounders, past treatment, and not censored.

$$\lambda^A(t|\overline{L}(t), \overline{A}(t-), \underline{Y}^{\overline{a}}(t+), B, C(t) = Y(t) = 0) = \lambda^A(t|\overline{L}(t), \overline{A}(t-), B, C(t) = Y(t) = 0).$$

- **Positivity**: $\tilde{P} << P$[1] [10, 11, 19].
For $P(\overline{l}(t), \overline{a}(t-), C(t) = Y(t) = 0, B) > 0$, the hypothetical treatment has $\tilde{P}(dA(t)|\overline{a}(t-), C(t) = Y(t) = 0) > 0$, i.e.
$\tilde{\lambda}^A(t|\overline{a}(t-), C(t) = Y(t) = 0) > 0$. Then the positivity assumption requires $P(dA(t)|\overline{l}(t), \overline{a}(t-), C(t) = Y(t) = 0, B) > 0$, i.e
$\lambda^A(t|\overline{l}(t), \overline{a}(t-), C(t) = Y(t) = 0, B) > 0$.
- **Consistency**:

$$\text{If } \overline{A}(t-) = \overline{a}(t-) \text{ then } \overline{Y}(t) = \overline{Y}^{\overline{a}}(t).$$

[1] This ensures Radon-Nikodym derivative $\frac{d\tilde{P}}{dP}$ exists.

IPW-MSMs

- Idea: create a pseudo-population that mimics a situation under randomisation [12].
- Weight[10]:

$$R(t_k) = \prod_{u=0}^{k} \frac{P(A(t_u)|\bar{A}(t_{u-1}), T^Y \geq u)}{P(A(t_u)|\bar{A}(t_{u-1}), \bar{L}(t_u), B, T^Y \geq t_u)} = \prod_{u=1}^{k} \left\{ \frac{\tilde{\lambda}^A(t_u)}{\lambda^A(t_u)} \right\}^{\Delta N^A(t_u)} \frac{\prod_{u=1}^{k} \left\{1 - \tilde{\lambda}^A(t_u)\right\}^{1-\Delta N^A(t_u)}}{\prod_{u=1}^{k} \{1 - \lambda^A(t_u)\}^{1-\Delta N^A(t_u)}}$$

$$R(t) = \prod_{u \leq t} \left\{ \frac{\tilde{\lambda}^A(u)}{\lambda^A(u)} \right\}^{dN^A(u)} \exp\left\{ \int_0^t \lambda^A(s) - \tilde{\lambda}^A(s) ds \right\} \tag{1}$$

- Marginal mean model:

$$\tilde{h}^Y\left(t|\bar{a}(t-)\right) = E_{\tilde{P}}\left(dN^Y(t)|\bar{A}(t-) = \bar{a}(t-)\right). \tag{2}$$

IPW-MSMs

- Model fitting:

	marginal treatment hazard	conditional treatment hazard	outcome hazard
Additive hazard model[10, 13]	$h^A(t) = \beta_0(t)$	$h^A(t) = \beta_0(t) + \beta_1 B_1 + \beta_2 B_2 + \beta_3(t)L(t)$	$\tilde{h}^Y(t) = \gamma_0(t) + \gamma_1(t)A(t-)$
Cox-PH model[5]	$h^A(t) = h_0^A(t)$	$h^A(t) = h_0^a(t) \times \exp(\beta_1 B_1 + \beta_2 B_2 + \beta_3 L(t))$	$\tilde{h}^Y(t) = h_0(t)\exp(\gamma_1 A(t-))$
Case-base sampling[5, 14, 15]	$\texttt{logit}(h^A(t)) = \beta_0 + \sum_{k\in\{1,2\}} \beta_{1k}(t) + f^A$	$\texttt{logit}(h^A(t)) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 L(t) + \sum_{k\in\{1,2,3\}} \beta_{5k}(t) + f^A$	$\texttt{logit}(\tilde{h}^Y(t)) = \gamma_0 + \gamma_1 A(t-) + \sum_{k\in\{1,2,3\}} \gamma_{2k}(t) + f^Y$

Table 1: Models fitted for the treatment process and outcome process.

- Weight estimation:

Additive	Recursively estimated by the Doléans-Dade theorem [1, 17][2](R package `ahw`)
Cox	Breslow estimator for cumulative baseline hazard
Case-base	multiplication

Table 2: Ways to estimate $R(t)$.

[2] $R^{i,A}(t) = R_0^{i,A} + \int_0^t R^{i,A}(s-)dK^i(s)$, $K^i(t) = \int_0^t (\theta^{i,A}(s) - 1)dN^{i,A}(s) + \int_0^t G^{i,A}(s)X^i(s)^T dB(s) - \int_0^t G^{i,A}(s)\tilde{X}^i(s)^T d\tilde{B}(s)$

IPW-MSM

Causal parameters:

- Additive: $H_0 : \gamma_1(t) = 0$ with P-value being 0.033; $H_0 : \gamma_1(t) = \gamma$ for some constant γ with p-value being 0.178.
- Cox: The causal parameter $\hat{\gamma}_1 = 0.279$ with p-value 0.119.
- Case-base: The causal parameter $\hat{\gamma}_1 = 0.338$ with p-value 0.091.

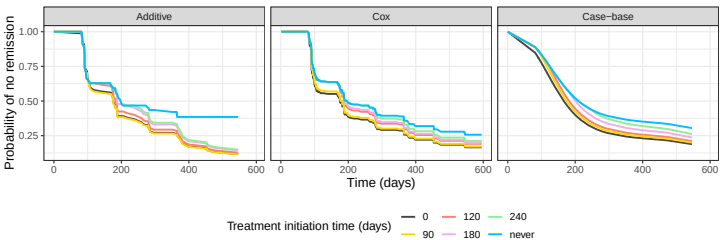


Figure 3: Survival curves estimated using IPW-MSMs with additive hazard models, Cox models, and case-base sampling methods.

High-dose methotrexate is beneficial for remission of RA. Earlier treatment initiation leads to higher probability of remission over time, but the evidence is in not strong in general.

IPW-MSMs: covariate balance diagnostics

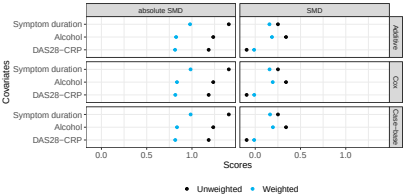


Figure 4: Aggregated standardised mean difference (SMD) and Absolute SMD [3].

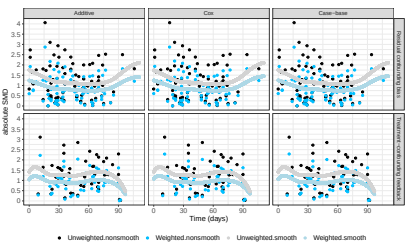


Figure 5: Absolute SMD scores for DAS28-CRP scores: nonsmoothed vs smoothed by kernel regression.

[3] $s_L(t_a, t_l) = \left(\mathbb{E}[R(t_a)L(t_l)\mathbb{I}(A(t_a)=1)|\overline{A}(t_a-)=0, C(t_a)=Y(t_a)=0] - \mathbb{E}[R(t_a)L(t_l)\mathbb{I}(A(t_a)=0)|\overline{A}(t_a-)=0, C(t_a)=Y(t_a)=0]\right) / \sigma_{pool}(t_a, t_l)$

Parametric g-formula

$$E_{\tilde{P}}[dN^Y(T)] = \int_0^T \int_{\bar{l}(t)} \lambda^Y(u; \bar{L}(u-), \bar{a}(u-), B) \prod_{0<s<u} \Big(1 - \lambda^Y(s; \bar{L}(s-), \bar{a}(s-), B)\Big) \times$$
$$\prod_{0<s\leq u} \Big(\lambda^L(s; \bar{L}(s-), \bar{a}(s-), B) P(L(s)|\bar{L}(s-), \bar{a}(s-), B)\Big)^{dN^L(t)}$$
$$\Big(1 - \lambda^L(s; \bar{L}(s-), \bar{a}(s-) = 0, B)\Big)^{1-dN^L(t)} P(L(0), B) \tilde{\lambda}^A(s|\bar{a}(s-), \bar{l}(s), B) d\bar{l} du. \tag{3}$$

Multistates model!

- Artificially manipulate the transition intensities to impose the intervention.
- The transition matrix $\tilde{\boldsymbol{P}}(s,t) = \prod_{s<t_i\leq t}(\boldsymbol{I} + d\tilde{\boldsymbol{\Lambda}}(t_i))$.

Parametric g-formula

- Categorise $L(t)$ into three levels: remission ($L(t) \in (0, 2.6]$), mild condition ($L(t) \in (2.6, 5.1]$), severe condition ($L(t) \in (5.1, 9.6)$).
- Compare $\overline{a} = 1$ with $\overline{a} = 0$.

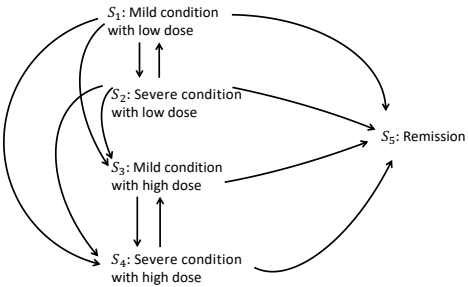


Figure 6: Multi-state diagram for TACERA study.

Parametric g-formula

$\overline{a} = 1$:

$$\tilde{\lambda}^{\overline{1}}(t) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -(\lambda_{34}(t) + \lambda_{35}(t)) & \lambda_{34}(t) & \lambda_{35}(t) \\ 0 & 0 & \lambda_{43}(t) & -(\lambda_{43}(t) + \lambda_{45}(t)) & \lambda_{45}(t) \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4}$$

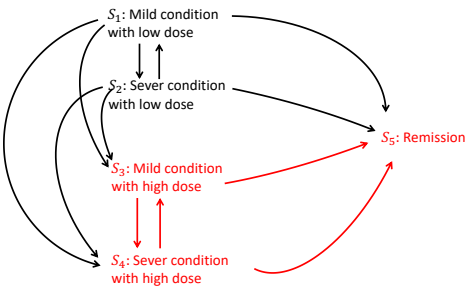


Figure 7: Multi-state diagram for TACERA study.

Parametric g-formula

$\overline{a} = 0$:

$$\bar{\lambda}^{\overline{0}}(t) = \begin{pmatrix} -(\lambda_{12}(t) + \lambda_{15}(t)) & \lambda_{12}(t) & 0 & 0 & \lambda_{15}(t) \\ \lambda_{21}(t) & -(\lambda_{21}(t) + \lambda_{25}(t)) & 0 & 0 & \lambda_{25}(t) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5}$$

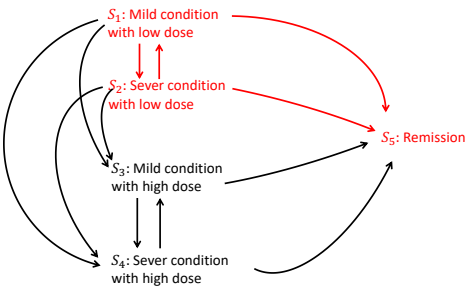
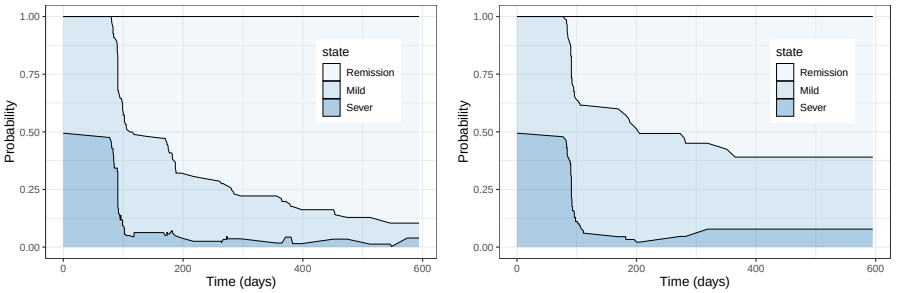


Figure 8: Multi-state diagram for TACERA study.

Parametric g-formula



(a) $\tilde{P}^{\overline{1}}(0,t)$.

(b) $\tilde{P}^{\overline{0}}(0,t)$.

Figure 9: The estimated transition probabilities by g-formula with the multi-states model using `mstate` R package.

Parametric g-formula

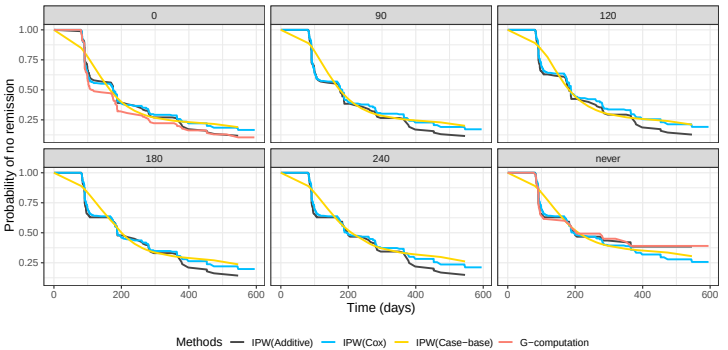


Figure 10: Comparing survival curves estimated by different methods for different treatment initiation times.

SNFTMs

Let Y, C denote time to remission and censoring.

- Idea: model the effect of skipping the "last" treatment given the past
  - baseline treatment: $a = 0$
  - at t, we consider removing an "instantaneous blip" – that is removing the treatment $a(t)$ – on $[t, t+h)$ with $h > 0$ and $h \downarrow 0$. We contrast two treatment regimes:

$$\overline{a}_{[0,t+h)} = (\overline{a}((t+h)-), \underline{0}(t+h)) \tag{6}$$
$$\overline{a}_{[0,t)} = (\overline{a}(t-), \underline{0}(t)) \tag{7}$$

    with $h \downarrow 0$.
  - SNMs model the infinitesimal effect of the treatment given in $[t, t+h)$ as $h \downarrow 0$ by comparing the following counterfactuals in distribution:

$$Y^{\overline{a}_{[0,t+h)}} | \overline{A}(t) = \overline{a}(t), \overline{L}(t), B \tag{8}$$
$$Y^{\overline{a}_{[0,t)}} | \overline{A}(t) = \overline{a}(t), \overline{L}(t), B. \tag{9}$$

- Model the effect of skipping $a(t)$ at t given the past, define the distributional relationship

$$F_{Y^{\overline{a}_{[0,t+h)}} | \overline{A}(t)=\overline{a}(t), \overline{L}(t), B}(y_{t+h}) = F_{Y^{\overline{a}_{[0,t)}} | \overline{A}(t)=\overline{a}(t), \overline{L}(t), B}(y_t) \tag{10}$$

SNFTMs

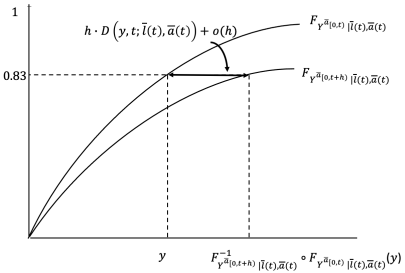


Figure 11: Illustration of D [8]

$$\frac{y_{t+h} - y_t}{h} = \frac{F^{-1}_{Y^{\bar{a}[0,t+h)}|\overline{A}(t)=\bar{a}(t),\overline{L}(t),B} \circ F_{Y^{\bar{a}[0,t)}|\overline{A}(t)=\bar{a}(t),\overline{L}(t),B}(y_t) - y_t}{h} \tag{11}$$

which, with $h \downarrow 0$, can be formulated as

$$\mathcal{D}(y, t; \overline{A}(t), \overline{L}(t), B) = \frac{\partial}{\partial h}\bigg|_{h=0} \left(F^{-1}_{Y^{\overline{A}[0,t+h)}|\overline{A}(t),\overline{L}(t),B} \circ F_{Y^{\overline{A}[0,t)}|\overline{A}(t),\overline{L}(t),B} \right)(y). \tag{12}$$

This is the infinitesimal shift function.

SNFTMs

Define a continuous variable $\mathcal{Y}(t)$ that mimics $Y^{\overline{a}[0,t)}$:

$$\mathcal{Y}(t) \sim Y^{\overline{a}[0,t)} | \overline{A}(t), \overline{L}(t), B$$

$$\frac{d\mathcal{Y}(t)}{dt} = \mathcal{D}(\mathcal{Y}(t), t; \overline{A}(t), \overline{L}(t), B)$$

with $\mathcal{Y}(T) = Y$.

For survival outcomes, [9] suggested

$$\mathcal{D}_\psi(y, t; \overline{A}(t), \overline{L}(t), B) = (1 - \exp(\gamma_\psi(y, t; \overline{A}(t), \overline{L}(t), B)))\mathbb{I}_{y>t} \quad (13)$$

The form of (13) ensures $D < 1$ that avoids infinite survival time.

$$Y^{\overline{0}} \sim \mathcal{Y}(0) = \int_0^Y \exp(\psi A(s)) ds. \quad (14)$$

In this special case, the residual lifetime had treatment been stopped at t given past observations has the same distribution as the residual lifetime had treatment being continued till the end of study multiplied by a factor $\exp(\psi)$.

- Problem of artificial censoring: $C_\psi = \int_0^C \exp(\psi A(t)) dt$, let $X_\psi = \min\{C_\psi, Y_\psi\}$, $X_\psi = \int_0^X \exp(\psi A(t)) dt$.
- if counterfactual failure comes after counterfactual censoring for individuals who are observed to fail, then this raises the problem of artificial censoring where observed failures are artificially censored $\rightarrow$ information loss, nonsmoothness in estimation function.

G-estimation

- Sequentially "blipping off" treatment backwards $\Longrightarrow$ a random variable $\mathcal{X}(0)$ *mimicking* $X^{\overline{0}}$
- Fitting a model for $P_\theta(A(t)|\overline{L}(t), \overline{A}(t-), \mathcal{X}_\psi)$ with parameters θ, under no unmeasured confounding assumption, the coefficient of $\mathcal{X}_\psi$ should 0. – Grid search or solving estimating equation.
- An unbiased estimating equation can be constructed to estimate ψ. Martingale theory guides the formulation of such an estimating equation by defining a $\sigma(\overline{A}(t), \overline{Z}(t), \overline{Y}^{\overline{A}_{[0,t)}})$-predictable process $h_t(Y^{\overline{A}_{[0,t-)}}, \overline{A}(t-), \overline{Z}(t-))$ with regularity conditions.

$$P_n \int_0^\tau h_t(\mathcal{X}_\psi(t), \overline{A}(t), \overline{L}(t), B)(dN^A(t) - \lambda_\theta^A(t)dt) = 0, \tag{15}$$

  where $P_n X = 1/n \sum_{i=1}^n X_i$ is the empirical measure.
- Doubly robust estimator see [18][4]
- The causal parameter $\exp(\psi)$ has a closed form.

4

$$\mathbb{E}\left\{ \frac{\Delta}{P(C > X|\overline{Z}(X), \overline{A}(X))} \int_0^\tau c^{\text{eff}}(t, \overline{Z}(t), \overline{A}(t-)) \left[X_\psi - \mathbb{E}[X_\psi|\overline{Z}(t), \overline{A}(t-) = 0, V \geq t] \right] dM^A(t) \right\} = 0, \tag{16}$$

where $\Delta = \mathbb{I}(Y > C)$, $V = \min\{T^A, C, Y\}$ and $\{V \geq t\} = \{T^A \geq t, C \geq t, Y \geq t\}$.

SNFTMs and G-estimation

- Fitting a Cox-PH model for the hazard of treatment initiation: $h^A(t) = h_0^A(t)\exp(\beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 L(t))$, then derive the martingale via $d\hat{M}^A(t) = dN^A(t) - \hat{h}^A(t)G^A(t)$. Secondly, fit a Cox-PH model for the censoring process to estimate the hazard of censoring: $h^C(t) = h_0^C(t)\exp(\nu_0 + \nu_1 B_1 + \nu_2 B_2 + \nu_3 L(0))$ and estimate $\hat{P}(C > X\overline{Z}(X), \overline{A}(X)) = \exp(-\int_0^X \hat{h}^C(t)G^C(t)dt)$.
- $\exp(\hat{\psi}) = 1.388$,p-value$= 0.025$. Time to remission would be prolonged by a factor of 1.388 had the treatment not initiated compared to initiated from the beginning of follow-up. Note that IPCWs appear to be unstable. Extremely large weights frequently appear when estimating the parameter for each bootstrapped sample. We truncate the censoring weights larger than 10.

Summary

- A conceptual comparison:

	IPW-MSMs	SNFTMs& g-estimation	Parametric g-formula
Interpretability	✓✓	✓	✓
Software available	✓✓	✓	✗
Existing application	✓✓	✓	✗
Robust to model misspecification	✓	✓	✗
Robust to violation of positivity	✗	✓✓	✗
Censoring	✓✓	✓	✗

Table 3: Comparison of three methods in continuous-time setting with time-varying treatment and time-varying confounding. Strength is ranked from high to low, indicated by ✓✓,✓ respectively. ✗ labels the weakness or non-published area.

- A quantitative comparison is rarely studied.
- Gaps:
  - simulation [16];
  - complicated treatment processes;
  - issues of positivity assumption violation, artificial censoring.

References I

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding.
Statistical models based on counting processes.
Springer Science & Business Media, 2012.

[2] R.-M. Consortium.
Characterization of disease course and remission in early seropositive rheumatoid arthritis: results from the tacera longitudinal cohort study.
Therapeutic advances in musculoskeletal disease, 13:1759720X211043977, 2021.

[3] V. Didelez.
Graphical models for marked point processes based on local independence.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(1):245–264, 2008.

[4] V. Didelez.
Asymmetric separation for local independence graphs.
arXiv preprint arXiv:1206.6841, 2012.

[5] Y. Dong.
Continuous-time marginal structural models for adverse drug effects in pharmacoepidemiology.
University of Toronto (Canada), 2021.

[6] S. Ferreira Guerra, M. E. Schnitzer, A. Forget, and L. Blais.
Impact of discretization of the timeline for longitudinal causal inference methods.
Statistics in medicine, 39(27):4069–4085, 2020.

[7] M. A. Hernan and J. M. Robins.
Causal Inference: What If.
Chapman Hall/CRC, 2020.

[8] J. J. Lok.
Statistical modeling of causal effects in continuous time.
2008.

[9] J. M. Robins.
Structural nested failure time models.
Wiley StatsRef: statistics reference online, 2014.

[10] K. Røysland.
A martingale approach to continuous-time marginal structural models.
2011.

[11] K. Røysland, P. Ryalen, M. Nygård, and V. Didelez.
Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses.
arXiv preprint arXiv:2202.02311, 2022.

References II

[12] P. C. Ryalen, M. J. Stensrud, S. Fosså, and K. Røysland.
Causal inference in continuous time: an example on prostate cancer therapy.
Biostatistics, 21(1):172–185, 2020.

[13] P. C. Ryalen, M. J. Stensrud, and K. Røysland.
The additive hazard estimator is consistent for continuous-time marginal structural models.
Lifetime data analysis, 25:611–638, 2019.

[14] O. Saarela.
A case-base sampling method for estimating recurrent event intensities.
Lifetime data analysis, 22:589–605, 2016.

[15] O. Saarela and Z. Liu.
A flexible parametric approach for estimating continuous-time inverse probability of treatment and censoring weights.
Statistics in medicine, 35(23):4238–4251, 2016.

[16] S. R. Seaman and R. H. Keogh.
Simulating data from marginal structural models for a survival time outcome.
arXiv preprint arXiv:2309.05025, 2023.

[17] J. A. Wellner.
A heavy censoring limit theorem for the product limit estimator.
The Annals of Statistics, pages 150–162, 1985.

[18] S. Yang, K. Pieper, and F. Cools.
Semiparametric estimation of structural failure time models in continuous-time processes.
Biometrika, 107(1):123–136, 2020.

[19] A. Ying.
Causal inference for complex continuous-time longitudinal studies.
arXiv preprint arXiv:2206.12525, 2022.