Proceedings of Machine Learning Research vol 236:665–687, 2024 3rd Conference on Causal Learning and Reasoning

Causality for Functional Longitudinal Data

Andrew Ying AYING9339@GMAIL.COM
Irvine, CA, U.S.A.

Editors: Francesco Locatello and Vanessa Didelez

Abstract

"Treatment-confounder feedback" is the central complication to resolve in longitudinal studies, to infer causality. The existing frameworks of identifying causal effects for longitudinal studies with repeated measures hinge heavily on assuming that time advances in discrete time steps or data change as a jumping process, rendering the number of "feedbacks" finite. However, medical studies nowadays with real-time monitoring involve functional time-varying outcomes, treatment, and confounders, which leads to an uncountably infinite number of "feedbacks". Therefore more general and advanced theory is needed. We generalize the definition of causal effects under user-specified stochastic treatment regimes to functional longitudinal studies with continuous monitoring and develop an identification framework for a end-of-study outcome. We provide sufficient identification assumptions including a generalized consistency assumption, a sequential randomization assumption, a positivity assumption, and a novel "intervenable" assumption designed for the continuous-time case. Under these assumptions, we propose a g-computation process and an inverse probability weighting process, which suggest a g-computation formula and an inverse probability weighting formula for identification. For practical purposes, we also construct two classes of population estimating equations to identify these two processes, respectively, which further suggest a doubly robust identification formula with extra robustness against process misspecification.

Keywords: Causal Inference; Stochastic Process; Panel Data; Functional Data; Continuous Time.

1. Introduction

Causality addresses the definition of quantifiable causal relationships and the undertaking of valid causal inferences. While double-blinded randomized controlled trials stand as the gold standard for both aspects, practical considerations, ethical concerns, and costs often lead statisticians towards observational studies. Various causal frameworks have been developed, including potential outcomes (Neyman, 1923; Rubin, 1974; Holland, 1986), graphical theory (Dawid, 1979; Lauritzen and Wermuth, 1989; Cox and Wermuth, 2014), structural equation models (Jöreskog, 1978; Pearl, 2009), dynamical models (Commenges and Gégout-Petit, 2009), and decision-theoretic frameworks (Dawid, 2000; Geneletti, 2005). A common thread across these approaches, in our understanding, is the principle of physical causality, which asserts that future events cannot influence past events.

Complex longitudinal studies, involving panel data or repeated measures, where treatment, confounders, and sometimes the outcome evolve over time, present formidable challenges to valid causal inference. This is due to the influence of past confounders, including previous outcomes, on past treatment allocation, which subsequently impacts the current state of patients and the progression of the disease. This phenomenon, referred to as "treatment-confounder feedback" (Hernán and Robins, 2020), renders traditional adjustment methods, such as regression with a history of treatment and confounders, ineffective. Intervening on past treatment values necessitates changes in future confounders, which, in turn, are conditioned upon for confounding correction.

Noteworthy contributions have been made by Greenland and Robins (1986); Robins (1986, 1987, 1989, 1997, 1998) in developing extensive theories for causal inference in complex longitudinal studies. However, these methods are limited to well-structured longitudinal data with predetermined and fixed discrete time steps, which may not accurately represent the random nature of visit times. We provisionally refer to these as "regular longitudinal data (RLD)" and "regular longitudinal studies."

In contrast, Lok (2001); Johnson and Tsiatis (2005); Røysland (2011, 2012); Hu and Hogan (2019); Rytgaard et al. (2022); Yang (2022); Røysland et al. (2022) have made strides in formally establishing identification frameworks that allow for continuous time steps. However, their assumptions hinge on treatment and confounder processes with stepwise paths featuring finite jumps, akin to point processes. For instance, in pharmacoeconomics studies as considered in Rytgaard et al. (2022), each patient is assumed to have a random finite number of visits at random time points, during which they may alter their treatment and certain confounders. This approach accommodates longitudinal data with irregular, randomly distributed visit times. We refer to this category as "irregular longitudinal data (ILD)" and "irregular longitudinal studies."

In contemporary medical studies, there is a growing need for more comprehensive and advanced causal inference theories tailored for longitudinal data, particularly when confounders and treatments are continuously measured over time. For instance, in intensive care and in-patient settings, vital status is continually monitored as part of standard practice (Johnson et al., 2016, 2018). Additionally, the advent of wearable devices has led to the increasing use of real-time monitoring for the long-term management of chronic diseases, such as continuous glucose monitoring for diabetes (Mastrototaro, 2000; Klonoff, 2005; Rodbard, 2016). As healthcare providers increasingly rely on real-time monitoring reports for disease management decisions, real-time monitoring may introduce confounding factors in longitudinal studies. Notably, real-time monitoring generates functional data, which may be observed discretely over time, but the underlying causal mechanism operates continuously. This leads to an infinite number of treatment-confounder feedback loops across the timeline and the absence of a joint density. We informally refer to this as "functional longitudinal data (FLD)" and "functional longitudinal studies." This presents a substantial challenge to existing causal inference methodologies.

As previously discussed, these three types of longitudinal data not only differ in their sources, motivations, and backgrounds, but also in their inherent nature. See Figure 1 for a comparison between possible simulated realizations of a stochastic process across three scenarios: fixed visit time frames (Greenland and Robins, 1986; Robins, 1986, 1987, 1989, 1997, 1998), irregular visit time frames (Lok, 2001; Johnson and Tsiatis, 2005; Røysland, 2011, 2012; Hu and Hogan, 2019; Rytgaard et al., 2022; Yang, 2022; Røysland et al., 2022), and real-time monitoring data. Regardless of the type of variable (binary, categorical, or continuous), both time and path are finite for RLD. In the case of ILD, time is allowed to vary infinitely, with the random value subject to change at any time point. However, the jumps of a path remain finite. FLD are truly continuous, as both time and path are permitted to vary infinitely. It is evident that both RLD and ILD are special cases of FLD.

Given the growing needs of rigorous causality theory around FLD, the fact that current methodologies designed for RLD or ILD all fail for FLD, and rare development of methods around FLD, this paper seeks to develop a novel causal framework for FLD. This paper presents an initial exploration into this topic. It serves as a precursor to a more comprehensive study, which is being developed for a detailed journal paper. The journal paper will extend the discussions and findings presented here, incorporating a wider perspective on its generality, nonparametric property,

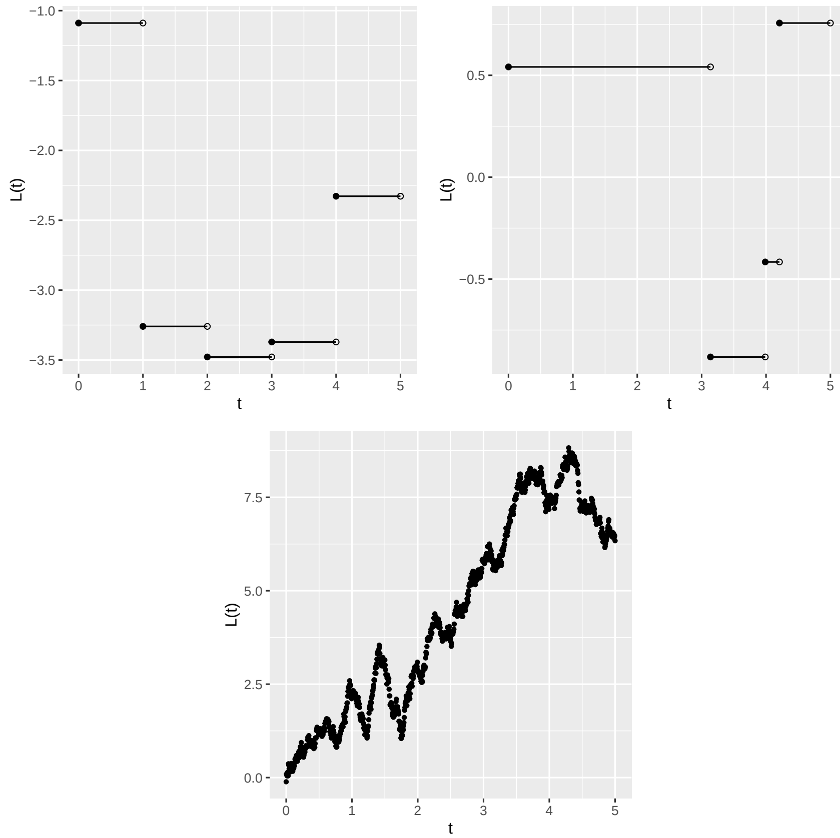

Figure 1: Possible realizations from three type of longitudinal data. The top left is from a regular longitudinal study with fixed visit times. The top right is from a irregular longitudinal study with irregular and random visit times. The bottom one is from a functional longitudinal study with real-time monitoring data.

and accommodating time-varying outcome of interest, mortality and right censoring, as typically complications in longitudinal studies.

We first give an intensive literature review over RLD, ILD and FLD. This helps the readers to understand the common standard and structure of causal inference within these settings, and the research upfront of these directions. With the dimensionality challenge posed by the temporal aspect in longitudinal data, it is intuitive to consider strategies for dimension reduction in statistical inference. This leads to two primary directions: model-based methods and estimand-based methods. The key distinction lies in whether the identification of the parameter of interest is nonparametric. Model-based methods typically focus on a parameter that indexes a specific model, be it parametric or semiparametric. In contrast, estimand-based methods target simpler parameters, such as marginal means, without imposing restrictions on the underlying distribution. For a comprehensive discussion on these statistical strategies, see Vansteelandt and Dukes (2022). In this draft, we proceed with estimand-based methods and review the literature on causal inference for three types of longitudinal data, placing a particular emphasis on estimand-based approaches.

In the case of RLD, a joint density of the random vectors exists, allowing statisticians to potentially intervene in the treatment process by substituting the treatment density with another density. Within this context, well-established methods for nonparametrically inferring the marginal distribution of potential outcomes include the g-computation formula (Greenland and Robins, 1986), inverse probability weighting (IPW) estimators (Rosenbaum and Rubin, 1983; Hernán et al., 2000, 2001, 2002), and doubly robust estimators, all designed to correct for potential time-varying confounding. Additionally, there exist model-based frameworks such as marginal structural models (Robins, 1998; Ying et al., 2023) and structural nested models (Vansteelandt and Joffe, 2014).

For ILD, Lok (2001); Johnson and Tsiatis (2005); Røysland (2011, 2012); Hu and Hogan (2019); Yang (2022); Røysland et al. (2022) have proposed and developed structural models. More recently, Rytgaard et al. (2022) thoroughly characterized assumptions and identification strategies by leveraging marked point process theories, with a focus on marginal parameters. In ILD, the point process under consideration is the simplest type of continuous-time stochastic process, for which a Riemann-Stieltjes measure exists pathwise.

Data originating from either fixed visit time frames or irregular random visit time frames exhibit a finite number of treatment-confounder feedback loops and admit a joint density, significantly simplifying theoretical and computational complexities. Due to these characteristics, while there may be some distinctions between the two frameworks in terms of data characteristics, they are not fundamentally distinct in their underlying mathematical theories. Both frameworks draw upon similar mathematical concepts and principles, and share many of the same foundational assumptions. Most importantly, a joint likelihood exists for both RLD and ILD. In contrast, no likelihood but only probability measure exists for FLD, which hence necessitates much more advanced mathematical theory to be effectively analyzed. Existing frameworks tailored for RLD or ILD cannot be straightforwardly applied to FLD. Within the potential outcomes framework, the causal quantity of interest is scarcely well defined for functional data, let alone identification procedures and associated plausible identification assumptions. Notable exceptions that have considered FLD include continuous-discrete state-space models (Singer, 2008), dynamic models (Commenges and Gégout-Petit, 2009), and structural rank-preserving models (Sun and Crawford, 2022). However, these approaches are model-based and typically rely on complex statistical models with low-dimensional parameters indexing an uncountable number of potential functional data points through stochastic differential equations, potentially leading to unrealistic restrictions on the observed data. Our framework, in contrast, commences by postulating an estimand rather than a statistical model. Our assumptions are purely for causal interpretation, imposing no restrictions on the observed data. This aligns with recent efforts to provide assumption-free causal inference (van der Laan, 2010; Vansteelandt and Dukes, 2022).

There is also literature on causal inference dealing with functional treatment or functional covariates in point observational studies (Miao et al., 2020; Zhang et al., 2021; Tan et al., 2022). The data format under consideration aligns with our setting. However, our paper distinguishes itself by addressing the temporal dimension introduced by longitudinal studies, whereas these studies primarily focus on point exposure. As a result, in principle, if we concentrate on the same quantity and disregard the treatment-confounder feedback, our paper should extend their identification results.

In this paper, we leverage stochastic process theory and measure theory (Bhattacharya and Waymire, 2007; Durrett, 2019) to establish a novel estimand-based causal framework for longitudinal studies where confounders and treatments are continuously measured over time under a stochastic treatment regime. This framework encompasses causal interpretation and identification, operating within the potential outcome (or counterfactual) framework (Neyman, 1923; Rubin, 1974; Holland, 1986). Unlike existing literature (Greenland and Robins, 1986; Robins, 1986, 1987, 1989, 1997, 1998) or (Lok, 2004, 2008; Røysland, 2011; Rytgaard et al., 2022), our framework does not impose path restrictions on confounder, treatment, and outcome processes, except for them being càdlàg. The framework is built on assumptions including a generalized consistency assumption, a generalized sequential randomization assumption, and a generalized positivity assumption, typically required in RLD and ILD for identification. In addition, we introduce a novel "intervenable" assumption tailored for the continuous case, designed to ensure compatibility between the target treatment regime and the observed data distribution. While this set of assumptions is sufficient for identification, it may be challenging to interpret. To facilitate a more straightforward understanding, we present an additional set of stronger assumptions based on generalized "coarsening at random" assumption (Heitjan and Rubin, 1991). We define a g-computation formula through a g-computation process, and an inverse probability weighting (IPW) formula through an IPW process. Notably, our framework provides a unified comprehension of g-computation formula and IPW formula in terms of projections and Radon-Nikodym derivatives over filtration, respectively. For practical applications, we construct two classes of population estimating equations that identify these two processes, respectively. These population estimating equations also suggest a doubly robust formula, providing additional safeguarding against process misspecification.

Our work establishes a rigorous framework for defining the estimand in causal inference for FLD, addressing a significant limitation of existing approaches. Unlike existing frameworks, ours can handle longitudinal data where both time and state space (the space of càdlàg paths) are uncountably infinite. Because of our path assumption being only càdlàg, our framework naturally unifies identification theory for longitudinal data encompassing both discretely repeated measures and continuous monitoring. Also, our IPW characterization is particularly valuable in generalizing the propensity score, a challenging task given the typical ill-posed nature of propensity scores with functional data, where probability density functions for random functions are often non-existent. Our framework also immediately implies identification for RLD and FLD when at each observation, functional data were to be recorded, which has only been developed for point observation studies (Miao et al., 2020; Zhang et al., 2021; Tan et al., 2022). To our best knowledge, this paper is the first attempt of building an estimand-based framework without any modeling or path assumptions.

2. Causal Identification in Functional Longitudinal Studies

Suppose that there is a functional longitudinal study during 0 to τ, where $\tau > 0$ is a positive constant representing the end of the study.

- Define $(\Omega, \mathscr{F}, \mathbb{P})$ as the underlying sample space, σ-algebra, and probability measure.
- $A(t)$ and $L(t)$ are a treatment received and measured confounders at time t. The treatment $A(t)$ at time t can be binary, categorical, continuous or even itself functional. $L(t)$ is also allowed to take any form. We write $\bar{A}(t) = \{A(s) : 0 \leq s \leq t\}$ and $\bar{L}(t) = \{L(s) : 0 \leq s \leq t\}$. We abbreviate $\bar{A} = \{A(s) : 0 \leq s \leq \tau\}$ and $\bar{L} = \{L(s) : 0 \leq s \leq \tau\}$.
- A subset of $Y \subset L(\tau)$ measured at the end of study is considered as the outcome of interest, similar to Choi et al. (2002); Ying et al. (2023).
- $\mathcal{A}$ is the set of all possible values of $\bar{a}$.

- Also, write the counterfactual outcome $Y_{\bar{a}}$ and counterfactual confounders $L_{\bar{a}}(t)$, for any $\bar{a} \in \mathcal{A}$ and any $t \in [0, \tau]$, as the outcome and confounders if the treatment were to set as $\bar{a}$.

- When there is no confusion, we may write $\mathbb{P}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})$ to represent the distribution on the path space induced by the stochastic processes and the probability measure $\mathbb{P}$ on the sample space Ω. Note that this is not a density or a likelihood. This notation is well adopted by probabilists (Bhattacharya and Waymire, 2007; Durrett, 2019) and also statisticians (Gill and Robins, 2001).

- Define $\mathscr{F}_t = \sigma\{A(s), L(s) : \forall s \leq t\}$ as a filtration of information collected before time t. Also we write $\mathscr{F}_{t-} = \sigma(\cup_{0 \leq s < t} \mathscr{F}_t)$ and $\mathscr{G}_t = \sigma(\mathscr{F}_{t-}, A(t))$. We define $\mathscr{G}_{\tau+} = \mathscr{F}_\tau$. We write $\mathscr{F}_{0-}$ and $\mathscr{G}_{0-}$ as the trivial sigma algebra for convenience.

- We assume the event space is Polish so that conditional probability can be chosen to be regular. We understand conditional distribution as a function over a sigma algebra multiplied with a path set. For instance, $\mathbb{P}(\mathrm{d}\bar{a}|\bar{l})$ can be seen as a function over Borel space generated by $\{\bar{a}\}$ and the path set $\{\bar{l}\}$. Importantly, conditional distribution is defined almost surely and one needs to take extra caution when replacing and intervening treatment distributions when conducting causal inference.

- We use the upper case for random variables and the lower case for their realized values. An important caveat is that conditional probability is only uniquely defined almost surely. Therefore throughout this draft, otherwise stated, for the measure zero subset where the conditional probability is not uniquely defined, we set the conditional probability to be zero.

- A partition $\Delta_K[0, \tau]$ on $[0, \tau]$ is a finite sequence of $K + 1$ numbers of the form $0 = t_0 < \cdots < t_K = \tau$. The mesh $|\Delta_K[0, \tau]|$ of a partition $\Delta_K[0, \tau]$ is $\max_{i=0,\cdots,K-1}(t_{j+1} - t_j)$, representing the maximum gap length of the partition.

- We use $\|\cdot\|_{\mathrm{TV}}$ to represent the total variation norm over the space of signed measures of the path space, which is a Banach space.

Here are two real-life examples:

Example 1 (Intensive care unit) *MIMIC-III ('Medical Information Mart for Intensive Care') is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital (Johnson et al., 2016, 2018). The timing and duration of treatment are important concepts for researchers seeking to understand issues that relate to the intensity of an administered intervention. For instance, $A(t)$ can be antibiotics usage at time t, and $L(t)$ may include the severity of illness scores, immediate vital signs, laboratory values, blood gas values, urine output, weight, height, age, gender, service type, total fluid intake, total fluid output, etc, at time t. The user-specified treatment regime can be chosen to understand the effect of antibiotics on certain illnesses progression Y.*

Example 2 (Continuous glucose monitoring) *Continuous glucose monitoring (CGM) provides information unattainable by intermittent capillary blood glucose (Rodbard, 2016). $A(t)$ typically is insulin dosage at time t, and $L(t)$ may include glucose level immediate changes in behaviors such as diet, medications, physical activity, etc, at time t. The user-specified treatment regime can be chosen* *to understand the effect of insulin dosage and usage frequency on the glucose level Y measured at the end.*

Suppose one is interested in evaluating the average treatment effect of one treatment regime versus another one. One way that he/she could do this is to identify the mean of some transformation of the potential outcome under both treatment regimes and compare them. As an illustration, throughout the draft, suppose we are interested in learning a marginal mean of a transformed potential outcome under a user-specified treatment regime,

$$\int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}), \tag{1}$$

where ν is a user-specified function and $\mathbb{G}$ is a priori defined (signed) measure on $\mathcal{A}$ and we assume $\mathbb{E}\{\nu(Y_{\bar{a}})\}$ is integrable against $\mathbb{G}$. This estimand include those considering the marginal mean of outcomes under a static treatment regime, either deterministic (Robins, 1997; Sun and Crawford, 2022) or stochastic (Cain et al., 2010; Young et al., 2011). We list some examples of the intervention $\mathbb{G}$ for FLD below:

- When the causal outcome under a specific regime $\bar{a}$ is of interest, for instance, all patient were under treatment, the point mass (delta) measure $\mathbb{G} = \mathbb{1}(\bar{A} = \bar{a})$ can be considered. If a contrast between a specific regime $\bar{a}$ and the control $\bar{0}$ needs investigation, then $\mathbb{G} = \mathbb{1}(\bar{A} = \bar{a}) - \mathbb{1}(\bar{A} = \bar{0})$.
- Though the data are allowed to be functional and the underlying data generating mechanism can have uncountably infinite number of treatment-confounder feedbacks, a finite-dimensional distribution intervention can still be considered, for example, intervening dosage of certain drug hourly or daily.
- Likewise, a marked point process considered in Rytgaard et al. (2022) measure represents intervening both dosage and frequency of usage for certain drugs.
- If considering certain fluid intake that is continuously used, one might leverage stationary process measure that allows noise of fluid usage yet conforms to time regularity. One may also consider continuous Gaussian process (including Wiener measure, also known as Brownian motion) as a typical example considered in stochastic processes.

2.1. Weak identification assumptions

The paramount principle of adjusting for "treatment-confounders" feedbacks and creating a pseudo-population under which treatment distribution were to follow $\mathbb{G}$ is intervening the treatment distribution for each time point iteratively. This principle is the central idea for literature on RLD and ILD. However, given the uncountably infinite number of feedbacks "treatment-confounders" in FLD, such adjustment becomes impossible. Instead, we adopt a net convergence idea (like in the definition of the Riemannian integral) to overcome this complication.

For any sequences of partitions $\{\Delta_K[0,\tau]\}_{K=1}^{\infty}$ with $|\Delta_K[0,\tau]| \to 0$ as $K \to \infty$, we have the following decomposition

$$\mathbb{P}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) = \prod_{j=0}^{K} \left[\mathbb{P}\{\mathrm{d}\bar{l}(t_j)|\bar{a}(t_j), \bar{l}(t_{j-1})\}\, \mathbb{P}\{\mathrm{d}\bar{a}(t_j)|\bar{a}(t_{j-1}), \bar{l}(t_{j-1})\}\right].$$

We define

$$\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) = \prod_{j=0}^{K} \left[\mathbb{P}\{\mathrm{d}\bar{l}(t_j)|\bar{a}(t_j), \bar{l}(t_{j-1})\}\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{a}(t_{j-1})\}\right].$$

Assumption 1 (Intervenable) *The measures $\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})$ converges to the same (signed) measure in the total variation distance on the path space, regardless of the choices of partitions, in which case we may write the limit as $\mathbb{P}_{\mathbb{G}} = \mathbb{P}_{\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})$. That is, $\| \mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}} - \mathbb{P}_{\mathbb{G}} \|_{\mathrm{TV}} \to 0$, as $|\Delta_K[0, \tau]| \to 0$. We call $\mathbb{P}_{\mathbb{G}}$ the target distribution.*

We note that Assumption 1 is a novel assumption imposed uniquely for the continuous framework yet holds trivially for discrete-time cases. Intuitively, Assumption 1 claims any sequence of discrete-time studies represented by a sequence of partitions that approximates our continuous study, the substitution in probability measures approximates the same limit, regardless of the choices of partitions on time. We provide plausible sufficient assumption based on a generalized "coarsening at random" assumption (Heitjan and Rubin, 1991) in Section 2.3.

Next, we impose the positivity assumption.

Assumption 2 (Positivity) *The target distribution $\mathbb{P}_{\mathbb{G}}$ induced by the target regime $\mathbb{G}$ is absolutely continuous against $\mathbb{P}$, that is, $\mathbb{P}_{\mathbb{G}} \ll \mathbb{P}$, where we may write $\frac{\mathrm{d}\,\mathbb{P}_{\mathbb{G}}}{\mathrm{d}\,\mathbb{P}} = \frac{\mathrm{d}\,\mathbb{P}_{\mathbb{G}}}{\mathrm{d}\,\mathbb{P}}(\bar{a}, \bar{l})$ as the corresponding Radon-Nikodym derivative.*

This assumption implies that there are enough data to infer the expectation $\mathbb{E}_{\mathbb{G}}(Y)$ of the outcome Y under the target distribution, where $\mathbb{E}_{\mathbb{G}}$ is the mathematical expectation under $\mathbb{P}_{\mathbb{G}}$. The following generalized consistency assumption and sequential randomization assumption ensures us to show that $\mathbb{E}_{\mathbb{G}}(Y)$ has a causal interpretation, that is, it equals the target parameter (1).

Assumption 3 (Consistency) *$Y = Y_{\bar{A}}$, almost surely.*

Like in the discrete case, the consistency assumption links the observed outcome and the potential outcome via the treatment actually received. It says that if an individual receives the treatment $\bar{A} = \bar{a}$, then his/her observed outcome Y matches $Y_{\bar{a}}$. Before stating the last assumption, we define $\|\cdot\|_p$, $\|\cdot\|_{p,\mathbb{G}}$ as the L^p distance and we denote $L^p(\mathbb{P})$ as the space of p-th order integrable functions over $\mathbb{P}$ and $\mathbb{P}_{\mathbb{G}}$.

Assumption 4 (Continuous-time randomization) *There exists a bounded function $\varepsilon(t, \eta) > 0$ with $\int_0^\tau \varepsilon(t, \eta)dt \to 0$ as $\eta \to 0$, such that for any $\bar{a} \in \mathcal{A}$, $t \in [0, \tau]$, $\eta > 0$,*

$$\left\|\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{A}(t), \bar{L}(t)\} - \mathbb{E}\{\nu(Y_{\bar{a}})|\bar{A}(t+\eta), \bar{L}(t)\}\right\|_1 < \varepsilon(t, \eta).$$

Assumption 4 states that given the information history $\mathcal{F}_t$ up to time t, the expectation of $\nu(Y_{\bar{a}})$ merely depends on a short period of treatment assignment between $(t, t+\eta]$. Indeed, such dependence is upper bounded by this bounded function $\varepsilon(t, \eta)$ whose integral over $t \in (0, \tau)$ tends to zero, as the gap η goes to zero. The condition $\int_0^\tau \varepsilon(t, \eta)dt \to 0$ is needed to handle longitudinal studies on time range $[0, \tau]$. The assumption is likely to hold if the treatment assignment in a short amount of time $(t, t+\eta]$ purely depends on the history of treatment and covariates.

2.2. Identification formulas

Definition 1 (G-computation process) *Under Assumptions 1 and 2, define*

$$H_{\mathbb{G}}(t) = \mathbb{E}_{\mathbb{G}}\{\nu(Y)|\mathscr{G}_t\},$$

as a projection process, which is apparently a $\mathbb{P}_{\mathbb{G}}$-martingale. We call $H_{\mathbb{G}}(t)$ the g-computation process. Note that $H_{\mathbb{G}}(\tau) = \mathbb{E}_{\mathbb{G}}\{\nu(Y)|\mathscr{G}_{\tau+}\} = \nu(Y)$ and $H_{\mathbb{G}}(0-) = \mathbb{E}_{\mathbb{G}}\{\nu(Y)\}$.

The g-computation process intuitively serves as a consecutive adjustment of the target Y from τ to 0.

Theorem 1 (G-computation formula) *Under Assumptions 1 - 4, (1) is identified via a g-computation formula as*

$$\int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}) = H_{\mathbb{G}}(0-).$$

Definition 2 (Inverse probability treatment weighting process) *Under Assumptions 1 and 2, define*

$$Q_{\mathbb{G}}(t) = \mathbb{E}\left\{\frac{\mathrm{d}\,\mathbb{P}_{\mathbb{G}}}{\mathrm{d}\,\mathbb{P}}\middle|\bar{A}(t), \bar{L}(t)\right\},$$

as a Radon-Nikodym derivative at time t, which is apparently a $\mathbb{P}$-martingale. We call $Q_{\mathbb{G}}(t)$ the inverse probability weighting (IPW) process. Note that $Q_{\mathbb{G}}(\tau) = Q_{\mathbb{G}}(\tau+) = \frac{\mathrm{d}\,\mathbb{P}_{\mathbb{G}}}{\mathrm{d}\,\mathbb{P}}(\bar{A}, \bar{L})$ and $Q_{\mathbb{G}}(0-) = 1$.

The IPW process intuitively serves as a continuous adjustment of the treatment process $\bar{A}$ from 0 to τ.

Theorem 2 (Inverse probability treatment weighting formula) *Under Assumptions 1 - 4, (1) is identified via an inverse probability weighting formula as*

$$\int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}) = \mathbb{E}\left\{Q_{\mathbb{G}}(\tau)\nu(Y)\right\}.$$

We have defined two processes $H_{\mathbb{G}}(t)$, $Q_{\mathbb{G}}(t)$ in Definitions 1, 2, which provided two identification formulas for (1) in Theorems 1, 2. However, such definitions of $H_{\mathbb{G}}(t)$, $Q_{\mathbb{G}}(t)$ via projections and Radon-Nikodym derivatives are not useful for identifying themselves because recursive identification approaches across time for the discrete-time version g-computation process and IPW process cannot be utilized for FLD. Furthermore, unlike a density, a Borel measure cannot be learned from data, to our best knowledge. In this subsection, we establish identification (also can be seen as alternative definitions) of both processes via two classes of population estimating equations which can be more useful for practice and may inspire Z-estimators (Ghassami et al., 2022) or M-estimators (Kompa et al., 2022) in the future. Identification for $H_{\mathbb{G}}(t)$ and $Q_{\mathbb{G}}(t)$ also turns out to be useful for constructing the doubly robust formula later.

For any two $\mathscr{G}_t$-adapted processes $H(t)$ and $Q(t)$, and a partition $\Delta_K[0, \tau] = \{0 = t_0 < \cdots < t_K = \tau\}$, we define

$$\begin{aligned}&\Xi_{\text{out},\Delta_K[0,\tau]}(H, Q)\\ =&Q(\tau)\{\nu(Y) - H(\tau)\} + \sum_{j=0}^{K-1} Q(t_j)\left[\int H(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H(t_j)\right].\end{aligned}$$

We also define $\Xi_{\text{out}}(H, Q)$ as the limit of $\Xi_{\text{out},\Delta_K[0,\tau]}(H, Q)$ in probability whenever it exists, as $|\Delta_K[0,\tau]| \to 0$. We define

$$\mathcal{M}_{\text{out}} := \left\{(H, Q) : \Xi_{\text{out}}(H, Q) \text{ exists and } \mathbb{E}\left\{\Xi_{\text{out}}(H, Q)\right\} = \lim_{|\Delta_K[0,\tau]|\to 0} \mathbb{E}\left\{\Xi_{\text{out},\Delta_K[0,\tau]}(H, Q)\right\}\right\}.$$

With some regularity conditions,

$$(H_{\mathbb{G}}, Q) \in \mathcal{M}_{\text{out}},$$

at least for any locally bounded process Q. This can happen, for example, defining such limit Ξ_{out} is similar to defining stochastic integral in L^2 limit or Riemann-Stieltjes integral. The discussion on this details is beyond the scope of this paper but definitely of probabilistic interest. To prove the next theorem, we need a further assumption on the rate of how fast $\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}$ approaches $\mathbb{P}_{\mathbb{G}}$ in a tiny interval.

Assumption 5 (Approximating rate) *For any $\mathscr{G}_t$-adapted process $H(t)$, and any $0 \le s < t \le \tau$,*

$$\left|\mathbb{E}\left[\int H(t)\mathbb{G}\{\mathrm{d}\bar{a}(t)|\bar{A}(s)\}\,\mathbb{P}\{\mathrm{d}\bar{l}(t)|\mathscr{G}_s\} - \mathbb{E}_{\mathbb{G}}\{H(t)|\mathscr{G}_s\}\right]\right| \le \kappa\|H(s)\|_1(t-s)^{\alpha},$$

for some constant $\kappa > 0$ and $\alpha > 1$.

With the help of this assumption, one may obtain the following theorem, which states that a process is the g-computation process if and only if it is the solution to some class of estimating equations. Therefore, if one imposes a parametric or semiparametric form on $H_{\mathbb{G}}(t)$ as $H_{\mathbb{G}}(t;\theta)$, one might estimate θ via some Z-estimation. Or, this theorem may inspire nonparametric estimation of $H_{\mathbb{G}}(t)$ via minimax estimation (Ghassami et al., 2022) or maximum moment restriction (Kompa et al., 2022).

Proposition 1 (Identification of the g-computation process) *Under Assumptions 1, 2, and 5, for any $\mathscr{G}_t$-adapted process $Q(t)$ with $(H_{\mathbb{G}}, Q) \in \mathcal{M}_{\text{out}}$ and $\sup_t \|H_{\mathbb{G}}(t)Q(t)\|_1 < \infty$, $\Xi_{\text{out}}(H_{\mathbb{G}}, Q)$ is unbiased for zero, that is,*

$$\mathbb{E}\left\{\Xi_{\text{out}}(H_{\mathbb{G}}, Q)\right\} = 0. \tag{2}$$

Moreover, suppose there exists an $\mathscr{G}_t$-adapted process $H(t)$ and $\sup_t \|H(t)\|_{1,\mathbb{G}} < \infty$, so that for any $\mathscr{G}_t$-adapted process $Q(t)$ with $\sup_t \|H(t)Q(t)\|_1 < \infty$, we have $(H, Q) \in \mathcal{M}_{\text{out}}$ and

$$\mathbb{E}\left\{\Xi_{\text{out}}(H, Q)\right\} = 0. \tag{3}$$

Then $H(t)$ equals the g-computation process $H_{\mathbb{G}}(t)$ in Definition 1 for any $t \in [0,\tau]$ almost surely.

For any two $\mathscr{G}_t$-adapted processes $H(t)$ and $Q(t)$, and a partition $\Delta_K[0,\tau] = \{0 = t_0 < \cdots < t_K = \tau\}$, we define

$$\begin{aligned}&\Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q)\\ =&\sum_{j=1}^{K}\left[Q(t_j)H(t_j) - Q(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\\ &+\left[Q(0)H(0) - \int H(0)\mathbb{G}\{\mathrm{d}\bar{a}(0)\}\right].\end{aligned}$$

We also define $\Xi_{\text{trt}}(H, Q)$ as the limit of $\Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q)$ in probability whenever it exists, as $|\Delta_K[0, \tau]| \to 0$. We define

$$\mathcal{M}_{\text{trt}} := \left\{(H, Q) : \Xi_{\text{trt}}(H, Q) \text{ exists and } \mathbb{E}\left\{\Xi_{\text{trt}}(H, Q)\right\} = \lim_{|\Delta_K[0,\tau]|\to 0} \mathbb{E}\left\{\Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q)\right\}\right\}.$$

Likewise with some regularity conditions, we have

$$(H, Q_{\mathbb{G}}) \in \mathcal{M}_{\text{trt}},$$

for any locally bounded processes H and Q. Here we provide a theorem for identifying the IPW process like Theorem 1.

Proposition 2 (Identification of the IPW process) *Under Assumptions 1, 2, and 5, for any $\mathscr{G}_t$-adapted process $H(t)$ with $(H, Q_{\mathbb{G}}) \in \mathcal{M}_{\text{trt}}$ and $\sup_t \|H(t)Q_{\mathbb{G}}(t)\|_1 < \infty$, $\Xi_{\text{trt}}(H, Q_{\mathbb{G}})$ is unbiased for zero, that is,*

$$\mathbb{E}\left\{\Xi_{\text{trt}}(H, Q_{\mathbb{G}})\right\} = 0. \tag{4}$$

Moreover, suppose there exists a $\mathscr{G}_t$-adapted process $Q(t)$ and $\sup_t \|Q(t)\|_1 < \infty$, so that for any $\mathscr{G}_t$-adapted process $H(t)$ with $\sup_t \|H(t)Q(t)\|_1 < \infty$, we have $(H, Q) \in \mathcal{M}_{\text{trt}}$ and

$$\mathbb{E}\left\{\Xi_{\text{trt}}(H, Q)\right\} = 0. \tag{5}$$

Then $Q(t)$ equals the IPW process $Q_{\mathbb{G}}(t)$ in Definition 2 for any $t \in [0, \tau]$ almost surely.

Finally, we establish a doubly robust formula identifying (1). For any two $\mathscr{G}_t$-adapted processes $H(t)$ and $Q(t)$, and a partition $\Delta_K[0, \tau] = \{0 = t_0 < \cdots < t_K = \tau\}$, we define

$$\begin{aligned}\Xi_{\Delta_K[0,\tau]}(H, Q) =& \Xi_{\text{out},\Delta_K[0,\tau]}(H, Q) + \int H(0)\mathbb{G}\{\mathrm{d}\bar{a}(0)\} \\ =& Q(\tau)\nu(Y) - \Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q).\end{aligned}$$

We also define $\Xi(H, Q)$ as the limit of $\Xi_{\Delta_K[0,\tau]}(H, Q)$ in probability whenever it exists. Then clearly $\Xi(H, Q) = \Xi_{\text{out}}(H, Q) + \int H(0)\mathbb{G}\{\mathrm{d}\bar{a}(0)\} = Q(\tau)\nu(Y) - \Xi_{\text{trt}}(H, Q)$. The following theorem is immediate from the previous two theorems and fact that $\sup_t \|H_{\mathbb{G}}(t)Q_{\mathbb{G}}(t)\|_1 = \mathbb{E}_{\mathbb{G}} |\nu(Y)| < \infty$.

Theorem 3 (Doubly robust formula) *Under Assumptions 1 - 5, and assuming $(H_{\mathbb{G}}, Q_{\mathbb{G}}) \in \mathcal{M}_{\text{out}} \cap \mathcal{M}_{\text{trt}}$, (1) is identified via a doubly robust formula as*

$$\int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}) = \mathbb{E}\left\{\Xi(H_{\mathbb{G}}, Q_{\mathbb{G}})\right\}. \tag{6}$$

Furthermore, $\Xi(H, Q)$ is doubly robust in the sense that (6) remains true when either H or Q is correct but not necessarily both. That is, for any $\mathscr{G}_t$-adapted processes $Q(t)$ and $H(t)$ with $(H_{\mathbb{G}}, Q) \in \mathcal{M}_{\text{out}}$, $(H, Q_{\mathbb{G}}) \in \mathcal{M}_{\text{trt}}$ and $\sup_t \|H_{\mathbb{G}}(t)Q(t)\|_1 < \infty$, $\sup_t \|H(t)Q_{\mathbb{G}}(t)\|_1 < \infty$, we have

$$\int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}) = \mathbb{E}\left\{\Xi(H_{\mathbb{G}}, Q)\right\} = \mathbb{E}\left\{\Xi(H, Q_{\mathbb{G}})\right\}.$$

2.3. Strong but interpretable identification assumptions

We outline a sufficient set of conditions for our Assumptions 1 - 4 for easier interpretation. We start with a generalized "coarsening at random" assumption (Heitjan and Rubin, 1991; Gill et al., 1997; Tsiatis, 2006).

Assumption 6 (Coarsening at random) *The treatment assignment is independent of the all potential outcomes and covariates given history, in the sense that there exists a bounded function* $\varepsilon(t,\eta) > 0$ *with* $\int_0^\tau \varepsilon(t,\eta)dt \to 0$ *as* $\eta \to 0$*, such that for any* $\bar{a}' \in \mathcal{A}$*,* $t \in [0,\tau]$*,* $\eta > 0$*,* $\mathbb{E}(\| \mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}'} | \bar{A}(t+\eta), \bar{L}(t)\} - \mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}'} | \bar{A}(t), \bar{L}(t)\} \|_{\mathrm{TV}}) < \varepsilon(t,\eta)$.

This assumption claims that, the treatment distribution, or equally, the probability of coarsening, in a small period of time around t, only depends on the observed data up to time t and independent of other unobserved counterfactuals. We additional define counterfactual confounders $L_{\bar{a}}(t)$ and assuming that the future cannot affect the past, that is, $L_{\bar{a}}(t) = L_{\bar{a}'}(t)$ whenever $\bar{a}(t) = \bar{a}'(t)$. We strengthen the consistency assumption into:

Assumption 7 (Strong consistency) $L(t) = L_{\bar{A}}(t)$*, for any* t*, almost surely.*

Assumption 8 (Positivity) $\mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}})\mathbb{G}(\mathrm{d}\bar{a}) \ll \mathbb{P}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})$*, almost surely.*

Proposition 3 *Under Assumptions 6 and 7,*

$$\| \mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}})\mathbb{G}(\mathrm{d}\bar{a}) \|_{\mathrm{TV}} \to 0,$$

whenever $|\Delta_K[0,\tau]| \to 0$*. It is immediate that Assumption 2 is equivalent to Assumption 8. Assumptions 1 and 4 hold for any bounded function* $\nu(\cdot)$*.*

Consequently, one may, in principle, substitute Assumptions 6 - 8 for Assumptions 1 - 4 as foundational premises if Assumption 6 is found to be more intuitive and easier to grasp. Nevertheless, Assumptions 1 to 4 persist as the most minimalistic set of assumptions in this paper. It is worth noting a subtle distinction, namely that Assumption 4 is applicable solely to bounded functions ν, as implied by Assumption 6. This arises from the fact that Assumption 6 operates within the scope of total variation, necessitating boundedness for the dominated convergence theorem to hold.

3. Discussion

We establish a comprehensive framework for valid causal inference in continuous-time longitudinal studies, allowing for both continuous time progression and continuous data operation, without imposing constraints on the observed data distribution. We introduce two sets of sufficient assumptions for causal identification, which significantly expand upon the existing literature, encompassing scenarios involving both mortality and censoring. Additionally, we present three distinct identification approaches: the g-computation formula, IPW formula, and DR formula. Furthermore, we furnish a population estimating equation to discern the g-computation and IPW processes. While this study provides foundational insights into causal inference for FLD, our ongoing research is expanding these findings. A more extensive investigation, which is currently being prepared for journal submission, will explore additional facets such as official claims of its generality, nonparametric property. It will also accommodate additional complications in longitudinal studies like time-varying outcome of interest like mortality and right censoring. This forthcoming journal paper aims to provide a deeper understanding.

While numerical results often play a crucial role in validating statistical methods and providing empirical insights, it is essential to recognize the value of rigorous theoretical and population-level investigations. The functional data estimation framework is still in its early stages for longitudinal causal inference and there is a substantial theoretical gap that needs to be addressed. In fact, functional estimation has barely been developed for cross-sectional studies just in 2020s (Miao et al., 2020; Zhang et al., 2021; Tan et al., 2022), and it has not been considered for RLD or ILD (where if one considers data at each time point is functional), let alone FLD. Even without numerical results, theoretical contributions can establish fundamental concepts, identify key assumptions, and delineate the limitations of statistical methods. Such theoretical insights serve as a foundation for future empirical research and guide the development of more robust and effective statistical tools. Our paper focuses on the theoretical underpinnings of longitudinal causal inference, specifically for the FLD setting. This also helps to avoid shifting the focus from the theoretical framework's construction using stochastic process theory and measure theory, refraining from introducing sample-level notation and empirical process theory. This significant gap warrants dedicated attention in a separate, comprehensive paper.

Given the novelty of this direction, a multitude of unanswered questions persists. Firstly, delving deeper into Assumption 1, including exploring alternative sufficient conditions and conducting sensitivity analyses, is of paramount interest. Secondly, the positivity assumption in longitudinal studies may face challenges, as only a limited number of subjects in the observed study population may adhere to any given regimen. In practice, amalgamating information from diverse regimes becomes imperative. Hence, extending to semiparametric models, such as marginal structural models (Robins, 1998; Røysland, 2011) and structural nested models (Robins, 1999; Lok, 2008), holds promise. To further relax the positivity assumption, one can consider dynamic treatment regimes (Fitzmaurice et al., 2008; Young et al., 2011; Rytgaard et al., 2022) or incremental interventions (Kennedy, 2017). Generalizing the framework in cases where Assumption 4 falters, including situations involving time-dependent instrumental variables (Tchetgen Tchetgen et al., 2018) and time-dependent proxies (Tchetgen Tchetgen et al., 2020; Ying et al., 2023), is also a viable avenue for exploration. While the primary focus has been on extending the identification results of causal inference from discrete to continuous longitudinal studies, the formalization of estimation and inferential outcomes has garnered considerable interest and is currently under active investigation. We posit that with a judicious estimation of the g-computation process, the IPW process, and leveraging the doubly robust formula (6), one can achieve both model double-robustness and rate double-robustness (Smucler et al., 2019) for estimation. While the real functional data are typically stored in a discrete manner, we do believe the discretely observed data can identify the causal effects, but it needs careful assumptions on how the sample size approaches infinity and grid size approaches to zero. Characterizing the efficiency lower bound for the quantity of interest, leveraging semiparametric theory, is an enticing prospect. Additionally, a formal delineation of assumptions and theory underpinning the limit in probability of pairs (H, Q) in $\mathcal{M}_{\text{out},\mathbb{G}}$ and $\mathcal{M}_{\text{trt},\mathbb{G}}$ is of profound probabilistic interest. From a simulation standpoint, investigating if, and under what conditions, we can consistently estimate (1) using discrete-time observed processes with diminishing meshes poses intriguing questions. Moreover, exploring the extent of partial identification using a discrete-time observed process constitutes an area of continued interest.

Appendix A. Proofs

A.1. Proof of Theorem 1

Since $\mathbb{P}_{\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})$ is a limit of measures in total variation norm of

$$\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}),$$

whenever $|\Delta_K[t,\tau]| \to 0$, we have

$$\int f(\bar{a},\bar{l})\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) \to \mathbb{E}_{\mathbb{G}}\{f(\bar{A},\bar{L})\},$$

for any indicator functions $f(\bar{a},\bar{l})$. Therefore, with linearity of expectations and possibly monotone convergence theorem, we have

$$\begin{aligned}
&\left|H_{\mathbb{G}}(0-) - \int \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a})\right| \\
\leq &\left|\int \nu(y)\,\mathbb{P}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})\right| \\
&+\left|\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \int \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a})\right| \\
= &\left|\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \int \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a})\right| + o(1) \\
\leq &\left|\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \int \mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{a}(t_1),\bar{L}(t_0)\}]\mathbb{G}(\mathrm{d}\bar{a})\right| \\
&+\left|\int \mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{a}(t_1),\bar{L}(t_0)\}]\mathbb{G}(\mathrm{d}\bar{a}) - \int \mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{L}(t_0)\}]\mathbb{G}(\mathrm{d}\bar{a})\right| + o(1) \\
\leq &\left|\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \int \mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{a}(t_1),\bar{L}(t_0)\}]\mathbb{G}(\mathrm{d}\bar{a})\right| \\
&+\varepsilon(t_0,t_1-t_0) + o(1) \\
\leq &\Bigg|\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) \\
&-\int \mathbb{E}(\mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{a}(t_1),\bar{L}(t_1)\}|\bar{a}(t_1),\bar{L}(t_0)])\mathbb{G}(\mathrm{d}\bar{a})\Bigg| \\
&+\varepsilon(t_0,t_1-t_0) + o(1) \\
&\cdots \\
\leq &\Bigg|\int \mathbb{E}(\mathbb{E}[\mathbb{E}\{\nu(Y_{\bar{a}})|\bar{a}(t_K),\bar{L}(t_{K-1})\}|\cdots|\bar{a}(t_0),\bar{L}(t_{-1})])\mathbb{G}(\mathrm{d}\bar{a}) \\
&-\int \nu(y)\,\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l})\Bigg| \\
&+\sum_{j=0}^{K-1}\varepsilon(t_j,t_{j+1}-t_j) + o(1) \\
= &\sum_{j=0}^{K-1}\varepsilon(t_j,t_{j+1}-t_j) + o(1) \to 0,
\end{aligned}$$

as $K \to \infty$ and hence the conclusion.

A.2. Proof of Theorem 2

The proof is immediate by noting that

$$\mathbb{E}\left\{Q_{\mathbb{G}}(\tau)\nu(Y)\right\} = \mathbb{E}_{\mathbb{G}}\{\nu(Y)\} = \int_{\mathcal{A}} \mathbb{E}\{\nu(Y_{\bar{a}})\}\mathbb{G}(\mathrm{d}\bar{a}),$$

by Theorem 1.

A.3. Proof of Proposition 1

We first prove the first part of the theorem that $H_{\mathbb{G}}$ is a solution to the class of estimating equations (2) given in the theorem. To that end, we need to show that for any $\mathscr{G}_t$-adapted process $Q(t)$ with $(H_{\mathbb{G}}, Q) \in \mathcal{M}_{\text{out}}$ and $\sup_t \|H_{\mathbb{G}}(t)Q(t)\|_1 < \infty$, (2) holds. Indeed,

$$\begin{aligned}
&|\mathbb{E}\{\Xi_{\text{out}}(H_{\mathbb{G}}, Q)\}| \\
\leq & \left|\mathbb{E}\left\{\Xi_{\text{out},\Delta_K[0,\tau]}(H_{\mathbb{G}}, Q)\right\}\right| + \left|\mathbb{E}\left\{\Xi_{\text{out},\Delta_K[0,\tau]}(H_{\mathbb{G}}, Q)\right\} - \mathbb{E}\left\{\Xi_{\text{out}}(H_{\mathbb{G}}, Q)\right\}\right| \\
= & \Bigg| \mathbb{E}\left(Q(\tau)[\nu(Y) - \mathbb{E}_{\mathbb{G}}\{\nu(Y)|\bar{A}, \bar{L}\}]\right) \\
& + \sum_{j=1}^{K-1} \mathbb{E}\left(Q(t_j)\left[\int H_{\mathbb{G}}(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H_{\mathbb{G}}(t_j)\right]\right)\Bigg| + o(1) \\
= & \left|0 + \sum_{j=1}^{K-1} \mathbb{E}\left(Q(t_j)\left[\int H_{\mathbb{G}}(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H_{\mathbb{G}}(t_j)\right]\right)\right| + o(1) \\
\leq & \sum_{j=1}^{K-1} \left|\mathbb{E}\left(Q(t_j)\left[\int H_{\mathbb{G}}(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H_{\mathbb{G}}(t_j)\right]\right)\right| + o(1) \\
\leq & \sum_{j=1}^{K-1} \left|\mathbb{E}\left(Q(t_j)\left[\int H_{\mathbb{G}}(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - \mathbb{E}_{\mathbb{G}}\left\{H_{\mathbb{G}}(t_{j+1})|\mathscr{G}_{t_j}\right\}\right]\right)\right| + o(1) \\
\leq & \sum_{j=0}^{K} \kappa\|H_{\mathbb{G}}(t_j)Q(t_j)\|_1(t_{j+1} - t_j)^{\alpha} + o(1) \\
\leq & \kappa \sup_t \|H_{\mathbb{G}}(t)Q(t)\|_1 \sum_{j=1}^{K-1}(t_{j+1} - t_j)^{\alpha} + o(1) \\
\to & 0,
\end{aligned}$$

when $|\Delta_K[0,\tau]| \to 0$.

We now prove the second part of the theorem. Now suppose $H(t)$ is a solution to the class of estimating equations (3). To prove $H(t) = H_{\mathbb{G}}(t)$, it suffices to show that $H(t)$ is a $\mathbb{P}_{\mathbb{G}}$ martingale with initial condition $H(\tau) = \nu(Y)$, that is, by Doob's theorem, also equivalent to showing

$$\mathbb{E}_{\mathbb{G}}\left\{\int_0^{\tau} Q(t)\mathrm{d}H(t)\right\} = 0,$$

for a sufficiently rich set of $Q(t)$. We plug in a process $Q'(t) = Q_{\mathbb{G}}(t)Q(t)$ for some bounded process $Q(t)$. We have $\sup_t \|H(t)Q'(t)\|_1 = \sup_t \|H(t)Q(t)\|_{1,\mathbb{G}} < \infty$. It hence follows that $(H, Q) \in \mathcal{M}_{\text{out}}$ and

$$\mathbb{E}\left\{\Xi_{\text{out}}(H, Q')\right\} = 0.$$

Therefore we have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{G}}\left\{\int_0^\tau Q(t)\mathrm{d}H(t)\right\} \\
=&\mathbb{E}_{\mathbb{G}}\left[\lim_{|\Delta_K[0,\tau]|\to 0}\sum_{j=1}^{K-1} Q(t_j)\left\{H(t_{j+1}) - H(t_j)\right\}\right] \\
=&\lim_{|\Delta_K[0,\tau]|\to 0}\mathbb{E}_{\mathbb{G}}\left[\sum_{j=1}^{K-1} Q(t_j)\left\{H(t_{j+1}) - H(t_j)\right\}\right] \\
=&\lim_{|\Delta_K[0,\tau]|\to 0}\mathbb{E}_{\mathbb{G}}\left(\sum_{j=1}^{K-1} Q(t_j)\left[\mathbb{E}_{\mathbb{G}}\{H(t_{j+1})|\mathscr{G}_{t_j}\} - H(t_j)\right]\right) \\
=&\lim_{|\Delta_K[0,\tau]|\to 0}\mathbb{E}_{\mathbb{G}}\left(\sum_{j=1}^{K-1} Q(t_j)\left[\int H(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\}\,\mathbb{P}\{\mathrm{d}\bar{l}(t_{j+1})|\mathscr{G}_{t_j}\} - H(t_j)\right]\right) \\
=&\lim_{|\Delta_K[0,\tau]|\to 0}\mathbb{E}\left(\sum_{j=1}^{K-1} Q'(t_j)\left[\int H(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\}\,\mathbb{P}\{\mathrm{d}\bar{l}(t_{j+1})|\mathscr{G}_{t_j}\} - H(t_j)\right]\right) \\
=&\lim_{|\Delta_K[0,\tau]|\to 0}\mathbb{E}\left(\sum_{j=1}^{K-1} Q'(t_j)\left[\int H(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H(t_j)\right]\right) \\
=&\mathbb{E}\left(\lim_{|\Delta_K[0,\tau]|\to 0}\sum_{j=1}^{K-1} Q'(t_j)\left[\int H(t_{j+1})\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{A}(t_j)\} - H(t_j)\right]\right) \\
=&\mathbb{E}\left\{\Xi_{\text{out}}(H, Q')\right\} = 0.
\end{aligned}$$

By Doob's Theorem, $H(t)$ is a $\mathbb{P}_{\mathbb{G}}$-martingale with respect to $\mathscr{F}_t$ with $H(\tau+) = H_{\mathbb{G}}(\tau+) = \nu(Y)$. Therefore we have $H(t) = \mathbb{E}_{\mathbb{G}}\{H(\tau)|\mathscr{G}_t\} = \mathbb{E}_{\mathbb{G}}\{\nu(Y)|\mathscr{G}_t\} = H_{\mathbb{G}}(t)$ for any $t \in [0, \tau]$ almost surely.

A.4. Proof of Proposition 2

We prove the first part of the theorem that $Q_{\mathbb{G}}$ is a solution to the class of estimating equations (4) given in the theorem. To that end, we need to show that for any $\mathbb{P}_{\mathbb{G}}$-integrable process $H(t)$ with

$(H, Q_{\mathbb{G}}) \in \mathcal{M}_{\text{trt}}$ and $\sup_t \|H(t)Q_{\mathbb{G}}(t)\|_1 < \infty$, (4) holds. Indeed,

$$
\begin{aligned}
&|\mathbb{E}\{\Xi_{\text{trt}}(H, Q_{\mathbb{G}})\}| \\
\leq &\left|\mathbb{E}\left\{\Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q_{\mathbb{G}})\right\}\right| + \left|\mathbb{E}\{\mathbb{E}\left\{\Xi_{\text{trt},\Delta_K[0,\tau]}(H, Q_{\mathbb{G}})\right\} - \Xi_{\text{trt}}(H, Q_{\mathbb{G}})\}\right| \\
= &\Bigg|\mathbb{E}\left(\sum_{j=0}^{K}\left[Q_{\mathbb{G}}(t_j)H(t_j) - Q_{\mathbb{G}}(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\right) \\
&- \mathbb{E}\left[Q_{\mathbb{G}}(0)H(0) - \int H(0)\mathbb{G}\{\mathrm{d}\bar{a}(0)\}\right]\Bigg| + o(1) \\
= &\left|\mathbb{E}\left(\sum_{j=0}^{K}\left[Q_{\mathbb{G}}(t_j)H(t_j) - Q_{\mathbb{G}}(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\right) + 0\right| + o(1) \\
\leq &\sum_{j=0}^{K}\left|\mathbb{E}\left[Q_{\mathbb{G}}(t_j)H(t_j) - Q_{\mathbb{G}}(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\right| + o(1) \\
= &\sum_{j=0}^{K}\Bigg|\mathbb{E}\Bigg[Q_{\mathbb{G}}(t_{j-1})\,\mathbb{E}_{\mathbb{G}}\{H(t_j)|\mathcal{G}_{t_{j-1}}\} \\
&- Q_{\mathbb{G}}(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\,\mathbb{P}\{\mathrm{d}\bar{l}(t_j)|\mathcal{G}_{t_{j-1}}\}\Bigg]\Bigg| + o(1) \\
\leq &\kappa\sum_{j=0}^{K}\|H(t_{j-1})Q_{\mathbb{G}}(t_{j-1})\|_1(t_j - t_{j-1})^{\alpha} + o(1) \\
\leq &\kappa\sup_t\|H(t)Q_{\mathbb{G}}(t)\|_1\sum_{j=0}^{K}(t_j - t_{j-1})^{\alpha} + o(1) \to 0,
\end{aligned}
$$

when $|\Delta_K[0,\tau]| \to 0$.

We now prove the second part of the theorem. Now suppose $Q(t)$ is a solution to the class of estimating equations (5). To prove $Q(t) = Q_{\mathbb{G}}(t)$, it suffices to show that $Q(t)$ satisfies the property of a Radon-Nikodym derivative process, that is, for any u,

$$\mathbb{E}\{H(u)Q(u)\} = \mathbb{E}_{\mathbb{G}}\{H(0)\},$$

for a sufficiently rich set of $H(t)$. To that end, we plug in any bounded process $H(t)$ with $H(t) = 0$ for any $t > u$ and $H(t) = \mathbb{E}_{\mathbb{G}}\{H(u)|\mathscr{G}_t\}$ for any $t \leq u$, for some $u > 0$, treating $H(t_{-1}) = H(0-) = \mathbb{E}_{\mathbb{G}}\{H(0)\}$. We have $\sup_t \|H(t)Q(t)\|_1 < \infty$ and it hence follows that $(H, Q) \in \mathcal{M}_{\text{trt}}$ and

$$\mathbb{E}\{\Xi_{\text{trt}}(H, Q)\} = 0.$$

Therefore we have

$$
\begin{aligned}
&\mathbb{E}\{H(u)Q(u)\} - \mathbb{E}_{\mathbb{G}}\{H(0)\} \\
=& \lim_{|\Delta_K[0,u]|\to 0} \mathbb{E}\left[\sum_{j=0}^{K}\{Q(t_j)H(t_j) - Q(t_{j-1})H(t_{j-1})\}\right] \\
=& \lim_{|\Delta_K[0,u]|\to 0} \mathbb{E}\left(\sum_{j=0}^{K}\left[Q(t_j)H(t_j) - Q(t_{j-1})\,\mathbb{E}_{\mathbb{G}}\{H(t_j)|\mathscr{G}_{t_{j-1}}\}\right]\right) \\
=& \lim_{|\Delta_K[0,u]|\to 0} \mathbb{E}\left(\sum_{j=0}^{K}\left[Q(t_j)H(t_j) - Q(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\right) \\
=& \mathbb{E}\left(\lim_{|\Delta_K[0,u]|\to 0}\sum_{j=0}^{K}\left[Q(t_j)H(t_j) - Q(t_{j-1})\int H(t_j)\mathbb{G}\{\mathrm{d}\bar{a}(t_j)|\bar{A}(t_{j-1})\}\right]\right) \\
=& \mathbb{E}\{\Xi_{\text{trt}}(H,Q)\} = 0.
\end{aligned}
$$

Therefore $M(u)$ satisfies the definition of Radon-Nikodym derivative and by the uniqueness of the Radon-Nikodym derivative we conclude that $Q(t) = Q_{\mathbb{G}}(t)$ for any $t \in [0,\tau]$ almost surely.

A.5. Proof of Proposition 3

Below we use Assumption 3, 6, and a triangular inequality,

$$
\begin{aligned}
&\left\|\mathbb{P}_{\Delta_K[0,\tau],\mathbb{G}}(\mathrm{d}\bar{a}\mathrm{d}\bar{l}) - \mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}})\mathbb{G}(\mathrm{d}\bar{a})\right\|_{\text{TV}} \\
=&\left\|\prod_{j=0}^{K-1}\left[\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{a}(t_j)\}\,\mathbb{P}\{\mathrm{d}\bar{l}(t_{j+1})|\bar{a}(t_{j+1}),\bar{l}(t_j)\}\right] - \mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}\in\mathcal{A}})\mathbb{G}(\mathrm{d}\bar{a})\right\|_{\text{TV}} \\
=&\left\|\prod_{j=0}^{K-1}\left[\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{a}(t_j)\}\,\mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{a}(t_{j+1}),\bar{l}_{\bar{a}}(t_j)\}\right] - \mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}})\mathbb{G}(\mathrm{d}\bar{a})\right\|_{\text{TV}} \\
\le&\left\|\prod_{j=0}^{K-1}\left[\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{a}(t_j)\}\,\mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{l}_{\bar{a}}(t_j)\}\right] - \mathbb{P}(\mathrm{d}\bar{l}_{\bar{a}})\mathbb{G}(\mathrm{d}\bar{a})\right\|_{\text{TV}} \\
&+\left\|\prod_{j=0}^{K-1}\left(\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{a}(t_j)\}\left[\mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{a}(t_{j+1}),\bar{l}_{\bar{a}}(t_j)\} - \mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{l}_{\bar{a}}(t_j)\}\right]\right)\right\|_{\text{TV}} \\
=&\left\|\prod_{j=0}^{K-1}\left(\mathbb{G}\{\mathrm{d}\bar{a}(t_{j+1})|\bar{a}(t_j)\}\left[\mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{a}(t_{j+1}),\bar{l}_{\bar{a}}(t_j)\} - \mathbb{P}\{\mathrm{d}\bar{l}_{\bar{a}}(t_{j+1})|\bar{l}_{\bar{a}}(t_j)\}\right]\right)\right\|_{\text{TV}} \\
&\to 0.
\end{aligned}
$$

References

Rabindra Nath Bhattacharya and Edward C Waymire. *A Basic Course in Probability Theory*, volume 69. Springer, 2007.

Lauren E Cain, James M Robins, Emilie Lanoy, Roger Logan, Dominique Costagliola, and Miguel A Hernán. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *The international journal of biostatistics*, 6(2):1–26, 2010.

Hyon K Choi, Miguel A Hernán, John D Seeger, James M Robins, and Frederick Wolfe. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *The Lancet*, 359 (9313):1173–1177, 2002.

Daniel Commenges and Anne Gégout-Petit. A general dynamical statistical model with causal interpretation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3): 719–736, 2009.

David Roxbee Cox and Nanny Wermuth. *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press, 2014.

A Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):1–15, 1979.

A Philip Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, 2000.

Rick Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 2019.

Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. *Longitudinal Data Analysis*. CRC press, 2008.

Sara Gisella Geneletti. *Aspects of Casual Inference in a Non-counterfactual Framework*. University of London, University College London (United Kingdom), 2005.

AmirEmad Ghassami, Andrew Ying, Ilya Shpitser, and Eric Tchetgen Tchetgen. Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *International Conference on Artificial Intelligence and Statistics*, pages 7210–7239. PMLR, 2022.

Richard D Gill and James M Robins. Causal inference for complex longitudinal data: the continuous case. *The Annals of Statistics*, 29(2):1785–1811, 2001.

Richard D Gill, Mark J Van Der Laan, and James M Robins. Coarsening at random: Characterizations, conjectures, counter-examples. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pages 255–294. Springer, 1997.

Sander Greenland and James M Robins. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, 15(3):413–419, 1986.

Daniel F Heitjan and Donald B Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19(4): 2244–2253, 1991.

Miguel A Hernán and James M Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Miguel A Hernán, Babette Brumback, and James M Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448, 2001.

Miguel A Hernán, Babette A Brumback, and James M Robins. Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in Medicine*, 21(12):1689–1709, 2002.

Miguel Ángel Hernán, Babette Brumback, and James M Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, 11 (5):561–570, 2000.

Paul W Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

Liangyuan Hu and Joseph W Hogan. Causal comparative effectiveness analysis of dynamic continuous-time treatment initiation rules with sparsely measured outcomes and death. *Biometrics*, 75(2):695–707, 2019.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):1–9, 2016.

Alistair EW Johnson, David J Stone, Leo A Celi, and Tom J Pollard. The mimic code repository: enabling reproducibility in critical care research. *Journal of the American Medical Informatics Association*, 25(1):32–39, 2018.

Brent A Johnson and Anastasios A Tsiatis. Semiparametric inference in observational duration-response studies, with duration possibly right-censored. *Biometrika*, 92(3):605–618, 2005.

Karl G Jöreskog. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4): 443–477, 1978.

Edward H Kennedy. Semiparametric theory. *arXiv preprint arXiv:1709.06418*, 2017.

David C Klonoff. Continuous glucose monitoring: roadmap for 21st century diabetes therapy. *Diabetes care*, 28(5):1231–1239, 2005.

Benjamin Kompa, David R Bellamy, Thomas Kolokotrones, James M Robins, and Andrew L Beam. Deep learning methods for proximal inference via maximum moment restriction. *arXiv preprint arXiv:2205.09824*, 2022.

Steffen Lilholt Lauritzen and Nanny Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, 17(1):31–57, 1989.

Judith J Lok. *Statistical modelling of causal effects in time*. PhD thesis, Vrije Universiteit Amsterdam, 2001. Naam instelling promotie: VU Vrije Universiteit Naam instelling onderzoek: VU Vrije Universiteit.

Judith J Lok. Mimicking counterfactual outcomes for the estimation of causal effects. *arXiv preprint math.ST/0409045*, 2004.

Judith J Lok. Statistical modeling of causal effects in continuous time. *The Annals of Statistics*, 36 (3):1464–1507, 2008.

John J Mastrototaro. The minimed continuous glucose monitoring system. *Diabetes technology & therapeutics*, 2(1, Supplement 1):13–18, 2000.

Rui Miao, Wu Xue, and Xiaoke Zhang. Average treatment effect estimation in observational studies with functional covariates. *arXiv preprint arXiv:2004.06166*, 2020.

Jersey Neyman. Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10(1):1–51, 1923.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jamees M Robins. Marginal structural models. In *1997 Proceedings of the Section on Bayesian Statistical Science*, pages 1–10. Alexandria, VA: American Statistical Association, 1998.

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

James M Robins. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases*, 40:139S–161S, 1987.

James M Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, pages 113–159, 1989.

James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.

James M Robins. Association, causation, and marginal structural models. *Synthese*, 121(1/2):151–179, 1999.

David Rodbard. Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes Technology & Therapeutics*, 18(S2):S2–3, 2016.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Kjetil Røysland. A martingale approach to continuous-time marginal structural models. *Bernoulli*, 17(3):895–915, 2011.

Kjetil Røysland. Counterfactual analyses with graphical models based on local independence. *The Annals of Statistics*, 40(4):2162–2194, 2012.

Kjetil Røysland, Pål Ryalen, Mari Nygård, and Vanessa Didelez. Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses. *arXiv preprint arXiv:2202.02311*, 2022.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

Helene C Rytgaard, Thomas A Gerds, and Mark J van der Laan. Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. *The Annals of Statistics*, 50(5): 2469–2491, 2022.

Hermann Singer. Nonlinear continuous time modeling approaches in panel research. *Statistica Neerlandica*, 62(1):29–57, 2008.

Ezequiel Smucler, Andrea Rotnitzky, and James M Robins. A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. *arXiv preprint arXiv:1904.03737*, 2019.

Jinghao Sun and Forrest W Crawford. Causal identification for continuous-time stochastic processes. *arXiv preprint arXiv:2211.15934*, 2022.

Ruoxu Tan, Wei Huang, Zheng Zhang, and Guosheng Yin. Causal effect of functional treatment. *arXiv preprint arXiv:2210.00242*, 2022.

Eric J. Tchetgen Tchetgen, Haben Michael, and Yifan Cui. Marginal structural models for time-varying endogenous treatments: A time-varying instrumental variable approach. *arXiv preprint arXiv:1809.05422*, 2018.

Eric J. Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*, 2020.

Anastasios A Tsiatis. *Semiparametric Theory and Missing Data*. Springer, 2006.

Mark J van der Laan. Targeted maximum likelihood based causal inference: Part i. *The International Journal of Biostatistics*, 6(2), 2010.

Stijn Vansteelandt and Oliver Dukes. Assumption-lean inference for generalised linear model parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):657–685, 2022.

Stijn Vansteelandt and Marshall Joffe. Structural nested models and g-estimation: the partially realized promise. *Statistical Science*, 29(4):707–731, 2014.

Shu Yang. Semiparametric estimation of structural nested mean models with irregularly spaced longitudinal observations. *Biometrics*, 78(3):937–949, 2022.

Andrew Ying, Wang Miao, Xu Shi, and Eric J Tchetgen Tchetgen. Proximal causal inference for complex longitudinal studies. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):684–704, 2023.

Jessica G Young, Lauren E Cain, James M Robins, Eilis J O'Reilly, and Miguel A Hernán. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences*, 3(1):119–143, 2011.

Xiaoke Zhang, Wu Xue, Qiyue Wang, et al. Covariate balancing functional propensity score for functional treatments in cross-sectional observational studies. *Computational Statistics & Data Analysis*, 163(C), 2021.