INCREMENTAL CAUSAL EFFECT FOR TIME TO TREATMENT INITIALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We consider time to treatment initialization. This can commonly occur in preventive medicine, such as disease screening and vaccination; it can also occur with non-fatal health conditions such as HIV infection without the onset of AIDS. While traditional causal inference focused on 'when to treat' and its effects, including their possible dependence on subject characteristics, we consider the incremental causal effect when the intensity of time to treatment initialization is intervened upon. We provide identification of the incremental causal effect without the commonly required positivity assumption, as well as an estimation framework using inverse probability weighting. We illustrate our approach via simulation, and apply it to a rheumatoid arthritis study to evaluate the incremental effect of time to start methotrexate on joint pain.

1 INTRODUCTION

As a motivating example we borrow one from Bonvini et al. (2023) on behavioral health services for probationers, in order to reduce their chances of re-arrest. The question of interest, to be elaborated below, is about shifting the odds of receiving such a treatment. This can be achieved via affecting probationers' likelihood of attending services by, for example, providing transportation stipends. In public health, guidelines can be provided in order to affect people's likelihood of getting disease screening or vaccination.

An intervention such as in the above examples on the treatment distribution, is referred to as incremental intervention. The resulting causal effect of an incremental intervention is called the incremental causal effect (Kennedy, 2019; Naimi et al., 2021; Kim et al., 2021; Sarvet et al., 2023; Bonvini et al., 2023). In general the incremental causal effect has the interpretation of a policy effect on the population.

To better understand incremental intervention we provide a brief overview of the types of interventions in causal inference, which centers around answering "what happens if a particular treatment assignment changes ..." type of questions. An intervention on the treatment can change the value of a treatment, like from control to treated, or it can change its distribution, which is the case for incremental intervention. Causal inference literature started with investigating *static deterministic* interventions (Rubin, 1974; Hernán & Robins, 2020). Here 'static' is in contrast with 'dynamic', where the (possibly time-varying) treatment assignment may depend on covariates (and prior treatments and outcomes) (Robins, 1986; Murphy, 2003; Robins, 2004; Moodie et al., 2007; Young et al., 2011; Haneuse & Rotnitzky, 2013; Rytgaard et al., 2022; Ying, 2024b). In contrast, under a static intervention the subjects are either all treated or all untreated; this is typically the case when considering an average treatment effect (ATE). A second aspect about intervention is 'deterministic' versus 'stochastic'. Deterministic interventions assign a particular value of treatment to each individual, which may depend on the covariates, in the case of a deterministic dynamic intervention. On the other hand, stochastic interventions do not assign a particular value of the treatment, but rather a probability of receiving a particular treatment value (Cain et al., 2010; Díaz & van der Laan, 2012; Díaz & Hejazi, 2020; van der Laan et al., 2018). More specifically, a stochastic intervention assigns treatment randomly based on a probability distribution. Bonvini et al. (2023) describes the example of a Bernoulli distribution for a binary point treatment, where a single parameter $p \in [0, 1]$ indices such an intervention. Finally, an intervention can be 'pre-specified' versus 'otherwise'. For example, Young et al. (2011) described a pre-specified intervention as: "start combined antiretroviral therapy

(cART) within 6 months of CD4 cell count first dropping below x or diagnosis of an AIDS-defining illness, whichever happens first". This is a pre-specified dynamic treatment regime as whether an individual starts treatment depends on their own health related history captured by the covariates.

An incremental intervention (Kennedy, 2019) is a special stochastic dynamic intervention. It is not pre-specified, and it intervenes on the observed treatment distribution. Here by observed treatment distribution we mean the default, or baseline, or 'natural', treatment distribution, without the incremental intervention. In addition to the causal interpretation as policy effects above, incremental causal effects have the notable advantage of not requiring the *positivity* assumption typically required of the more traditional causal estimands.

The literature referenced above has studied incremental causal effect in the setting of a binary treatment. For discrete time longitudinal studies, the incremental intervention leads to modifying the odds of receiving the binary treatment at each time point Kennedy (2019); Naimi et al. (2021); Kim et al. (2021); Sarvet et al. (2023). To our best knowledge, however, there is currently no investigation in the literature on incremental causal effects by modifying the distribution of time to treatment. The examples described at the start of this paper fall under time to treatment; that is, given a well defined time zero, how long a subject has to or choose to wait before receiving the treatment.

In this paper, we generalize existing work on incremental causal effects to the time to (initializing) treatment and prove that one can avoid the positivity assumption in this case as well. We first define incremental causal effects by shifting hazard function via hazard ratio. We show that they can be identified by the observed data with only sequential randomization assumption and consistency assumption, without the often needed positivity assumption in causal inference. We construct a generic inverse probability weighted estimator accommodating numerous estimators for the hazard function, coupled with root-n inference results via random weighting bootstrap. Our estimation tools can leverage off-the-shell software from survival analysis. We examine the finite-sample performance of these estimators via extensive Monte-Carlo simulations. We apply our estimator onto a rheumatoid arthritis study to evaluate the effect of methotrexate.

The remainder of the article is organized as follows. In Section 1.1 we review the positivity assumption, and we review related work in Section 2. We introduce notation, the causal estimand, and key assumptions for identification in Section 3.1, where we construct an inverse probability weighting identification approach. In Section 3.2, we propose an inverse probability weighted estimator, together with inferential results yielding confidence intervals. In Section 4.1, we conduct extensive simulations examining finite-performance of our estimator and inference results. In Section 4.2, we apply our framework to answer what happens if we change the probability of receiving methotrexate on rheumatoid arthritis. We end the paper with a discussion in Section 5.

1.1 Positivity assumption

To identify estimands such as the average treatment effect from the observed data distribution, we typically need at least three common assumptions: consistency, positivity, and no unmeasured confounding or exchangeability (Hernán & Robins, 2020). Consistency is needed in order to bridge the observed data with the potential outcomes, as well as to exclude the possibility of interference where one subject's treatment affects another subject's outcome. No unmeasured confounding has been studied extensively in the literature, including instrumental variables (Angrist et al., 1996; Ying et al., 2019; Ying & Tchetgen Tchetgen, 2023) and proximal approaches (Miao et al., 2018; Tchetgen Tchetgen et al., 2024; Ying et al., 2023), as well as sensitivity analysis towards the violation of this assumption (Rosenbaum, 2002). Here we focus on the positivity assumption; that is, each subject in the population must have a positive probability of receiving each treatment level. In addition, for estimation approaches such as inverse probability weighting (IPW) or doubly robust approaches to have valid asymptotic inference, the stronger *strict positivity or overlap* assumption is needed; that is, the probability of receiving each treatment level needs to be bounded away from zero. These assumptions, while testable using observed data, can be violated in practice. Bonvini et al. (2023) provided a recividism example where some probationers might be required to obtained the treatment under study by attending behavioral health services, therefore their probability of receiving treatment is one and not receiving treatment is zero. Similar cases can be found for vaccination where some people may not be suitable for receiving the vaccine. D'Amour et al. (2021) showed mathematically that overlap is hard to satisfy when many covariates included in order to meet the

unconfoundedness assumption that is also required in many causal inference approaches. For studies involving time-varying treatment, covariates, and outcomes, the positivity assumption is even more of a challenge to meet.

Therefore, many traditional causal quantities of interest might be vetoed by absence of positivity alone after checking the data. On the other hand, incremental causal effects bridge this gap, by not requiring positivity. This is because for subjects with degenerate probabilities (of zero or one) of receiving treatments, perturbing the odds does not change their degenerate probabilities, therefore the positivity can be seen as always satisfied (Kennedy, 2019). More generally speaking, to non-parametrically identify the effect of intervention on the treatment value, such as 0 or 1, requires data on subjects receiving both treatments 0 and 1 over the population of interest. On the other hand, to identify the effect of intervention on the treatment distribution, and not specific values, as will be seen below, does not require data on subjects receiving all treatment values over the population of interest.

2 Related Work

As mentioned above Kennedy (2019) proposed for binary treatment incremental intervention which fully resolved the positivity issue mentioned above. A disadvantage of the incremental causal effect, quoting Kennedy (2019), is: "First, we expect incremental causal effects to play a more descriptive than prescriptive role compared to other approaches. Specifically, they give an interpretable picture of what would happen if exposure were increased or decreased in a natural way, but will likely be less useful for informing specific treatment decisions." As will be seen, contrary to what is said here, the incremental causal effect in our setting can yield useful policy-making decisions.

Díaz & van der Laan (2012), Haneuse & Rotnitzky (2013), Díaz & van der Laan (2013), Young et al. (2014), Díaz et al. (2023a), and Díaz et al. (2023b) considered general interventions that depend on the observed treatment process, and are referred to as modified treatment policies or interventions that depend on the natural value of treatment, the latter referring to a treatment that a subject would 'naturally' receive in the absence of it being assigned by an intervention. An example from Haneuse & Rotnitzky (2013) concerns the surgery time for a patient, and asks the question of what if every patient's operating time is reduced by say 5 minutes. Note that each patient's (observed) operating time depends potentially on their individual covariates as well as additional unobserved attributes. In this sense the observed operating time contains more information than what is captured by the measured covariates. In a similar vein, the natural value of treatment also emphasizes the belief that it contains more information than what is captured by the measured covariates (Stensrud et al., 2024). Because these interventions involve more information than the measured covariates, they are different from the pre-specified dynamic intervention introduced in the previous section (Haneuse & Rotnitzky, 2013). Our work here also involves the observed treatment distribution, and considers the continuous time to treatment initialization that has not been previously studied in the literature.

Time to treatment or treatment initialization, or to a large extent similarly, time to treatment termination, has been consider in the dynamic treatment setting (Johnson & Tsiatis, 2005; Yang et al., 2018; Nie et al., 2021). In particular, the first two references (Johnson & Tsiatis, 2005; Yang et al., 2018) considered continuous time to treatment termination, while (Nie et al., 2021) considered discrete time. All of the above are for deterministic interventions.

Continuous time causal inference in longitudinal studies for stochastic interventions, a broader setting than time to treatment initialization studies, is general challenging, and limited work has been done in the field (Rytgaard et al., 2022; Ying, 2022a; 2024a;b; Sun & Crawford, 2022). Among these, Ying (2022a; 2024a) and Sun & Crawford (2022) only considered static intervention, Ying (2024b) obtained only identification results without any estimation procedure, and none of them (Rytgaard et al., 2022) considered incremental causal effects .

Finally, the literature on causal inference until recently have paid less attention to the positivity assumption than, for example, unmeasured confounding. Petersen et al. (2012), Zhu et al. (2021), and Léger et al. (2022) have recently investigated the importance and implications of this often-overlooked assumption. Hwang et al. (2024) and Matsouaka & Zhou (2024) aimed at relaxing the positivity assumption. Our paper here aims at a new class of estimands which completely avoid positivity, and this in turn opens up more causal questions that one can answer in a study.

3 PROPOSED APPROACH

3.1 ESTIMAND AND INVERSE PROBABILITY WEIGHTED IDENTIFICATION

Suppose we observe $\{Y, T \wedge \tau, \Delta = \mathbb{1}(T < \tau), L\}$, where Y is an outcome of interest measured at time τ, T is time to (initializing) a certain treatment, and L are baseline covariates. We use $\mathbb{P}$ and $\mathbb{E}$ to represent the distribution of $\{Y, T \wedge \tau, \Delta = \mathbb{1}(T < \tau), L\}$ and expectation with respect to $\mathbb{P}$. We assume T to be absolutely continuous, and write $\lambda(t|l) = \lim_{h \to 0} \mathbb{P}(T < t + h | T \geq t, L)/h$ and $\Lambda(t|l) = \int_0^t \lambda(s|L) ds$ as its hazard function and cumulative hazard function at time t given $L = l$, respectively. As reviewed in the literature above, the hazard function is a natural quantity used to describe the intensity of time to treatment (Yang et al., 2018; 2020). Nonetheless we also write $f(l)$, $f(t|l)$, and $f(y|t, l)$ as the (conditional) densities or probability functions.

We may define the potential outcome Y_t (Neyman, 1923; Rubin, 1974; Holland, 1986) if the treatment were initialized at time t. Since Y is measured at time τ, whenever $t \geq \tau$ we denote $Y_t = Y_\tau = Y_\infty$ for someone who is never treated.

Unlike the previous literature, we are interested in answering the question like, "what would be the expected outcome Y if the hazard function $\lambda(t|l)$ were to be doubled?" In general, we may consider shifting the hazard by a positive quantity $\theta(t, l) > 0$. We consider this to be a natural 'hazard preserving' stochastic intervention; that is, to multiply the hazard function for time to treatment by a positive quantity. We note that when θ is a constant, we have a *proportional hazards* intervention, which has the obvious advantage of simple interpretation.The fact that it coincides with the familiar Cox model is another added advantage. More generally, we might consider recommendation of intensified disease screening for certain high risk groups described via l, or after a certain age t.

For the rest of the paper we denote $T(\theta)$ as the random draw following the hazard function $\theta(t, l) \cdot \lambda(t|l)$. For example, suppose the time to a certain treatment initialization, T, follows the Cox proportional hazards model including a single continuous covariate L, with the baseline hazard corresponding to the hazard of a Weibull random variable. Specifically, the hazard function of the time to the treatment initialization at time t given $L = l$ is $\lambda(t|l) = \lambda_0(t) \exp(0.2l)$, where the baseline hazard is $\lambda_0(t) = 0.9t^{0.5}$. Figure 1 shows the effects of shifting the hazard $\lambda(t|l)$ by a constant positive quantity θ, where we plotted the resulting hazard functions and density functions of the time to the treatment initialization $T(\theta)$ under interventions with various constant θ values.

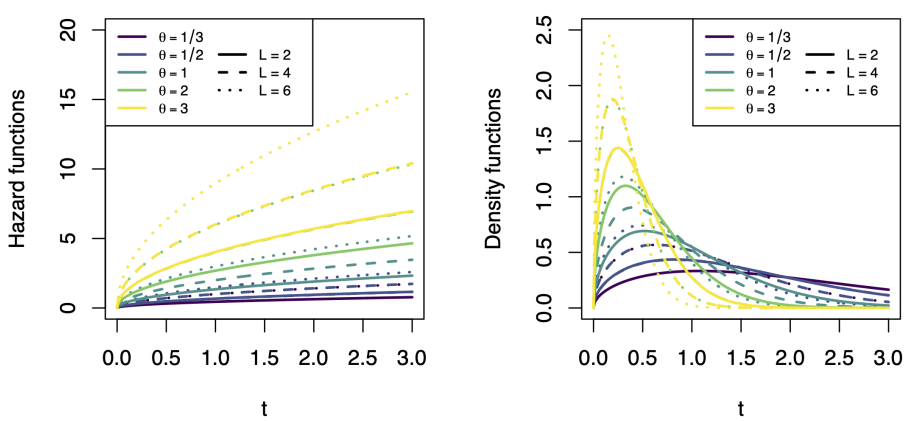

Figure 1: Hazard and density functions of the time to treatment initialization $T(\theta)$ under interventions with various constant θ values.

The corresponding potential outcome under $T(\theta)$ is then by plugging in $T(\theta)$ into Y_t as $Y_{T(\theta)}$. The incremental causal effect is defined as

$$\psi(\theta) = \mathbb{E}(Y_{T(\theta)}). \tag{1}$$

In particular when $\theta \equiv 1$, $\psi(1) = \mathbb{E}(Y_{T(1)}) = \mathbb{E}(Y)$ corresponds to the factual distribution of T that we have observed.

Assumption 1 (Consistency)**.**

$$Y_{T \wedge \tau} = Y.$$

This assumption links the observed outcome and the potential outcome via the treatment regime. It says that if an individual receives the treatment at time T, then his/her observed outcome Y matches Y_T. The second assumption is typically referred to as the "sequential randomization" assumption (SRA), "sequential exchangeability", or "no unmeasured confounders" assumption.

Assumption 2 (No Unmeasured Confounding)**.** *The time to decide treatment and treatment decision is independent of the potential outcomes given the covariate,*

$$T \perp Y_t \mid L, \quad \forall t \in [0, \tau].$$

It requires that time to treatment initialization is determined solely by the collected baseline information L.

We prove that Equation (1) can be identified:

Theorem 1 (IPW Identification)**.** *Under Assumptions 1 and 2, the incremental causal effect $\psi(\theta)$ can be identified by the observed data distribution via*

$$\psi(\theta) = \mathbb{E}(Y_{T(\theta)}) = \mathbb{E}\left\{\theta(T, L)^{\Delta} e^{-\int_0^{T \wedge \tau} (\theta(t,L)-1) d\Lambda(t|L)} Y\right\}.$$

This identification generalizes the classical IPW identification because the weights $\theta(T, L)^{\Delta} e^{-\int_0^{T \wedge \tau} (\theta(t,L)-1) d\Lambda(t|L)}$ are the Radon-Nikodym derivatives (or known as likelihood ratio) between the targeted incremental intervention distribution and observed data distribution (both can be found in the appendix), as the classical inverse probability weights.

As noted in the Introduction, while positivity is a fundamental assumption in causal inference for common estimands such as the ATE, here we do not need positivity for identification of the incremental causal effect from the observed data distribution. More specifically, compared with the proof of identification here, traditional pre-specified stochatic dynamic interventions (including deterministic static interventions as special cases) operationalize by replacing $\lambda(T|L)e^{-\int_0^T \lambda(t|L)dt}$ by some pre-specified conditional distribution of T. This necessitates to impose an additional positivity assumption. Or in other words, if any person with medical history L can never receive vaccine at time t, then we certainly cannot investigate questions like what happens if they do.

3.2 Inverse Probability Weighted Estimation and Inference

Suppose we observe a random sample of size n denoted as $\{Y_i, T_i \wedge \tau, \Delta_i = \mathbb{1}(T_i < \tau), L_i\}_{i=1}^n$. Inspired by Theorem 1, if we were to know the truth $\Lambda(t|L)$, then one may plug it in and obtain an estimate of $\psi(\theta)$ by $\frac{1}{n}\sum_{i=1}^n \theta(T_i, L_i)^{\Delta_i} e^{-\int_0^{T_i \wedge \tau} \{\theta(t,L_i)-1\} d\Lambda(t|L_i)} Y_i$. However in practice, $\Lambda(t|L)$ is hardly known to an analyst and must be estimated. Write the estimate as $\hat{\Lambda}(t|L)$, we may define the plug-in IPW estimator as

$$\hat{\psi}(\theta) = \frac{1}{n}\sum_{i=1}^n \theta(T_i, L_i)^{\Delta_i} e^{-\int_0^{T_i \wedge \tau} \{\theta(t,L_i)-1\} d\hat{\Lambda}(t|L_i)} Y_i.$$

With some regularity conditions on the nuisance estimate $\hat{\Lambda}(t|L)$ given in the appendix, we may derive the following asymptotic results on $\hat{\psi}(\theta)$.

Theorem 2 (Consistency)**.** *Under Assumptions 1, 2, and the regularity condition that $\hat{\Lambda}(t|L)$ is uniformly consistent for $\Lambda(t|L)$, that is,*

$$\mathbb{P}(\sup_t |\hat{\Lambda}(t|L) - \Lambda(t|L)| > \varepsilon) \to 0$$

as $n \to \infty$, we have that $\hat{\psi}(\theta)$ converges to $\psi(\theta)$ in probability, that is, for any constant $\varepsilon > 0$,

$$\mathbb{P}(|\hat{\psi}(\theta) - \psi(\theta)| > \varepsilon) \to 0,$$

when $n \to \infty$.

Theorem 3 (Asymptotic Normality)**.** *Under Assumptions 1, 2, and the regularity conditions that $\hat{\Lambda}(t|L)$ is uniformly consistent and asymptotically linear for $\Lambda(t|L)$, that is,*

$$\mathbb{P}(\sup_t |\hat{\Lambda}(t|L) - \Lambda(t|L)| > \varepsilon) \to 0$$

and

$$\hat{\Lambda}(t|L) - \Lambda(t|L) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t, L) + o_P(1)$$

for some random process $\xi_i(t, L)$, where $o_P(1)$ converges to zero as $n \to \infty$ uniformly over t, then the root-n scaled centered difference $\sqrt{n}\{\hat{\psi}(\theta) - \psi(\theta)\}$ is asymtotically linear and thus converges to a normal variable weakly, that is,

$$\sqrt{n}\{\hat{\psi}(\theta) - \psi(\theta)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Xi_i + o_P(1) \to \mathcal{N}(0, \mathrm{Var}(\Xi_1)),$$

for some random variable Ξ, in distribution.

We use bootstrap to construct variance estimator of $\hat{\psi}(\theta)$ and build Wald-type confidence intervals.

4 Experiment result

4.1 Simulation

In this section, we investigate the finite-sample performance of our estimator. We set $\tau = 2$ and generate n i.i.d. copies of (L_i, T_i, Y_i) as follows:

$$L_i \sim \text{Unif}(0, 1),$$
$$\mathbb{P}(T_i > t|L_i) = \exp\{-\exp(0.25L_i)t\},$$
$$Y_i \sim \mathcal{N}(\exp(1 - 1.5L_i - (2 - T_i \wedge 2)), 0.5^2),$$

where we observe $\{L_i, T_i \wedge 2, \Delta_i = \mathbb{1}(T_i < 2), Y_i\}_{1 \le i \le n}$.

We fit a Cox proportional hazard model Cox (1972) to get an estimate $\hat{\Lambda}(t|l)$ because the proportional hazard assumption holds for distribution of T_i given L_i. We examine the performance of the IPW estimator $\hat{\psi}(\theta)$ and its variance estimate when $\theta(t, l) \equiv 1/3, 1/2.5, 1/2, 1/1.5, 1.5, 2, 2.5, 3$ by reporting biases, percent biases (PBIAS), empirical standard errors (SEE), average estimated standard errors (SD) based on $B = 20$ multiplier bootstrap (van der Vaart & Wellner, 1996; Kosorok, 2008), and coverage probabilities (95% CP) of Wald type 95% confidence intervals using $R = 1000$ simulated data sets of size $n = 200, 1000, 5000$. The results are given in Table 1.

As the simulation results illustrate, $\hat{\psi}(\theta)$ perform well with small biases, across different constant functions $\theta(t, l)$, thus confirming our theoretical results. As expected from theory, variance estimates approach the Monte Carlo variance as sample sizes increase. Similarly, Wald type confidence intervals of $\psi(\theta)$ attain their nominal levels as sample sizes become larger.

4.2 Evaluating Incremental Causal Effect of Methotrexate on Rheumatoid Arthritis

We illustrate our framework by estimating incremental causal effects of the anti-rheumatic therapy Methotrexate (MTX) among patients with rheumatoid arthritis, using data from Choi et al. (2002). As a disease-modifying antirheumatic drug, MTX can slow joint damage and the progress of the disease. It is one of the most effective medications for treating inflammatory types of arthritis. This, as well as its long track record and inexpensive price tag, explains why it's usually the first drug prescribed for rheumatoid arthritis. Here we consider the average of reported number of tender joints at one year of follow-up Y as the outcome, an important measure of disease progression.

Methotrexate use was recorded at each monthly clinic visit. At each visit, we classified methotrexate exposure status as ever-treated or never-treated, i.e., once a patient starts methotrexate therapy, he

Table 1: Simulation results of the IPW estimator. We report bias ($\times 10^{-2}$), percent bias (PBIAS), empirical standard error (SEE) ($\times 10^{-2}$), average estimated standard erros (SD) ($\times 10^{-2}$) and coverage probability of Wald type 95% confidence intervals (95% CP) of $\hat{\psi}(\theta)$ by $B = 20$ multiplier bootstrap, for $n = 200, 1000, 5000$ sample sizes and $R = 1000$ Monte Carlo samples.

	$\theta(t\|l) \equiv$	1/3	1/2.5	1/2	1/1.5	1.5	2	2.5	3
	$\psi(\theta)$	0.957	0.893	0.808	0.694	0.404	0.336	0.297	0.273
$n = 200$	Bias	0.813	0.679	0.396	0.121	-0.222	-0.209	-0.175	-0.124
	PBIAS	0.85	0.76	0.49	0.174	-0.549	-0.622	-0.591	-0.456
	SEE	7.298	6.755	6.16	5.536	4.517	4.468	4.609	4.834
	SD	6.854	6.412	5.947	5.474	4.584	4.495	4.583	4.754
	95% CP	91.8	92.1	92.2	92.3	94.2	93.6	93.9	93.5
$n = 1000$	Bias	0.277	0.276	0.149	0.047	-0.104	-0.117	-0.116	-0.09
	PBIAS	0.289	0.308	0.184	0.067	-0.256	-0.348	-0.39	-0.331
	SEE	3.24	3.036	2.809	2.558	2.061	2.007	2.052	2.147
	SD	3.144	2.929	2.695	2.452	2.039	2.016	2.074	2.17
	95% CP	93.4	93	92.8	92.6	92.3	92.8	93.4	93.3
$n = 5000$	Bias	0.04	0.077	-0.002	-0.04	-0.049	-0.039	-0.032	-0.008
	PBIAS	0.042	0.086	-0.002	-0.058	-0.121	-0.116	-0.107	-0.029
	SEE	1.455	1.361	1.257	1.144	0.925	0.907	0.933	0.979
	SD	1.411	1.316	1.212	1.104	0.917	0.904	0.929	0.972
	95% CP	92.9	92.8	92.2	92.8	93.7	93.8	93.2	94.2

or she was considered on therapy for the rest of the follow-up. Our analysis includes individuals who survived and were followed up more than 12 months, with 1010 patients meeting our inclusion criteria. Figure 2 shows the probability of receiving Methotrexate treatment since patients' first visit to the center, estimated by the Kaplan–Meier estimator $\hat{\mathbb{P}}(T \leq t)$, with its pointwise 95% confidence intervals. As the Figure illustrates, nearly 12% of the patients initiated MTX usage at the first visit and around 25% of the patients were treated at the 12 month visit.

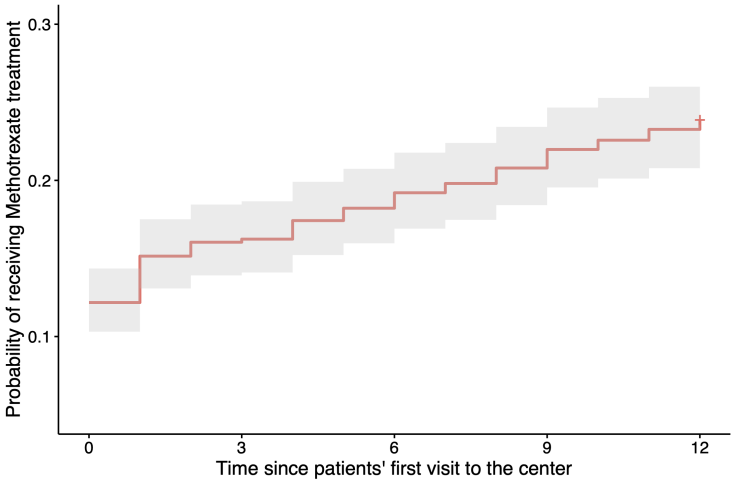

Figure 2: Estimated probability of receiving Methotrexate treatment since patients' first visit to the center $\hat{\mathbb{P}}(T \leq t)$, with pointwise 95% confidence intervals.

We used a Cox proportional hazards model (Cox, 1972) to estimate the conditional cumulative hazard function of the time to initialize Methotrexate $\Lambda(t|L = l)$, given the following variables L: age, sex, past smoking status, education level, rheumatoid arthritis duration, rheumatoid factor positive, health assessment questionnaire, patient's global assessment, erythrocyte sedimentation rate, prednisone use, and number of tender joints (measured at baseline). The proportional hazards assumption was rigorously assessed in the appendix and determined to be valid.

We consider estimands $\psi(\theta)$ by letting θ being constant functions and varying $\theta \in [0.2, 5]$. These are the incremental causal effects when the hazard rate for initiating MTX at any given time is scaled by θ across all patients. Such an analysis can illustrate how varying levels of aggressiveness or conservatism in prescribing MTX might influence average disease progression.

Plugging the estimated conditional cumulative hazards $\hat{\Lambda}(t|L)$ into our IPW estimator, we estimated the incremental treatment effect curve with $\hat{\psi}(\theta)$, where we also computed the pointwise 95% confidence intervals (with $B = 200$ multiplier bootstraps). The results are shown in Figure 3. If the hazards function of the time to Methotrexate treatment initialization were increased proportionally for all individuals, the average number of tender joints at one year of follow-up would drop. More specifically, the average number of tender joints at one year of follow-up would decrease by 1.23% from 1.90 if the hazards doubled ($\hat{\psi}(2)$=1.87, 95% CI: 1.81–1.93), and by 5.42% if the hazards were multiplied five-fold ($\hat{\psi}(5)$=1.79, 95% CI: 1.71–1.88). Conversely, the average number of tender joints would increase by 0.83% if the hazards were halved ($\hat{\psi}(0.5)$=1.91, 95% CI: 1.86–1.97), and by 1.51% if the hazards were reduced five-fold ($\hat{\psi}(0.2)$=1.92, 95% CI: 1.86–1.98). The decline in the average number of tender joints as the hazards ratio increases from 0.2 to 5 is consistent with the protective effect of Methotrexate found in Tchetgen Tchetgen et al. (2024), where the joint effects of Methotrexate use at baseline and month six on average of tender joints at month 12 of follow-up were evaluated under a marginal structural linear model and both IPW least squares and proximal recursive least-squares yielded results suggesting a protective effect of Methotrexate on disease progression.

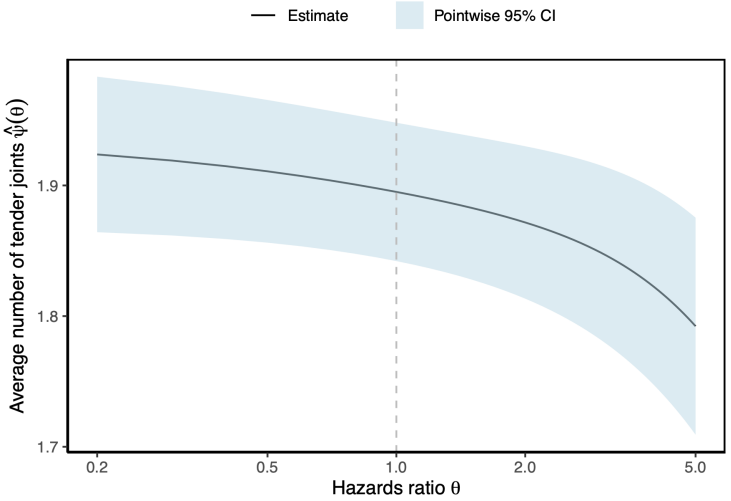

Figure 3: Estimated incremental causal effects $\hat{\psi}(\theta)$, if the hazards were multiplied by factor θ, with pointwise 95% confidence intervals.

5 Conclusion

In this paper, we extended the concept of incremental causal effects to continuous-time treatment initiation. By defining these effects through shifting the hazard function via the hazard ratio, we demonstrated that they can be identified using only sequential randomization and consistency, without requiring the positivity assumption. We developed a flexible inverse probability weighted estimator, accommodating various hazard function estimators, and provided root-n inference. The finite-sample performance of our framework and its potential real data value were assessed.

There are many open questions. First, one can consider augmenting the inverse probability weighted estimator through semiparametric theory to attain efficiency, using nonparametric estimation for nuisance parameters. This is currently under our investigation. Second, the idea can also be extended to complex longitudinal studies, including time-varying confounders and survival outcomes, or more generally, marked point processes (Rytgaard et al., 2022), càdlàg processes (Ying, 2022a). Third,

one might consider when Assumption 2 fails but either instrumental variable assumption (Ying et al., 2019; Ying & Tchetgen Tchetgen, 2023; Ying, 2022b) or proxy variable assumption (Miao et al., 2018; Tchetgen Tchetgen et al., 2024; Ying et al., 2022; 2023) holds. Last, conditional incremental effect (McClean et al., 2024) can also be of interest.

REFERENCES

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Matteo Bonvini, Alec McClean, Zach Branson, and Edward H. Kennedy. Incremental causal effects: An introduction and review. In *Handbook of Matching and Weighting Adjustments for Causal Inference*, pp. 349–372. Chapman & Hall/CRC, 2023.

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.

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

Iván Díaz and Nima S Hejazi. Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):661–683, 2020.

Iván Díaz and Mark J van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2012.

Iván Díaz and Mark J van der Laan. Assessing the causal effect of policies: an example using stochastic interventions. *The international journal of biostatistics*, 9(2):161–174, 2013.

Iván Díaz, Katherine L Hoffman, and Nima S Hejazi. Causal survival analysis under competing risks using longitudinal modified treatment policies. *Lifetime Data Analysis*, pp. 1–24, 2023a.

Iván Díaz, Nicholas Williams, Katherine L Hoffman, and Edward J Schenck. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 118(542):846–857, 2023b.

Alexander D'Amour, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics*, 221(2):644–654, 2021.

Sebastian Haneuse and Andrea Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32(30):5260–5277, 2013.

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

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

Inwoo Hwang, Yesong Choe, Yeahoon Kwon, and Sanghack Lee. On positivity condition for causal inference. In *Forty-first International Conference on Machine Learning*, 2024.

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.

Edward H Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.

Kwangho Kim, Edward H Kennedy, and Ashley I Naimi. Incremental intervention effects in studies with dropout and many timepoints#. *Journal of Causal Inference*, 9(1):302–344, 2021.

Michael R Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*, volume 61. Springer, 2008.

Maxime Léger, Arthur Chatton, Florent Le Borgne, Romain Pirracchio, Sigismond Lasocki, and Yohann Foucher. Causal inference in case of near-violation of positivity: comparison of methods. *Biometrical Journal*, 64(8):1389–1403, 2022.

Roland A Matsouaka and Yunji Zhou. Causal inference in the absence of positivity: The role of overlap weights. *Biometrical Journal*, 66(4):2300156, 2024.

Alec McClean, Zach Branson, and Edward H. Kennedy. Nonparametric estimation of conditional incremental effects. *Journal of Causal Inference*, 12(1), 2024.

Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

Erica E. M. Moodie, Thomas S. Richardson, and David A. Stephens. Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455, 2007.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):331–366, 2003.

Ashley I Naimi, Jacqueline E Rudolph, Edward H Kennedy, Abigail Cartus, Sharon I Kirkpatrick, David M Haas, Hyagriv Simhan, and Lisa M Bodnar. Incremental propensity score effects for time-fixed exposures. *Epidemiology (Cambridge, Mass.)*, 32(2):202, 2021.

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

Xinkun Nie, Emma Brunskill, and Stefan Wager. Learning when-to-treat policies. *Journal of the American Statistical Association*, 116(533):392–409, 2021.

Maya L Petersen, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J Van Der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, 2012.

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. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, pp. 189–326. Springer, 2004.

Paul R. Rosenbaum. *Observational Studies*. Springer, 2002.

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.

Aaron L Sarvet, Kerollos N Wanis, Jessica G Young, Roberto Hernandez-Alejandro, and Mats J Stensrud. Longitudinal incremental propensity score interventions for limited resource settings. *Biometrics*, 2023.

M. J. Stensrud, J. D. Laurendeau, and A. L. Sarvet. Optimal regimes for algorithm-assisted human decision-making. *Biometrika*, 00(0):1–20, 2024.

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

Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal inference. *Statistical Science*, 39(3):375–390, 2024.

Mark J van der Laan, Sherri Rose, Iván Díaz, and Mark J van der Laan. Stochastic treatment regimes. *Targeted learning in data science: Causal inference for complex longitudinal studies*, pp. 219–232, 2018.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996. ISBN 978-0-387-94640-6.

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

Shu Yang, Anastasios A Tsiatis, and Michael Blazing. Modeling survival distribution as a function of time to treatment discontinuation: A dynamic treatment regime approach. *Biometrics*, 74(3): 900–909, 2018.

Andrew Ying. Causal inference for complex continuous-time longitudinal studies. *arXiv preprint arXiv:2206.12525*, 2022a.

Andrew Ying. Structural cumulative survival models for robust estimation of treatment effects accounting for treatment switching in randomized experiments. *arXiv preprint arXiv:2204.13219*, 2022b.

Andrew Ying. Causality for functional longitudinal data. In *Causal Learning and Reasoning*, pp. 665–687. PMLR, 2024a.

Andrew Ying. Causality for complex continuous-time functional longitudinal studies with dynamic treatment regimes. *arXiv preprint arXiv:2406.06868*, 2024b.

Andrew Ying and Eric J Tchetgen Tchetgen. Structural cumulative survival models for estimation of treatment effects accounting for treatment switching in randomized experiments. *Biometrics*, 79(3):1597–1609, 2023.

Andrew Ying, Ronghui Xu, and James Murphy. Two-stage residual inclusion for survival data and competing risks—an instrumental variable approach with application to seer-medicare linked data. *Statistics in Medicine*, 38(10):1775–1801, 2019.

Andrew Ying, Yifan Cui, and Eric J. Tchetgen Tchetgen. Proximal causal inference for marginal counterfactual survival curves. *arXiv preprint arXiv:2204.13144*, 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.

Jessica G Young, Miguel A Hernán, and James M Robins. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic methods*, 3(1):1–19, 2014.

Yaqian Zhu, Rebecca A Hubbard, Jessica Chubak, Jason Roy, and Nandita Mitra. Core concepts in pharmacoepidemiology: Violations of the positivity assumption in the causal analysis of observational data: Consequences and statistical approaches. *Pharmacoepidemiology and drug safety*, 30(11):1471–1485, 2021.

A PROOFS

A.1 PROOF OF THEOREM 1

We have

$$
\begin{aligned}
\psi(\theta) &= \mathbb{E}(Y_{T(\theta)}) \\
&= \mathbb{E}[\mathbb{E}\{\mathbb{E}(Y_{T(\theta)}|T(\theta), L)|L\}] \\
&= \int \mathbb{E}(Y_t|T(\theta) = t, L = l) f(l) \{\theta(t|l)\lambda(t|l) e^{-\int_0^t \theta(s|l)\lambda(s|l)ds}\}^{\delta} \\
&\quad \{\mathbb{E}(Y_\tau|T(\theta) \geq \tau, L = l) e^{-\int_0^\tau \theta(t|l)\lambda(t|l)dt}\}^{1-\delta} dtdld\delta \\
&= \int \mathbb{E}(Y_t|L = l) f(l) \{\theta(t|l)\lambda(t|l) e^{-\int_0^t \theta(s|l)\lambda(s|l)ds}\}^{\delta} \\
&\quad \{\mathbb{E}(Y_\tau|L = l) e^{-\int_0^\tau \theta(t|l)\lambda(t|l)dt}\}^{1-\delta} dtdld\delta \\
&= \int \mathbb{E}(Y_t|T = t, L = l) f(l) \{\theta(t|l)\lambda(t|l) e^{-\int_0^t \theta(s|l)\lambda(s|l)ds}\}^{\delta} \\
&\quad \{\mathbb{E}(Y_\tau|T \geq \tau, L = l) e^{-\int_0^\tau \theta(t|l)\lambda(t|l)dt}\}^{1-\delta} dtdld\delta \\
&= \int \mathbb{E}(Y|T = t, L = l) f(l) \{\theta(t|l)\lambda(t|l) e^{-\int_0^t \theta(s|l)\lambda(s|l)ds}\}^{\delta} \\
&\quad \{\mathbb{E}(Y|T \geq \tau, L = l) e^{-\int_0^\tau \theta(t|l)\lambda(t|l)dt}\}^{1-\delta} dtdld\delta \\
&= \mathbb{E}\left\{\theta(T, L)^{\Delta} e^{-\int_0^{T\wedge\tau}(\theta(t,L)-1)d\Lambda(t|L)} Y\right\}.
\end{aligned}
$$

where the first equation follows by definition, the second by iteration of expectation, the third by definition of expectation, the fourth by the definition of $T(\theta)$, the fifth by Assumption 2, the sixth by Assumption 1, and the last by the Radon-Nikodym Theorem.

Another intuition of checking this is through densities: The observed data density is

$$f(Y, T, L) = f(L)\{\lambda(T|L)e^{-\int_0^T \lambda(t|L)dt} f(Y|T, L)\}^{\Delta}\{e^{-\int_0^\tau \lambda(t|L)dt} f(Y|T > \tau, L)\}^{1-\Delta}.$$

The target incremental intervention density is

$$f(Y, T, L) = f(L)\{\theta(T|L)\lambda(T|L)e^{-\int_0^T \theta(t|L)\lambda(t|L)dt} f(Y|T, L)\}^{\Delta}\{e^{-\int_0^\tau \theta(t|L)\lambda(t|L)dt} f(Y|T \geq \tau, L)\}^{1-\Delta}.$$

The density for the potential outcomes is

$$f(Y_{T(\theta)}, T, L) = f(L)\{\theta(T|L)\lambda(T|L)e^{-\int_0^T \theta(t|L)\lambda(t|L)dt} f(Y_T|L)\}^{\Delta}\{e^{-\int_0^\tau \theta(t|L)\lambda(t|L)dt} f(Y_\tau|L)\}^{1-\Delta}.$$

Integrating $Y_{T(\theta)}$ yields the parameter of interest (1). By Assumptions 1 and 2, the density for the potential outcomes is equal to the target incremental intervention density. Therefore, to identify (1) through the observed data density, one can use Radon-Nikodym weights to reweight Y.

A.2 PROOF OF THEOREM 2

We have

$$
\begin{aligned}
&= \left|\hat{\psi}(\theta) - \psi(\theta)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n} \theta(T_i, L_i)^{\Delta_i} e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\hat{\Lambda}(t|L_i)} Y_i - \psi(\theta)\right| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^{n} \theta(T_i, L_i)^{\Delta_i} \left\{e^{-\int_0^{T_i\wedge\tau}(\theta(t,L)-1)d\hat{\Lambda}(t|L_i)} - e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\Lambda(t|L_i)}\right\} Y_i\right| \\
&+ \left|\frac{1}{n}\sum_{i=1}^{n} \theta(T_i, L_i)^{\Delta_i} e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\Lambda(t|L_i)} Y_i - \psi(\theta)\right| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^{n} \theta(T_i, L_i)^{\Delta_i} \left\{e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\hat{\Lambda}(t|L_i)} - e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\Lambda(t|L_i)}\right\} Y_i\right| + o_P(1).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\left| e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\hat{\Lambda}(t|L_i)} - e^{-\int_0^{T_i\wedge\tau}(\theta(t,L_i)-1)d\Lambda(t|L_i)} \right| \\
&= \left| e^{-\theta(T_i\wedge\tau,L_i)\hat{\Lambda}(T_i\wedge\tau,L_i)+\int_0^{T_i\wedge\tau}\hat{\Lambda}(t|L_i)d\theta(t,L_i)} - e^{-\theta(T_i\wedge\tau,L_i)\Lambda(T_i\wedge\tau,L_i)+\int_0^{T_i\wedge\tau}\Lambda(t|L_i)d\theta(t,L_i)} \right| \\
&\leq \Bigg| \left\{ e^{-\theta(T_i\wedge\tau,L_i)\Lambda(T_i\wedge\tau,L_i)+\int_0^{T_i\wedge\tau}\Lambda(t|L_i)d\theta(t,L_i)} + o_P(1) \right\} \\
&\left\{ \theta(T_i\wedge\tau,L_i)\left|\hat{\Lambda}(T_i\wedge\tau,L_i)-\Lambda(T_i\wedge\tau,L_i)\right| + \int_0^{T_i\wedge\tau}\left|\hat{\Lambda}(t|L_i)-\Lambda(t|L_i)\right| d\theta(t,L_i) \right\} \Bigg| \\
&\leq C \sup_t \sup_{L_i} \left|\hat{\Lambda}(t|L_i)-\Lambda(t|L_i)\right|.
\end{aligned}
$$

A.3 Proof of Theorem 3

By a simple Taylor expansion, we have

$$
\hat{\psi}(\theta)-\psi(\theta) = \frac{1}{n}\sum_{i=1}^{n}\Xi_i(\delta) = \frac{1}{n}\sum_{i=1}^{n}\theta^{\Delta_i}\hat{\mathbb{P}}(T_i > t|L_i\}^{\theta-1}|_{t=X_i}Y_i + \frac{\partial\psi(\theta)}{\partial\Lambda(t)}\xi_i(t,L)|_{t=X_i} - \psi(\theta),
$$

and therefore by central limit theorem we have reached the conclusion. □

B Methotrexate Data Processing

The raw dataset included the information of 1240 patients. For all the patients, we extracted 11 baseline covariates described in Section 4.2, as well as T, the time to receive Methotrexate (MTX) treatment, and the number of tender joints at month $\tau = 12$, which is the outcome Y of interest. Some patients exited the study before reaching month $\tau = 12$, i.e., one year of follow-up, resulting in no recorded Y for these individuals. After excluding them, 1010 patients met our inclusion criteria.

Denote $L_{1,i}$ to $L_{11,i}$ as the baseline covariates, T_i as the time to receive MTX, and Y_i as the number of tender joints at one year of follow-up for each patient ($i = 1, 2, \ldots, 1010$). Define $\Delta_i = \mathbb{1}(T_i < \tau)$. For the ith patient, the individual data consists of $\{Y_i, T_i \wedge \tau, \Delta_i, L_{1,i}, \ldots, L_{11,i}\}$.

For constructing the Cox proportional hazards model, we tested the proportional hazards assumption by examining the Schoenfeld residuals and cumulative martingale residuals. For simplicity, we used age_0, sex, smoke_0, edu_0, duration_0, rapos, haqc_0, gsc_0, esrc_0, onprd2_0, and jc_0 as the abbreviations for the covariates age, sex, past smoking status, education level, rheumatoid arthritis duration, rheumatoid factor positive, health assessment questionnaire, patient's global assessment, erythrocyte sedimentation rate, prednisone use, and number of tender joints (measured at baseline), respectively. Figure 4 gives the Schoenfeld residuals plots of time-dependent Cox regression and the results of Schoenfeld individual tests and global Schoenfeld test. Table 2 shows the results of the test of proportionality for fitting an Cox-Aalen model. Figure 5 gives the cumulative martingale residuals plots for each covariate. The results show that the Cox proportional hazards model we constructed is appropriate.

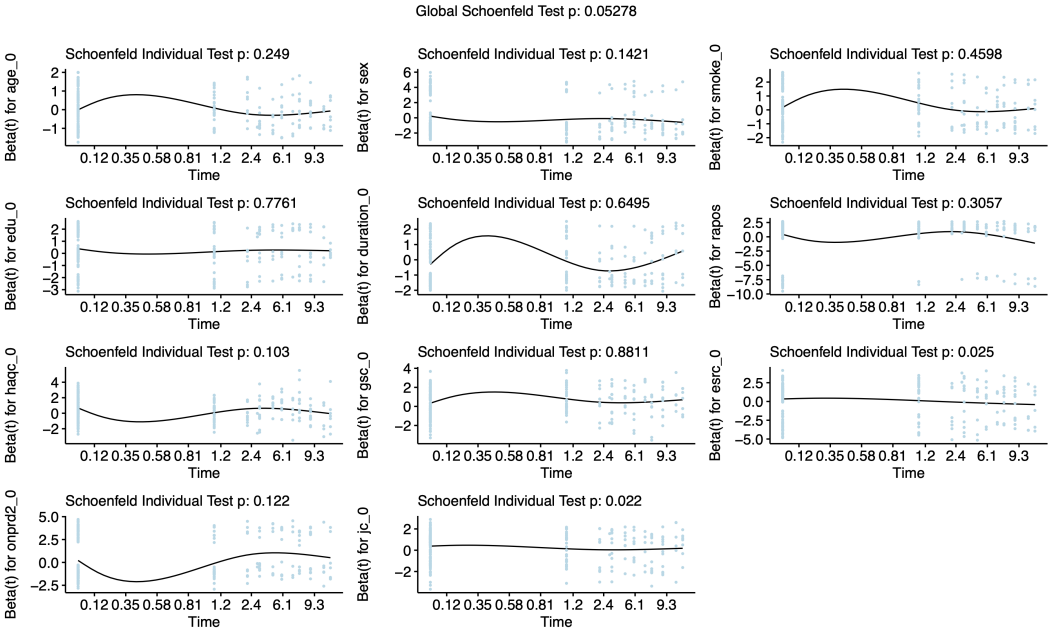

Figure 4: Results of Schoenfeld residuals examination.

Table 2: Results of the test of proportionality for fitting an Cox-Aalen model.

	$\sup \lvert\hat{U}(t)\rvert$	p-value
age_0	16.80	0.158
sex	6.30	0.248
smoke_0	10.30	0.406
edu_0	5.52	0.932
duration_0	10.60	0.306
rapos	4.68	0.294
haqc_0	9.52	0.320
gsc_0	6.28	0.774
esrc_0	8.40	0.078
onprd2_0	7.98	0.140
jc_0	14.00	0.094

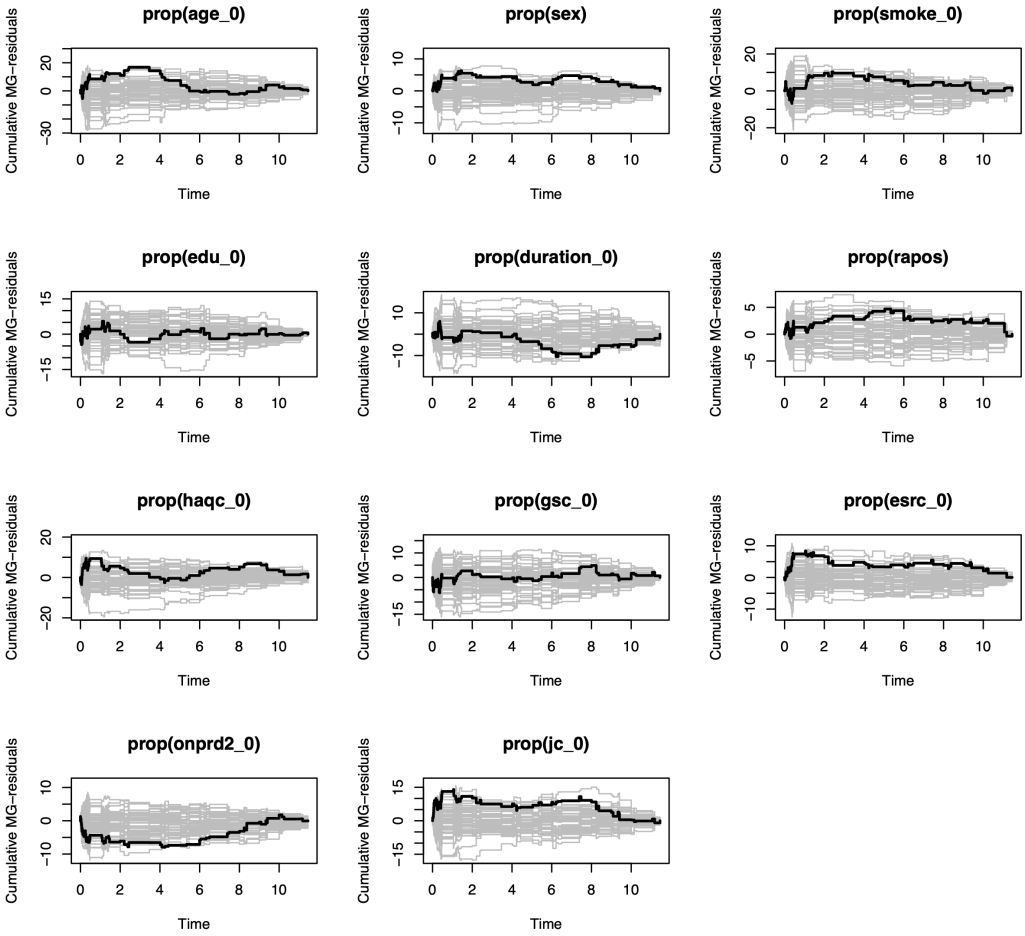

Figure 5: Cumulative martingale residuals plots for each covariate.