Estimating Treatment Effects for Recurrent Events in the Presence of Rescue Medications: An Application to the Immune Thrombocytopenia Study

Fei Gao[1] · Donglin Zeng[1] · Helen Wei[2] · Xuena Wang[2] · Joseph G. Ibrahim[1]

Received: 19 July 2016 / Accepted: 9 August 2016 / Published online: 29 August 2016

Abstract In many clinical studies, patients may experience the same type of event of interest repeatedly over time. However, the assessment of treatment effects is often complicated by the rescue medication uses due to ethical reasons. For example, in the motivating trial in studying the Immune Thrombocytopenia (ITP), when the interest lies in evaluating the treatment benefit of investigational product (IP) on reducing patient's repeated bleeding, rescue medication such as platelet transfusions may be allowed to raise platelet counts. Both the intention-to-treat analysis and treating the intermediate rescue medication as covariate tend to attenuate the treatment benefit, and the estimates can be biased if interpreted as causal. In this paper, we propose a general causal framework when intermediate rescue medications are informative. We adopt the inverse weighted estimation approach to estimate the treatment effect, where weights are constructed to reflect time-dependent medication use probabilities. The proposed estimators are shown to be asymptotically normal and are demonstrated to perform well in small-sample simulation studies. The application to the ITP studies reveals a stronger benefit of using IP in reducing bleeding.

Keywords Causal effect · Immune thrombocytopenia · Inverse weighted estimating equations · Recurrent events · Rescue medication

✉ Joseph G. Ibrahim
ibrahim@bios.unc.edu

[1] Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA

[2] Global Biostatistical Science, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA

1 Introduction

In many clinical studies, patients may experience the same type of event of interest repeatedly over time, which are referred as recurrent events. Examples of recurrent events include bleeding in studies of Immune Thrombocytopenia (ITP), transient myocardial ischemia in cardiovascular trials, and fractures in studies of osteoporosis. In practice, the investigator is often interested in the effect of covariates on the recurrent event. In order to model the effect, various models and methods have been proposed. The most popular types of methods include modeling the gap times between events [6], modeling the marginal hazards for individual recurrences [11], or modeling the intensity functions of the recurrent event process [1].

In our motivating case study of ITP, patients are randomized to receive investigational product (IP), placebo, or standard care. The interest lies in evaluating the treatment benefit of IP in reducing recurrent bleedings among ITP patients. However, the assessment of treatment effect is complicated by the use of rescue medications such as platelet transfusions. When the platelet counts are low for some subject, the rescue medication is likely to be applied with an intention of raising platelet counts for ethical reasons. Therefore, the rescue medication serves as an additional time-dependent treatment, which is affected by some time-dependent confounders such as platelet counts and previous bleeding and medication history. The primary goal of our proposed methodology is to estimate the causal effect of IP on reducing occurrence of recurrent bleeding events after balancing rescue medication use and other time-dependent confounders.

The marginal structure model (MSM) [3,7,8,10] is a widely used tool for estimating causal effect of a time-dependent exposure in the presence of time-dependent confounders. In contrast to the g-estimation in structural nested models (SNMs), MSM adopted the inverse-probability-of-treatment weighted (IPTW) estimators, where each observation is assigned a weight that is inversely proportional to the probability of treatment given the time-dependent confounders and previous treatment. It avoids over-adjustment of confounder by separating confounder control from the structural model [4]. MSM with survival outcome has been considered in previous literature [3,12], and has been applied to various clinical studies [2,5]. However, MSM for recurrent survival outcome has not been developed.

In this paper, we develop a general framework to infer the true treatment effect by balancing intermediate time-dependent medication. Our method is based on the inverse probability weight estimating equations where the time-dependent probability is the probability of receiving the rescue medication given the past event history. Due to random occurrence of interventions, we adopt a pooled data analysis to estimate this probability. The paper is constructed as follows. In Sect. 2 we give a detailed description of our method. The asymptotic results are given in Sect. 3. In Sect. 4, we conduct extensive simulation studies to examine the small-sample performance of the proposed method. In Sect. 5 we apply our method to our motivating study of ITP. A discussion follows in Sect. 6.

2 Method

2.1 Models and Assumptions

Let X denote the baseline covariates including treatment status, which is randomized and independent of the counterfactual outcomes. Let $\overline{R}(t)$ denote the medication use process up to time t. We use $\mathrm{d}N(t; x, \overline{r}(t))$ to denote the counterfactual outcome of the event process at $[t, t + \mathrm{d}t)$ if patients had baseline information $X = x$ and have medication use process $\overline{R}(t) = \overline{r}(t)$. We assume that the counterfactual outcome follows a proportional intensity model such that

$$E\left[\mathrm{d}N\left\{t; x, \overline{r}(t)\right\}\right] = \exp\left[\beta x + \gamma\phi\left\{\overline{r}(t)\right\}\right]\mathrm{d}\Lambda(t), \tag{1}$$

where $\phi(\cdot)$ is a functional of medication use $\overline{r}(t)$ and $\mathrm{d}\Lambda(t)/\mathrm{d}t$ is the baseline rate function. The parameter β, which is our point of interest, denotes the causal effect of X after balancing medication use. Here we assume a proportional intensity model for the recurrent event since the occurrence rate is of interest. The conditional independence of recurrent event and censoring time given the covariates is assumed.

In order to estimate the causal effect β, we impose the following assumptions.

(A1) The outcome of the event process at $[t, t + \mathrm{d}t)$, $N(t)$, satisfies $\mathrm{d}N(t) = \sum_{x,\overline{r}(t)} I\{X = x, \overline{R}(t) = \overline{r}(t)\}\mathrm{d}N\left\{t; x, \overline{r}(t)\right\}$.

(A2) The counterfactual outcome $\{\mathrm{d}N(t; x, \overline{r}(t)) : x, \overline{r}(t)\}$ is independent of $I(R(s) = r)$ given X and $\overline{W}(s-)$ for any $s \leq t$, where $R(s)$ denotes the process of medication use and $\overline{W}(s-)$ denotes the observed history up to time $s-$, including event history, medication use history, and other biomarker information.

Condition (A1) is the causal consistency assumption normally assumed in causal inference. Condition (A2) is equivalent to unobserved confounders assumption in causal inference; that is, whether subjects receive medication or not at time s is independent of their potential bleeding at any time after s given all available information up to time s.

2.2 Inference Procedure

Without loss of generality, we assume $\phi(\overline{r}(t)) = C(t)$, which is the cumulative medication use up to time t. We denote $p\left\{\overline{R}(t); X, \overline{W}(t-)\right\}$ as the conditional probability of the observed medication use up to time t given all available information up to time t. It is the product of the conditional probability of medication use at time s given all available information up to time s, for all $s \leq t$, i.e.,

$$p\left\{\overline{R}(t); X, \overline{W}(t-)\right\} = \prod_{s \leq t} P\left\{R(s) = r(s) | X, \overline{W}(s-)\right\}.$$

Furthermore, we use δN_k to denote binary outcome of having bleeding at time t_k, and similarly we use $\delta\Lambda_k$ to denote the jump of the baseline rate at t_k. Let $f(X)$ be a function of X. It follows from the derivation in the appendix that

$$E\left[\frac{f(X)}{p(\overline{R}(t);X,\overline{W}(t-))}\{\mathrm{d}N(t)-\exp\{\beta X+\gamma C(t)\}\mathrm{d}\Lambda(t)\}\,\Big|\,X=x,\overline{R}(t)=\overline{r}(t)\right]=0. \tag{2}$$

Therefore, we can construct estimating equations for β, γ, and the jump sizes of Λ as

$$\sum_{i=1}^{n}\int\begin{pmatrix}X_i\\C_i(t)\\I(t\le s)\end{pmatrix}Y_i(t)\Omega_i(t)\left[\mathrm{d}N_i(t)-\exp\{\beta X_i+\gamma C_i(t)\}\mathrm{d}\Lambda(t)\right]\mathrm{d}t=0 \tag{3}$$

for $s\in[0,\tau]$, where τ denotes the length of the observation period, $Y_i(t)$ is the at risk process for bleeding outcome for subject i, and

$$\Omega_i(t)=\frac{f(X_i)}{p\left\{\overline{R}_i(t);X_i,\overline{W}_i(t-)\right\}}$$

denotes the weight to adjust for medication use. When $f(X_i)=1$, the weight $\Omega_i(t)$ is the inverse of propensity score. The estimation of intervention probability $p(\overline{R}_i(t);X_i,\overline{W}_i(t-))$ may be unstable in real settings, so is the corresponding estimating equation. Therefore, we choose $f(X)=p(\overline{R}(t);X)$ to obtain the stabilized weight

$$\Omega_i(t)=\frac{p\left\{\overline{R}_i(t);X_i\right\}}{p\left\{\overline{R}_i(t);X_i,\overline{W}_i(t-)\right\}}$$

to improve stability.

To solve Eq. (3), we first obtain the estimator for baseline hazard function as

$$\mathrm{d}\widehat{\Lambda}(t)=\frac{\sum_{i=1}^{n}Y_i(t)\Omega_i(t)\mathrm{d}N_i(t)}{\sum_{i=1}^{n}Y_i(t)\Omega_i(t)\exp\{\beta X_i+\gamma C_i(t)\}}.$$

After plugging it back into (3), we then obtain the following estimating equation for (β,γ) as

$$\sum_{i=1}^{n}\int Y_i(t)\Omega_i(t)\left[\begin{pmatrix}X_i\\C_i(t)\end{pmatrix}-\frac{\sum_{j=1}^{n}(X_j,C_j(t))^{\mathrm{T}}Y_j(t)\Omega_j(t)\exp\{\beta X_j+\gamma C_j(t)\}}{\sum_{i=1}^{n}Y_j(t)\Omega_j(t)\exp\{\beta X_j+\gamma C_j(t)\}}\right]\mathrm{d}N_i(t)=0.$$

Finally, to compute the weight $\Omega(t)$, we need to estimate the propensity scores $p\{\overline{R}(t);X\}$ and $p\{\overline{R}(t);X,\overline{W}(t-)\}$. To this end, we propose models for $R(t)$. We assume a transition logistic regression model

$$\log\left[\frac{p\{R(t)=1|V(t)\}}{1-p\{R(t)=1|V(t)\}}\right]=\alpha_1^{\mathrm{T}}V(t), \tag{4}$$

where $V(t) \equiv (X, W(t-))$ includes the baseline covariates and the event and medication use history up to time $t-$. To estimate $p(\overline{R}(t); X)$, we propose a working model,

$$\log\left[\frac{p\{R(t)=1|X\}}{1-p\{R(t)=1|X\}}\right] = \alpha_2^{\mathrm{T}} X,$$

which may not be correct. We estimate the probabilities by $\widehat{p}\{\overline{R}(t); V(t)\}$ and $\widehat{p}\{\overline{R}(t); X\}$, and estimate $\Omega_i(t)$ by

$$\widehat{\Omega}_i(t) = \frac{\widehat{p}(\overline{R}(t); X)}{\widehat{p}\left\{\overline{R}(t); X, \overline{W}(t-)\right\}}.$$

We then estimate (β, γ) by solving the equation

$$\sum_{i=1}^{n} \int Y_i(t)\widehat{\Omega}_i(t)\left[\begin{pmatrix} X_i \\ C_i(t)\end{pmatrix} - \frac{\sum_{j=1}^{n}(X_j, C_j(t))^{\mathrm{T}} Y_j(t)\widehat{\Omega}_j(t)\exp\{\beta X_j + \gamma C_j(t)\}}{\sum_{j=1}^{n} Y_j(t)\widehat{\Omega}_j(t)\exp\{\beta X_j + \gamma C_j(t)\}}\right] \mathrm{d}N_i(t) = 0 \qquad (5)$$

using Newton–Raphson algorithm.

As a remark, in many applications such as our motivating example, at each time t, there may be very few subjects who receive interventions, i.e., $R(t) = 1$. Therefore, estimating the intervention probability may not be stable. In our numerical implementation, we partition the follow-up time into some finite and non-overlapping intervals and assume $R(t)$ to be constant within each partition. In this way, the estimation of $\alpha = (\alpha_1, \alpha_2)^{\mathrm{T}}$ is numerically more reliable.

3 Asymptotic Properties

Denote the true values of regression parameters $\theta \equiv (\beta, \gamma)$, Λ, and α_1 as θ_0, Λ_0, and α_{01}, respectively. We denote the corresponding estimators as $\widehat{\theta}$, $\widehat{\Lambda}$, and $\widehat{\alpha}_1$. Denote $\widehat{\alpha}_2$ as the estimator for α_2, and α_2^* as the limit of $\widehat{\alpha}_2$ when n tends to infinity. Let $N_R(t)$ be the counting process associated with the medication use, and $Z(t) = (X, C(t))^{\mathrm{T}}$. To establish the asymptotic properties of $\widehat{\theta}$, we impose the following conditions.

(C.1) With positive probability, $V(t)V(t)^{\mathrm{T}}$ is full rank for some t in the support of $N_R(t)$. The medication use process $N_R(t)$ is independent of $R(t)$ and $N(t)$ given covariates.
(C.2) With positive probability, $Z(t)Z(t)^{\mathrm{T}}$ is full rank for some t. The function $\Lambda_0'(t)$, the derivative of $\Lambda_0(t)$, is bounded from zero in $[0, \tau]$.
(C.3) With probability one, $V(t)$ and $C(t)$ have finite total bounded variation in $[0, \tau]$.
(C.4) The censoring time is independent of $R(\cdot)$ and $N(\cdot)$ given covariates.

All these conditions are standard to ensure the identifiability of the parameters for α, θ and $\Lambda(\cdot)$. Under these conditions, the following theorem holds.

Theorem 1 *Assume that model* (1) *and* (4) *hold. Under conditions* (C.1)–(C.4), $\widehat{\theta}$ *is consistent, and* $\sqrt{n}(\widehat{\theta} - \theta_0)$ *converges in distribution to a mean zero normal distribution with covariance matrix* $\Sigma = A^{-1}E\{I(O;\theta_0,\alpha_0)^{\otimes 2}\}A^{-1}$, *where* O *denotes the observation from a single subject, and* A *and* $I(O;\theta,\alpha)$ *is defined in the appendix.*

By Theorem 1, the covariance matrix of the estimator $\widehat{\theta}$ *can be consistently estimated as* $\widehat{A}^{-1}\widehat{I}(O;\theta_0,\alpha_0)^{\otimes 2}\widehat{A}^{-1}$.

4 Simulation Studies

We considered a simulation study with n independent subjects. For subject $i = 1,\ldots,n$, two independent baseline covariates $X_i = (X_{i1}, X_{i2})$ were simulated where X_{i1} is from Bernoulli (0.5) to represent treatment assignment and X_{i2} is from Uniform [0, 1]. A time-dependent covariate $U_i(t)$, which represents the value of some auxiliary marker, was generated in (0, 5] as piecewise constant in each time interval (0, 1], (1, 2], (2, 3], (3, 4], (4, 5] with $U_i(t) = -X_{i1} + X_{i2} + b_i + \epsilon_{ik}$ if $t \in (k-1, k]$. Here b_i is a subject-specific standard normal random variable and ϵ_{ik} is standard normal random variable independent of b_i.

We generated a censoring time from the discrete uniform distribution [1, 5]. To imitate the reality that bleeding process may depend on patient's history, we generate the bleeding event with intensity function

$$\lambda_i\left(t\middle|W_{i,t-}\right) = \exp\left\{-X_{i1} + 0.5X_{i2} - 0.8\overline{N}_i(t-) - 0.8\overline{R}_i(t-) + 0.5U_i(t)\right\},$$

where $\overline{N}(t)$ and $\overline{R}(t)$ denote the total numbers of bleeding event and medication use up to time t and $W_{i,t-} = \left(X_i, Z_i(t), \overline{N}(t-), \overline{R}_i(t-)\right)$. Furthermore, we assume that the medication use can only happen right after the bleeding event with probability

$$\begin{aligned}\operatorname{logit}\left\{P\left(R_i(t) = 1 | W_{i,t-}, \mathrm{d}N_i(t) = 1\right)\right\} = -2 - 0.5X_{i1} + 2X_{i2} + \overline{N}_i(t) \\ -\overline{R}_i(t-) + 2U_i(t).\end{aligned}$$

This corresponds to the common practice such that the rescue medication may be immediately used after the bleeding.

We are interested in the true causal effect for treatment, which is the difference on the outcome of bleeding event with and without primary treatment given the same medication use during the study. However, in the simulation setting, the intensity of bleeding event depends on the previous history of bleeding and medication use. Moreover, the medication use depends on the primary treatment. Therefore, the true causal treatment effect is not analytical available. We use the Monte-Carlo approach to calculate the true value of $\theta = (\beta, \gamma)$. Specially, we use the same simulation procedure to simulate the events with relatively large sample size (n = 20,000) and estimate the causal effects using the proposed method. The Monte-Carlo method gives that the true value for θ is $(-0.204, -0.145, 0.828)$.

Table 1 Simulation results

n	(P)				(N1)		(N2)	
	Bias	SE	SEE	CP	Bias	SE	Bias	SE
$n = 200$								
β_1	−0.013	0.120	0.120	0.955	−0.098	0.101	0.382	0.123
β_2	0.023	0.184	0.190	0.947	0.093	0.165	0.186	0.233
γ	0.004	0.097	0.095	0.943	−0.104	0.099	−1.22	0.163
$n = 400$								
β_1	−0.008	0.087	0.085	0.947	−0.101	0.074	0.370	0.088
β_2	0.008	0.150	0.134	0.938	0.086	0.114	0.171	0.164
γ	−0.003	0.066	0.066	0.951	−0.117	0.067	−1.21	0.116

(P) denotes our method. (N1) and (N2) denote the two naive methods. Bias and SE are the mean bias and standard error of the parameter estimator, respectively, SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95 % confidence intervals

For each simulated dataset, we apply our method to estimate the causal effect. The variance of the estimators is then estimated using the expressions after Theorem 1. For comparison, we also implement two naive estimating methods:

(N1) we set $\widehat{\Omega}_{ik} = 1$ and use the same estimating equation as in Eq. (5).
(N2) we set $\widehat{\Omega}_{ik} = 1$, and estimate effects using the estimating equation

$$\sum_{i=1}^{n} \int_s Y_i(s) \left[\begin{pmatrix} X_i \\ \overline{R}_i(s) \\ \overline{N}_i(s) \end{pmatrix} - \frac{\sum_{j=1}^{n} \begin{pmatrix} X_i \\ \overline{R}_j(s) \\ \overline{N}_j(s) \end{pmatrix} Y_j(s) \exp\{\beta X_j + \gamma \overline{R}_j(s) + \eta \overline{N}_j(s)\}}{\sum_{j=1}^{n} Y_j(s) \exp\{\beta X_j + \gamma \overline{R}_j(s) + \eta \overline{N}_j(s)\}} \right] \mathrm{d}N_i(s) = 0.$$

The naive method (N1) estimates the intent-to-treat effects for the model adjusting for medication use. The naive method (N2) estimates the conditional treatment effect after controlling the event history in the regression. Clearly, neither of these methods estimate the true underlying causal effect.

Simulation studies are conducted with sample sizes $n = 200$ and 400. Table 1 gives the simulation results from 1000 replicates. From the table, the two naive methods which estimate the intent-to-treat effects are highly biased. The result based on naive methods does not give valid inference on the causal effect of treatment after balancing medication use. For our approach, the bias is reasonably small in small samples, and the empirical variance agrees well with the estimated variance.

When the estimated treatment probability is small, the inverse probability weight can be large to cause numerical instability. One way to improve the computation is

Table 2 Simulation results of truncation at different quantiles ($n = 400$)

Truncation (%)	Bias	SE	SEE	CP
90				
β_1	−0.035	0.077	0.076	0.925
β_2	0.022	0.119	0.116	0.938
γ	−0.022	0.058	0.062	0.943
95				
β_1	−0.030	0.077	0.076	0.932
β_2	0.022	0.119	0.117	0.939
γ	−0.018	0.059	0.062	0.944

Bias and SE are the bias and standard error of the parameter estimator, SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95 % confidence interval

to truncate weights. To evaluate the sensitivity to the truncation of inverse probability weight, we use the same simulation dataset but truncate weights at different quantiles in the estimation. The result of truncation at 90 and 95 % quantile with $n = 400$ is shown in Table 2. With 90 % truncation, the inverse probability weight still gives reasonable bias and coverage probabilities.

5 Real Data Analysis

The motivation case study is from an integrated database consisting of three phase 3, randomized, placebo-controlled clinical trials and one phase 3b, randomized, standard of care (SOC) controlled clinical trial. The trials were conducted in multiple centers globally from 2005 to 2009. The primary goal is to evaluate the treatment effect of IP on subjects with ITP, which is an autoimmune disorder characterized by low platelet counts due to increased platelet destruction and suboptimal platelet production, with clinical manifestations ranging from being asymptomatic to having serious gastrointestinal or intracranial hemorrhage. The occurrence of bleeding events caused by low platelet counts is therefore an important clinical endpoint in evaluating treatment effect of IP.

A total of 393 subjects were enrolled in the four trials, where 131 subjects received placebo or SOC and 262 subjects received IP. We exclude five subjects from the analysis because of missing values of baseline covariates, to obtain a total of 388 subjects. Table 3 shows the distribution of baseline covariates, including age, sex, indicator of whether splenectomy performed prior to enrollment, years since ITP diagnosis, and baseline platelet counts for different treatment groups. The IP group and placebo/SOC group have similar distribution for baseline covariates. During the study, platelet counts were measured weekly with IP administration. Table 4 shows the distribution of number of bleeding event and rescue medication use for the two treatment groups. Compared to the placebo/SOC group, the IP group has less bleeding event and less rescue medication use.

To estimate the probability of medication use, the study period is partitioned into intervals of 2 weeks for the first 24 weeks and an interval post 24 weeks. This ensures at least 10 medication uses in each interval to avoid sparsity. We fit logistic regression

Table 3 Descriptive statistics for the baseline covariates

Covariate	IP group		Placebo/SOC group	
	N	Mean ± SD	*N*	Mean ± SD
Age	262	54.4 ± 17.4	126	53.9 ± 18.5
Years since ITP diagnosis	262	6.1 ± 8.4	126	5.9 ± 6.7
Baseline platelet	262	24.2 ± 13.2	126	21.6 ± 13.7
Covariate	IP group		Placebo/SOC group	
	N	Percentage	*N*	Percentage
Gender				
Female	154	58.8	80	63.5
Male	108	41.2	46	36.5
Splenectomy performed prior to enrollment				
Yes	52	19.8	26	20.6
No	210	80.2	100	79.4

Table 4 Distribution of number of bleeding event and medication use

Event	IP group Mean ± SD	Placebo/SOC group Mean ± SD
Bleeding	1.73 ± 2.91	1.96 ± 2.74
Rescue medication	1.10 ± 3.31	1.33 ± 3.34

models on medication use to calculate the weights. The parameter estimates for the medication use model are shown in Table 5, where average previous medication use $\left(\overline{R}_{i,k-1}\right)$ and average previous bleeding $\left(\overline{N}_{i,k-1}\right)$ are calculated as the proportion of previous partitions that adopted medication use or had bleeding event. Clearly, IP has significantly negative effect on the occurrence of present medication use, while sex, splenectomy performance, year since ITP diagnosis, baseline platelet, platelet at each visit, and previous medication use are strong predictors.

The estimated stabilized weight $\widehat{\Omega}_i^*(t)$ ranges from 0 to 98.0. We truncate the weight $\widehat{\Omega}_i^*(t)$ to its 90 % upper quantile (0.9946) and estimate the causal treatment effect. The results are given in Table 6. The estimated bleeding intensity ratio for patients receiving IP treatment versus placebo or standard care is $\exp\{-0.723\} = 0.485$, with a 95 % confidence interval of (0.282, 0.836). This implies that IP has significant effect on reducing bleeding occurrence. In comparison, we also estimated the unadjusted treatment effect where the effect of medication is ignored. The unadjusted bleeding intensity ratio is $\exp\{-0.654\} = 0.520$, with 95 % confidence interval (0.338, 0.799). The ignorance of medication use will underestimate the causal treatment effect for IP.

We also examine if different partitions of study period will affect the results. We partition the study period into intervals of 3 weeks and repeat the estimation procedure. The parameter estimates for the medication use model are different, but the estimated causal treatment effect does not change much [0.485 with 95 % interval (0.282, 0.836)].

Table 5 Parameter estimates for the medication use model

Parameter	Estimate	SE	p value
Intercept	−2.785	0.313	<0.0001
Age	0.003	0.004	0.435
Sex	0.430	0.153	0.005
IP	−0.532	0.155	0.0006
Splenectomy performed prior to enrollment	0.818	0.174	<0.0001
Years since ITP diagnosis	−0.055	0.013	<0.0001
Baseline platelet	−0.033	0.007	<0.0001
Average previous medication $\left(\overline{R}_{i,k-1}\right)$	5.157	0.220	<0.0001
Average previous bleeding $\left(\overline{N}_{i,k-1}\right)$	0.232	0.242	0.338
Platelet	−0.002	0.001	0.008

Standard care and placebo are pooled together

Table 6 Parameter estimates for causal model

Variable	Estimate	SE	p value
Age	0.006	0.007	0.371
Sex	0.315	0.285	0.270
IP	−0.723	0.277	0.009
Splenectomy performed prior to enrollment	−0.221	0.432	0.608
Years since ITP diagnosis	0.005	0.027	0.841
Baseline platelet	−0.003	0.010	0.750
Past 8-week medication use	0.027	0.014	0.048

Standard care and placebo are pooled together

We estimated the baseline intensity function for recurrent bleeding event. Figure 1 shows the estimated intensity function for a 54-year-old male in standard care or placebo group with baseline platelet count 24, no splenectomy performed, 6 years since ITP diagnosis, while receiving no rescue medication in the entire study period.

6 Discussion

We have proposed an inverse probability weighted estimating equations to estimate treatment causal effects when there are intermediate medication use during the trial. The proposed method works well in small-sample studies and the proposed estimators have been shown to be consistent and asymptotically normal. The application to the ITP study reveals a significant treatment effect.

The causal model in (1) aims for the treatment effect after balancing the medication use trajectory, i.e., what is the treatment effect if patients receive the same medication use. This model can be reduced to the model without balancing the medication use. Then the causal effect is the average causal effects of the treatment. Other conventional

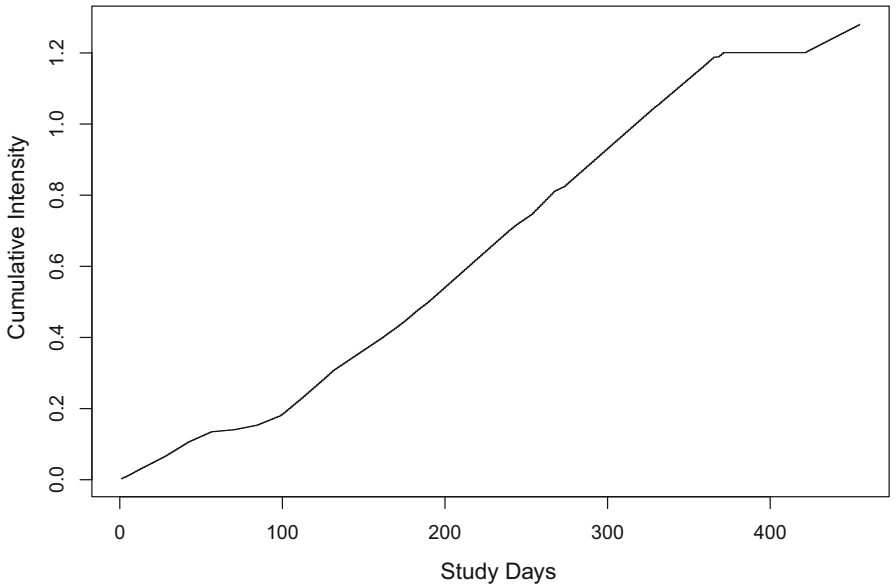

Fig. 1 Estimated intensity function

ways to obtain valid estimates for causal treatment effect with adjusted medication use include censoring the data at the first rescue medicine use, or incorporating the rescue medicine use into the endpoint definition. The former introduces informative censoring at the first rescue medication use, such that the medication use process needs to be modeled, while the latter involves more complicated causal effect interpretation.

In our estimation, we partition the follow-up period into a finite number of time interval and the propensity score is estimated as piecewise constant at each partition. A more smooth way is to use local smooth estimation, where the propensity score is estimated locally at time t by pooling observed data around time t and the pooling is governed using some kernel weights. Since the kernel smoothing methods requires frequent medication use, which is not the case in our real example, the smoothing method is not adopted.

The proposed method can be generalized to analyzing multiple types of recurrent events. Specifically, for estimating Eq. (5) it will be modified to

$$\sum_{i=1}^{n} \int Y_i(t)\widehat{\Omega}_i(t) \left[\begin{pmatrix} X_i \\ C_i(t) \end{pmatrix} - \frac{\sum_{j=1}^{n} (X_j, C_j(t))^{\mathrm{T}} Y_j(t)\widehat{\Omega}_j(t) \exp\{\beta_l X_j + \gamma_l C_j(t)\}}{\sum_{j=1}^{n} Y_j(t)\widehat{\Omega}_j(t) \exp\{\beta_l X_j + \gamma_l C_j(t)\}} \right] \mathrm{d}N_{il}(t) = 0,$$

where $N_{il}(t)$ is the event process for the lth type event for subject i and β_l are the corresponding regression coefficients. This is similar to marginal models for multivariate

survivals. We note that $\overline{W}(t-)$ can include the event history from all types of events prior to time t.

7 Derivation of Equation (2)

For discrete case, we have the conditional probability of the observed medication use given previous information,

$$p(\overline{r}(t_k);\, X,\, \overline{W}(t_k-)) = \prod_{j=1}^{k} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{j-1}),$$

which is the product of the transition probabilities. From condition (A1),

$$\begin{aligned}
& E\left[\frac{f(x)\,\{\delta N_k - \exp(\beta X + \gamma R_k)\delta\Lambda_k(t)\}}{\prod_{j=1}^{k} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{j-1})} \right. \\
& \qquad \left. \Big| X = x,\, R_0 = r_0, \ldots,\, R_k = r_k \right] \\
= & \; E\left(\frac{f(x)\left[\delta N_k(x, r_0, \ldots, r_k) - \exp\{\beta x + \gamma r(t)\}\delta\Lambda_k\right]}{\prod_{j=1}^{k} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{j-1})} \right. \\
& \qquad \left. I(R_0 = r_0, \ldots, R_k = r_k) \Big| X = x \right) \\
& \times P(R_0 = r_0, \ldots, R_k = r_k | X = x)^{-1},
\end{aligned}$$

where $\delta N_k(x, r_0, \ldots, r_k)$ is the counterfactual outcome. Now, according to condition (A2), if we denote $\delta F_k = \delta N_k(x, r_0, \ldots, r_k) - \exp\{\beta x + \gamma r(t)\}\delta\Lambda_k$, then we have

$$\begin{aligned}
& E\left\{ \frac{f(x) I(R_0 = r_0, \ldots, R_k = r_k)\delta F_k}{\prod_{j=1}^{k} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{j-1})} \,\middle|\, X = x \right\} \\
= & \; E\left[\frac{f(x) I(R_0 = r_0, \ldots, R_{k-1} = r_{k-1})}{\prod_{j=1}^{k-1} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{k-1})} \right. \\
& \times E\left\{ \frac{I(R_k = r_k)\delta F_k}{P(R_k = r_k | X,\, R_0 = r_0, \ldots,\, R_{k-1} = r_{k-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{k-1})} \right. \\
& \qquad \left. \left. \Big| X,\, R_0 = r_0, \ldots,\, R_{k-1} = r_{k-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{k-1} \right\} \Big| X = x \right] \\
= & \; E\left[\frac{f(x) I(R_0 = r_0, \ldots, R_{k-1} = r_{k-1})}{\prod_{j=1}^{k-1} P(R_j = r_j | X,\, R_0 = r_0, \ldots,\, R_{j-1} = r_{j-1},\, \widetilde{W}_0, \ldots,\, \widetilde{W}_{k-1})} \right. \\
& \times E\left\{ \frac{I(R_k = r_k)}{P(R_k = r_k | X, R_0 = r_0, \ldots,\ R_{k-1} = r_{k-1},\ \ \widetilde{W}_0, \ldots, \widetilde{W}_{k-1})} \right.
\end{aligned}$$

$$
\begin{aligned}
&\left| X, R_0 = r_0, \ldots, R_{k-1} = r_{k-1}, \widetilde{W}_0, \ldots, \widetilde{W}_{k-1} \right\} \\
&\times E\left\{ \delta F_k \middle| X, R_0 = r_0, \ldots, R_{k-1} = r_{k-1}, \widetilde{W}_0, \ldots, \widetilde{W}_{k-1} \right\} \middle| X = x \Bigg] \\
&= E\left\{ \frac{f(x) I(R_0 = r_0, \ldots, R_{k-1} = r_{k-1}) \delta F_k}{\prod_{j=1}^{k-1} P(R_j = r_j | X, R_0 = r_0, \ldots, R_{j-1} = r_{j-1}, \widetilde{W}_0, \ldots, \widetilde{W}_{k-1})} \middle| X = x \right\}.
\end{aligned}
$$

We repeat this derivation from $k-1$ to 0 to obtain

$$
\begin{aligned}
&E\Bigg(\frac{f(x) I(R_0 = r_0, \ldots, R_k = r_k)}{\prod_{j=1}^{k} P(R_j = r_j | X, R_0 = r_0, \ldots, R_{j-1} = r_{j-1}, \widetilde{W}_0, \ldots, \widetilde{W}_{j-1})} \\
&\qquad \times \left[\delta N_k(x, r_0, \ldots, r_k) - \exp\{\beta X + \gamma r(t)\} \delta \Lambda_k \right] \Bigg| X = x \Bigg) \\
&= f(x) E\left(\left[\delta N_k(x, r_0, \ldots, r_k) - \exp\{\beta X + \gamma r(t)\} \delta \Lambda_k \right] \middle| X = x \right) \\
&= f(x) E\left[\delta N_k(x, r_0, \ldots, r_k) - \exp\{\beta X + \gamma r(t)\} \delta \Lambda_k \right] \\
&= 0.
\end{aligned}
$$

We conclude that

$$
\begin{aligned}
E\Bigg[&\frac{f(x) \{\delta N_k - \exp(\beta X + \gamma R_k) \delta \Lambda_k(t)\}}{\prod_{j=1}^{k} P(R_j = r_j | X, R_0 = r_0, \ldots, R_{j-1} = r_{j-1}, \widetilde{W}_0, \ldots, \widetilde{W}_{j-1})} \\
&\left| X = x, R_0 = r_0, \ldots, R_k = r_k \right] = 0.
\end{aligned}
$$

Consequently, Eq. (2) holds.

8 Proof of Theorem 1

Let $\mathbf{P}_n$ denote the empirical measure, $\mathbf{P}$ denote the expectation, and $\mathbf{G}_n = \sqrt{n}(\mathbf{P}_n - \mathbf{P})$ denote the empirical process. We first derive the asymptotic distribution of $\widehat{\Omega}$. We define $\alpha_0 = (\alpha_{01}, \alpha_2^*)^{\mathrm{T}}$. The estimator $\widehat{\alpha} \equiv (\widehat{\alpha}_1, \widehat{\alpha}_2)^{\mathrm{T}}$ solves the estimating equations

$$
\sum_{i=1}^{n} \int V_i(t) \left\{ R_i(t) - \frac{e^{\alpha_1^{\mathrm{T}} V_i(t)}}{1 + e^{\alpha_1^{\mathrm{T}} V_i(t)}} \right\} \mathrm{d}N_R(t) = 0
$$

and

$$
\sum_{i=1}^{n} \int X \left\{ R_i(t) - \frac{e^{\alpha_2^{\mathrm{T}} X}}{1 + e^{\alpha_2^{\mathrm{T}} X}} \right\} \mathrm{d}N_R(t) = 0.
$$

It holds from (C.1) that

$$\sqrt{n}(\widehat{\alpha} - \alpha_0) = \mathbf{G}_n S_\alpha + o_p(1),$$

where $S_\alpha \equiv (S_{\alpha,1}, S_{\alpha,2})$ is the influence function for $\widehat{\alpha}$, with

$$S_{\alpha,1} = \left(E\left[\int V(t)V(t)^{\mathrm{T}} \frac{e^{\alpha_{01}^{\mathrm{T}} V(t)}}{\left\{1 + e^{\alpha_{01}^{\mathrm{T}} V(t)}\right\}^2} \mathrm{d}N_R(t) \right] \right)^{-1} \times \left[\int V(t) \left\{ R(t) - \frac{e^{\alpha_{01}^{\mathrm{T}} V(t)}}{1 + e^{\alpha_{01}^{\mathrm{T}} V(t)}} \right\} \mathrm{d}N_R(t) \right]$$

and

$$S_{\alpha,2} = \left(E\left[\int XX^{\mathrm{T}} \frac{e^{\alpha_{02}^{\mathrm{T}} X}}{\left\{1 + e^{\alpha_{02}^{\mathrm{T}} X}\right\}^2} \mathrm{d}N_R(t) \right] \right)^{-1} \left[\int X \left\{ R(t) - \frac{e^{\alpha_{02}^{\mathrm{T}} X}}{1 + e^{\alpha_{02}^{\mathrm{T}} X}} \right\} \mathrm{d}N_R(t) \right].$$

By definition, we rewrite the inverted weight $\widehat{\Omega}(t) = \Omega(t, R, V; \widehat{\alpha})$, with

$$\Omega(t, R, V; \alpha) = \exp\left(\int_0^t \left[R(s) \left\{ \alpha_2^{\mathrm{T}} X - \alpha_1^{\mathrm{T}} V(s) \right\} + \log\left\{1 + e^{\alpha_1^{\mathrm{T}} V(s)}\right\} - \log\left(1 + e^{\alpha_2^{\mathrm{T}} X}\right) \right] \mathrm{d}N_R(s) \right).$$

Therefore,

$$\sqrt{n}\left\{\Omega(t, r, v; \widehat{\alpha}) - \Omega(t, r, v; \alpha_0)\right\} = \frac{\partial \Omega(t, r, v; \alpha_0)}{\partial \alpha} \mathbf{G}_n S_\alpha + o_p(1), \tag{6}$$

with

$$\frac{\partial \Omega(t, r, v; \alpha)}{\partial \alpha} = \Omega(t, r, v; \alpha) \begin{pmatrix} \int_0^t V(s) \left\{ \dfrac{e^{\alpha_1^{\mathrm{T}} V(s)}}{1 + e^{\alpha_1^{\mathrm{T}} V(s)}} - R(s) \right\} \mathrm{d}N_R(s) \\ \int_0^t X \left\{ R(s) - \dfrac{e^{\alpha_2^{\mathrm{T}} X}}{1 + e^{\alpha_2^{\mathrm{T}} X}} \right\} \mathrm{d}N_R(s) \end{pmatrix},$$

where $o_p(1)$ is uniformly in r, v, and $t \in [0, \tau]$.

To derive the asymptotic distribution for $\widehat{\theta}$, for convenience, we define

$$g_n(t; \alpha, \theta) = \frac{\widetilde{\mathbf{P}}_n \widetilde{Z}(t) \widetilde{Y}(t) \Omega(t, \widetilde{R}, \widetilde{V}; \alpha) \exp\{\theta^{\mathrm{T}} \widetilde{Z}(t)\}}{\widetilde{\mathbf{P}}_n \widetilde{Y}(t) \Omega(t, \widetilde{R}, \widetilde{V}; \alpha) \exp\{\theta^{\mathrm{T}} \widetilde{Z}(t)\}}$$

and

$$g(t;\alpha,\theta)=\frac{\widetilde{\mathbf{P}}\widetilde{Z}(t)\widetilde{Y}(t)\Omega(t,\widetilde{R},\widetilde{V};\alpha)\exp\{\theta^{\mathrm{T}}\widetilde{Z}(t)\}}{\widetilde{\mathbf{P}}\widetilde{Y}(t)\Omega(t,\widetilde{R},\widetilde{V};\alpha)\exp\{\theta^{\mathrm{T}}\widetilde{Z}(t)\}},$$

where $\widetilde{\mathbf{P}}_n$ is the empirical expectation with respect to $(\widetilde{V}(t),\widetilde{Y}(t))$. Equation (5) can be written as

$$\mathbf{P}_n\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\widehat{\theta})\right\}\left\{\mathrm{d}N(t)-e^{\widehat{\theta}^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]=0,$$

Note that the left-hand side converges uniformly in θ in any compact set to

$$\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\alpha_0)\left\{Z(t)-g(t;\alpha_0,\theta)\right\}\left\{\mathrm{d}N(t)-e^{\theta^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right],$$

which has a unique solution at θ_0. The consistency of $\widehat{\theta}$ thus follows from Theorem 5.9 in [9].

We then obtain

$$\begin{aligned}
&\mathbf{G}_n\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\widehat{\theta})\right\}\left\{\mathrm{d}N(t)-e^{\widehat{\theta}^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&=-\sqrt{n}\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\widehat{\theta})\right\}\left\{\mathrm{d}N(t)-e^{\widehat{\theta}^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&=-\nabla_\theta\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\theta)\right\}\left\{\mathrm{d}N(t)-e^{\theta^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\Big|_{\theta=\theta^*}\\
&\quad\times\sqrt{n}(\widehat{\theta}-\theta_0)\\
&\quad-\sqrt{n}\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right],
\end{aligned}\tag{7}$$

where θ^* is some value between θ_0 and $\widehat{\theta}$. Since

$$\sqrt{n}g_n(t;\widehat{\alpha},\theta_0)=\sqrt{n}g(t;\alpha_0,\theta_0)+\widetilde{\mathbf{G}}_n\widetilde{S}(t)+o_p(1)$$

for some random variable $\widetilde{S}(t)$, we have that the last term of the right-hand side in (7) is equal to

$$\begin{aligned}
&\sqrt{n}\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g(t;\alpha_0,\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&-\widetilde{\mathbf{G}}_n\left(\int\tilde{S}(t)E\left[Y(t)\Omega(t)\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\right)+o_p(1).
\end{aligned}$$

On the other hand,

$$E\left[Y(t)\Omega(t)\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]=0$$

following from the property of the causal effect θ_0 shown in Sect. 2.1. Thus, we have

$$
\begin{aligned}
&\sqrt{n}\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&=\sqrt{n}\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g(t;\alpha_0,\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&\quad+o_p(1)\\
&=\sqrt{n}\mathbf{P}\left[\int Y(t)(\Omega(t,R,V;\widehat{\alpha})-\Omega(t,R,W;\alpha_0))\left\{Z(t)-g(t;\alpha_0,\theta_0)\right\}\right.\\
&\quad\left.\times\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]+o_p(1)\\
&=\mathbf{P}\left[\int Y(t)\frac{\partial\Omega(t,R,V;\alpha_0)}{\partial\alpha}\left\{Z(t)-g(t;\alpha_0,\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&\quad\mathbf{G}_nS_\alpha+o_p(1),
\end{aligned}
$$

where the last step follows from (6).

Combining all the results, we obtain

$$
\begin{aligned}
&\mathbf{G}_n\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\widehat{\theta})\right\}\left\{\mathrm{d}N(t)-e^{\widehat{\theta}^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&=-\nabla_\theta\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\widehat{\alpha})\left\{Z(t)-g_n(t;\widehat{\alpha},\theta^*)\right\}\left\{\mathrm{d}N(t)-e^{\theta^{*\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&\quad\sqrt{n}(\widehat{\theta}-\theta_0)\\
&\quad-\mathbf{P}\left[\int Y(t)\frac{\partial\Omega(t,R,V;\alpha_0)}{\partial\alpha}\left\{Z(t)-g(t;\alpha_0,\theta_0)\right\}\left\{\mathrm{d}N(t)-e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}\right]\\
&\quad\mathbf{G}_nS_\alpha+o_p(1).
\end{aligned}
$$

Using condition (C.3), it is easy to verify the Donsker property of the class of the functions on the left-hand side, which takes the form of

$$
\int Y(t)\Omega(t,R,V;\alpha)\{Z(t)-b(t)\}\left\{\mathrm{d}N(t)-e^{\theta^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\},
$$

where $b(t)$ has a finite total variation in $[0,\tau]$, θ is in a neighborhood of θ_0, and $\Omega(t,R,V;\alpha)$ is Lipschitz continuous in α. Therefore, we conclude

$$
A\sqrt{n}(\widehat{\theta}-\theta_0)=\mathbf{G}_nI(O;\theta_0,\Lambda_0,\alpha_0)+o_p(1),
$$

where

$$
A=\mathbf{P}\left[\int Y(t)\Omega(t,R,V;\alpha_0)\{Z(t)-g(t;\alpha_0,\theta_0)\}^{\otimes 2}e^{\theta_0^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right]
$$

and

$$I(O;\theta,\alpha) = \int Y(t)\Omega(t,R,V;\alpha_0)\{Z(t) - g(t;\alpha_0,\theta_0)\}\left\{\mathrm{d}N(t) - e^{\theta^{\mathrm{T}}Z(t)}\mathrm{d}\Lambda_0(t)\right\}$$
$$+ \widetilde{\mathbf{P}}\left[\int \widetilde{Y}(t)\frac{\partial\Omega(t,\widetilde{R},\widetilde{V};\alpha_0)}{\partial\alpha}\left\{\widetilde{Z}(t) - g(t;\alpha_0,\theta_0)\right\}\right.$$
$$\left.\left\{\mathrm{d}\widetilde{N}(t) - e^{\theta_0^{\mathrm{T}}\widetilde{Z}(t)}\mathrm{d}\Lambda_0(t)\right\}\right] S_\alpha.$$

By condition (C.2), A is non-singular. Therefore, we conclude that $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in distribution to a normal random variable with mean zero and covariance matrix $A^{-1}E\{I(O;\theta_0,\Lambda_0,\alpha_0)^{\otimes 2}\}A^{-1}$. Finally, to estimate the asymptotic variance, we replace $\mathbf{P}$ or $\tilde{\mathbf{P}}$ by the corresponding empirical averages and all the parameters by their estimators in A and $I(O;\theta_0,\Lambda_0,\alpha_0)$ to obtain $\widehat{A}$ and $\widehat{I}(O;\widehat{\theta},\widehat{\Lambda},\widehat{\alpha})$. Then, a consistent variance estimator is given by

$$\widehat{A}^{-1}\left\{n^{-1}\sum_{i=1}^{n}\widehat{I}(O_i;\widehat{\theta},\widehat{\Lambda},\widehat{\alpha})^{\otimes 2}\right\}\widehat{A}^{-1}.$$

References

1. Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. Ann Stat 10(4):1100–1120
2. Cole SR, Hernán MA, Robins JM, Anastos K, Chmiel J, Detels R, Ervin C, Feldman J, Greenblatt R, Lea Kingsley (2003) Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. Am J Epidemiol 158(7):687–694
3. Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. Epidemiology 11(5):561–570
4. Joffe MM, Ten Have TR, Feldman HI, Kimmel SE (2004) Model selection, confounder control, and marginal structural models. Am Stat 58(4):272–279
5. Miller JE, Molnar MZ, Kovesdy CP, Zaritsky JJ, Streja E, Salusky I, Arah OA, Kalantar-Zadeh K (2012) Administered paricalcitol dose and survival in hemodialysis patients: a marginal structural model analysis. Pharmacoepidemiol Drug Saf 21(11):1232–1239
6. Prentice RL, Williams BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. Biometrika 68(2):373–379
7. Robins JM (1999) Association, causation, and marginal structural models. Synthese 121(1):151–179
8. Robins JM, Hernan MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11(5):550–560
9. Van der Vaart AW (2000) Asymptotic statistics. Cambridge University Press, London
10. VanderWeele TJ (2009) Marginal structural models for the estimation of direct and indirect effects. Epidemiology 20(1):18–26
11. Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 84(408):1065–1073
12. Westreich D, Cole SR, Tien PC, Chmiel JS, Kingsley L, Funk MJ, Anastos K, Jacobson LP (2010) Time scale and adjusted survival curves for marginal structural cox models. Am J Epidemiol 171(6):691–700