

Semiparametric regression analysis for composite endpoints subject to componentwise censoring

BY GUOQING DIAO

*Department of Statistics, George Mason University, 4400 University Drive,
Fairfax, Virginia 22030, U.S.A.
gdiao@gmu.edu*

DONGLIN ZENG

*Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill,
North Carolina 27599, U.S.A.
dzeng@bios.unc.edu*

CHUNLEI KE, HAIJUN MA, QI JIANG

*Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320, U.S.A.
cke@amgen.com hma@amgen.com qjiang@amgen.com*

AND JOSEPH G. IBRAHIM

*Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill,
North Carolina 27599, U.S.A.
ibrahim@bios.unc.edu*

SUMMARY

Composite endpoints with censored data are commonly used as study outcomes in clinical trials. For example, progression-free survival is a widely used composite endpoint, with disease progression and death as the two components. Progression-free survival time is often defined as the time from randomization to the earlier occurrence of disease progression or death from any cause. The censoring times of the two components could be different for patients not experiencing the endpoint event. Conventional approaches, such as taking the minimum of the censoring times of the two components as the censoring time for progression-free survival time, may suffer from efficiency loss and could produce biased estimates of the treatment effect. We propose a new likelihood-based approach that decomposes the endpoints and models both the progression-free survival time and the time from disease progression to death. The censoring times for different components are distinguished. The approach makes full use of available information and provides a direct and improved estimate of the treatment effect on progression-free survival time. Simulations demonstrate that the proposed method outperforms several other approaches and is robust against various model misspecifications. An application to a prostate cancer clinical trial is provided.

Some key words: Composite endpoint; Hazard ratio; Progression-free survival; Semiparametric model.

1. INTRODUCTION

Composite endpoints, which consist of multiple component endpoints, are commonly used in clinical trials. Experiencing any of the events specified by the components is considered experiencing the composite endpoint. In particular, in oncology trials, progression-free survival is a composite endpoint with disease progression and death as the two components, and is defined as the time from randomization to the occurrence of disease progression or death from any cause. One example is the bone metastasis-free survival, with bone metastasis progression and death being the two components. Other examples of composite endpoints include cardiovascular death or hospitalization for heart failure, and major adverse cardiac events.

The censoring times for the individual components of a composite endpoint may not be the same. Using progression-free survival as an example, the two components, disease progression status and survival status, are often assessed using different schedules and methods, and are potentially subject to different censoring mechanisms, so the censoring times of the two components could differ for patients not experiencing the endpoint event. For example, assessment of objective disease progression requires radiological imaging evaluations. The progression-free status can only be confirmed up to the last evaluation, while the survival status can be assessed more frequently and thus confirmed up to a later time-point. Componentwise censoring can also occur when we retrieve information about survival status from public records for subjects who withdraw from clinical trials; for example, see the U.S. Food and Drug Administration's guidance document on data retention (FDA, 2008). However, not all death records are retrievable. When these events are included in a progression-free survival analysis, it is not clear how the censoring time should be defined for early-withdrawal subjects.

Componentwise censoring poses major challenges in the analysis of composite endpoints. Some conventional approaches include taking the minimum or the maximum of the censoring times of the two components. For example, in the U.S. Food and Drug Administration's guidance document on oncology clinical trials (FDA, 2007), it is recommended to censor progression-free survival time at the last assessment that determined a lack of progression. Death after two or more missed visits should be censored at the last assessment date before the missed visits. In practice, other methods may be used, such as censoring at the maximum censoring time for all component events, or treating withdrawal or change of therapy prior to disease progression as events (EMA, 2008). However, these methods either do not take advantage of the full information on all components, or impose assumptions that are hard to verify. As a result, conventional approaches may suffer from efficiency loss and could produce biased parameter estimates.

Composite endpoints could be handled by the method of Wei et al. (1989), where disease progression and death are analysed separately and a global treatment effect estimated by combining estimates from individual components. But this global treatment effect does not necessarily correspond to the effect on the composite endpoint. Wu & Cook (2012) proposed using copulas to link marginal distributions of the individual component endpoints with a proportional hazards structure. The treatment effect on the composite endpoint was not modelled directly but instead derived from the joint distribution of two individual components with a complicated form; therefore, it is not easy or intuitive to interpret the results. Quan et al. (2007) proposed a likelihood-based method for the analysis of a composite endpoint with missing data in its components. In their approach, instead of directly analysing the composite endpoint, the probabilities of all possible study outcomes were derived using all available data, and then the event rate was constructed for the composite endpoint based on these probabilities. Li & Zhang (2015) developed a parametric multistate model assuming a Weibull distribution for each possible transition from one state to another. None of the aforementioned methods considered componentwise censoring.

In a 2015 unpublished paper, Y. Y. Chen, C. Ke and J. Wang addressed different censoring schemes for component events through imputation. They proposed three multiple imputation-based methods: imputing the event time marginally using Kaplan–Meier estimates; imputing based on a Cox proportional hazards model; and imputing the event time of one component event based on a Kaplan–Meier estimate conditional on the other event. The risk set for sampling may become small towards the tail, leading to large variation. Furthermore, the inference procedure relies on simulation and can be computationally intensive. [Boruvka & Cook \(2016\)](#) developed sieve estimation in a Markov illness-death process under componentwise censoring, using a Cox-type proportional intensity model for each possible transition and assuming the baseline cumulative intensity functions to have piecewise parametric forms.

In this article, we develop a new likelihood-based approach to handle different censoring times for the individual components in a composite endpoint. The proposed model decomposes the endpoints and models both the progression-free survival time and the time from disease progression to death. The likelihood is derived, and the parameters including the treatment effect on progression-free survival time are estimated by maximizing the likelihood. The censoring times for different components are distinct and linked to individual components. Therefore, this approach makes full use of the available information and improves the estimate of the treatment effect on progression-free survival time. Furthermore, the proposed method provides insights into the covariate effects on the time from disease progression to death among patients who experience disease progression before death.

2. MODELS

Let T_s denote the time from randomization to disease progression and T_d the time from randomization to death. The composite endpoint progression-free survival time is defined as $Y = \min(T_s, T_d)$, i.e., the time to the first occurrence of disease progression or death. Our main interest is to study the treatment effect on this composite endpoint in the presence of right-censored data. We propose to model Y given X directly, which includes both treatment and baseline covariates, via the proportional hazards model

$$\lambda(t | X) = \lambda(t) \exp(X^T \beta), \quad (1)$$

where $\lambda(t)$ is the baseline hazard function and β represents the effect of X on the log hazard rate of Y . Model (1) explicitly gives treatment effects on composite endpoint and survival profiles. We propose to model the baseline cumulative hazard $\Lambda(t) = \int_0^t \lambda(s) ds$ parametrically. For example, we assume that $\Lambda(t) = (t/\alpha)^\theta$, which corresponds to a Weibull distribution. Hereafter we write $\Lambda(t; \alpha, \theta)$ and $\lambda(t; \alpha, \theta)$ to emphasize that the baseline cumulative hazard and baseline hazard functions for Y depend on parameters α and θ .

Since death is always the end observation for any subject, our second model will involve the conditional distribution of $G \equiv T_d - Y$ given X and Y ; that is, we model the distribution of the time from the occurrence of the composite event to death. Clearly, if a patient never experiences disease progression, T_d will be the same as Y ; otherwise, $Y = T_s$ is strictly less than T_d . This implies that G has a positive probability of being degenerate at zero. Therefore, we propose to model G given X and Y using a zero-inflated hazard model

$$\text{pr}(G > t | X, Y) = \{1 - q(X, Y; \xi)\} \exp[-A(t) \exp\{(X, Y)^T \gamma\}], \quad t \geq 0, \quad (2)$$

where $q(X, Y; \xi) = \text{pr}(G = 0 \mid X, Y; \xi)$ and $A(t)$ is the baseline cumulative hazard function for G in the subpopulation where $G > 0$. Although we assume a parametric form for $\lambda(t)$ in model (1), we leave $A(t)$ in (2) fully nonparametric, since we are interested in a model for Y but are willing to leave the model for G as nonparametric as possible. We model $q(X, Y; \xi)$ by the logistic regression model

$$q(X, Y; \xi) = \frac{\exp\{\xi_0 + (X, Y)^\top \xi_1\}}{1 + \exp\{\xi_0 + (X, Y)^\top \xi_1\}}, \quad (3)$$

so the unknown parameters are $\psi \equiv (\eta, A)$, where $\eta = (\alpha, \theta, \beta, \xi, \gamma)$ are the finite-dimensional parameters.

Define C_s and C_d to be the censoring times for T_s and T_d , respectively. Typically a study design ensures that $C_s \leq C_d$. Therefore, the observed data contain $O \equiv \{\tilde{Y}, \tilde{Z}, \Delta_1, \Delta_2, \Delta_3, X\}$, where $\tilde{Y} = \min(T_s, T_d, C_s)$, $\tilde{Z} = \min(T_d, C_d)$, $\Delta_1 = I(T_s \leq C_s \text{ and } T_s < T_d)$, $\Delta_2 = I(\tilde{Y} = T_d)$ and $\Delta_3 = I(T_d \leq C_d)$, with $I(\cdot)$ denoting the indicator function. Define $f_Y(t \mid X; \alpha, \theta, \beta) = \lambda(t; \alpha, \theta) \exp(X^\top \beta) \exp\{-\Lambda(t; \alpha, \theta) \exp(X^\top \beta)\}$ and $S_Y(t \mid X; \alpha, \theta, \beta) = \exp\{-\Lambda(t; \alpha, \theta) \exp(X^\top \beta)\}$ to be the density function and survival function of Y given X , respectively. Additionally, define $f_G(t \mid X, t_1; A, \gamma) = a(t) \exp\{(X, t_1)^\top \gamma\} \exp[-A(t) \exp\{(X, t_1)^\top \gamma\}]$ and $S_G(t \mid X, t_1; A, \gamma) = \exp[-A(t) \exp\{(X, t_1)^\top \gamma\}]$ to be respectively the density function and survival function of G given $Y = t_1$ and X in the subpopulation where $G > 0$. Here $a(\cdot)$ is the first derivative of $A(\cdot)$.

Assume conditional independent censoring, i.e., that the censoring times (C_s, C_d) are independent of (T_s, T_d) given X . We next derive the likelihood contribution from a random sample O under five possible scenarios, (a)–(e).

Scenario (a): $\Delta_1 = \Delta_3 = 1$. Both $Y = \tilde{Y}$ and $G = \tilde{Z} - \tilde{Y}$ are observed. Therefore the likelihood contribution is $W_1(O; \psi) = f_Y(\tilde{Y} \mid X; \alpha, \theta, \beta) \{1 - q(X, \tilde{Y}; \xi)\} f_G(\tilde{Z} - \tilde{Y} \mid X, \tilde{Y}; A, \gamma)$.

Scenario (b): $\Delta_2 = \Delta_3 = 1$. We observe death at $\tilde{Y} = \tilde{Z}$, but no disease progression was observed before \tilde{Y} . Therefore $Y = \tilde{Y}$ and $G = 0$. The likelihood contribution is $W_2(O; \psi) = q(X, \tilde{Y}; \xi) f_Y(\tilde{Y} \mid X; \alpha, \theta, \beta)$.

Scenario (c): $\Delta_1 = \Delta_2 = 0$ and $\Delta_3 = 1$. The patient was observed to stop disease progression assessment or to drop out at \tilde{Y} without disease progression, but later died at \tilde{Z} . Therefore $Y = \tilde{Z}$ with probability $q(X, \tilde{Z}; \xi)$ and $Y \in (\tilde{Y}, \tilde{Z})$ with probability $1 - q(X, \tilde{Y}; \xi)$. The likelihood contribution is

$$\begin{aligned} W_3(O; \psi) &= q(X, \tilde{Z}; \xi) f_Y(\tilde{Z} \mid X; \alpha, \theta, \beta) \\ &\quad + \int_{\tilde{Y}}^{\tilde{Z}-} f_Y(t \mid X; \alpha, \theta, \beta) \{1 - q(X, t; \xi)\} f_G(\tilde{Z} - t \mid X, t; A, \gamma) dt. \end{aligned}$$

Scenario (d): $\Delta_1 = 1$ and $\Delta_3 = 0$. We observe $Y = \tilde{Y}$ but G is censored at $\tilde{Z} - \tilde{Y}$. The likelihood contribution is $W_4(O; \psi) = \{1 - q(X, \tilde{Y}; \xi)\} f_Y(\tilde{Y} \mid X; \alpha, \theta, \beta) S_G(\tilde{Z} - \tilde{Y} \mid X, \tilde{Y}; A, \gamma)$.

Scenario (e): $\Delta_1 = \Delta_2 = \Delta_3 = 0$. In this case, $\tilde{Y} = C_s$, $\tilde{Z} = C_d$, and both Y and T_d are censored. The likelihood contribution is

$$\begin{aligned} W_5(O; \psi) &= \int_{\tilde{Z}}^{\infty} q(X, t; \xi) f_Y(t \mid X; \alpha, \theta, \beta) dt \\ &\quad + \int_{\tilde{Z}}^{\infty} \int_{\tilde{Y}}^{t_2} \{1 - q(X, t_1; \xi)\} f_Y(t_1 \mid X; \alpha, \theta, \beta) f_G(t_2 - t_1 \mid X, t_1; A, \gamma) dt_1 dt_2, \end{aligned}$$

where the first term corresponds to the case where $Y = T_d$ and $G = 0$, and the second term corresponds to the case where $Y = T_s$ and $G > 0$. One can show that

$$W_5(O; \psi) = S_Y(\tilde{Z} | X; \alpha, \theta, \beta) + \int_{\tilde{Y}}^{\tilde{Z}} \{1 - q(X, s_1; \xi)\} f_Y(s_1 | X; \alpha, \theta, \beta) S_G(\tilde{Z} - s_1 | X, s_1; A, \gamma) ds_1.$$

Define $I_1(O) = I(\Delta_1 = \Delta_3 = 1)$, $I_2(O) = I(\Delta_2 = \Delta_3 = 1)$, $I_3(O) = I(\Delta_1 = \Delta_2 = 0, \Delta_3 = 1)$, $I_4(O) = I(\Delta_1 = 1, \Delta_3 = 0)$ and $I_5(O) = I(\Delta_1 = \Delta_2 = \Delta_3 = 0)$. The likelihood function of ψ based on n independent observations $\{O_i \equiv (\tilde{Y}_i, \tilde{Z}_i, \Delta_{1i}, \Delta_{2i}, \Delta_{3i}, X_i)\}$ ($i = 1, \dots, n$) is $L_n(\psi) = \prod_{i=1}^n \prod_{j=1}^5 W_j(O_i; \psi)^{I_j(O_i)}$. Naturally, one would maximize $L_n(\psi)$ to obtain an estimator of ψ . However, for a function A with fixed values at the observed gap times, we can always let $a(\tilde{Z}_i - \tilde{Y}_i)$ go to ∞ for subject i with $I_1(O_i) = 1$. Therefore, we apply nonparametric maximum likelihood estimation by letting the estimator be a step function with jumps at the observed values for G . Suppose that $0 < \tau_1 < \dots < \tau_m$ are the m distinct observed gap times G . Using the nonparametric likelihood approach, $A(t) = \sum_{j: \tau_j \leq t} A\{\tau_j\}$, where $A\{t\}$ is the jump size of $A(\cdot)$ at t . For ease of notation, we denote the resulting nonparametric likelihood function by $L_n(\psi)$ and the corresponding nonparametric loglikelihood by $l_n(\psi)$.

We show how to calculate the integrals in $W_3(O; \psi)$ and $W_5(O; \psi)$ by using nonparametric maximum likelihood. Under scenario (c), the second term of the likelihood contribution based on an observation O is

$$\begin{aligned} & \sum_{j: \tau_j < \tilde{Z} - \tilde{Y}} \{1 - q(X, \tilde{Z} - \tau_j, X; \xi)\} A\{\tau_j\} \\ & \times \exp\{(X, \tilde{Z} - \tau_j)^\top \gamma\} \exp[-A(\tau_j) \exp\{(X, \tilde{Z} - \tau_j)^\top \gamma\}] \\ & \times \lambda(\tilde{Z} - \tau_j; \alpha, \theta) \exp(X^\top \beta) \exp[-\Lambda(\tilde{Z} - \tau_j; \alpha, \theta) \exp(X^\top \beta)]. \end{aligned}$$

We now turn our attention to scenario (e). Define $\tau_{m+1} = \infty$ and $0 = g_0 < g_1 < \dots < g_k \leq g_{k+1} = \tilde{Z} - \tilde{Y} < \tau_{k+1}$, where $g_j = \tau_j$ ($j = 1, \dots, k$). We can show that the second term of the likelihood contribution in $W_5(O; \psi)$ can be written as

$$\begin{aligned} & \int_{\tilde{Y}}^{\tilde{Z}} \{1 - q(X, s_1; \xi)\} f_Y(s_1 | X; \alpha, \theta, \beta) S_G(\tilde{Z} - s_1 | X, s_1; A, \gamma) ds_1 \\ & = \sum_{j=0}^k \int_{\tilde{Z} - g_{j+1}}^{\tilde{Z} - g_j} \{1 - q(s_1, X; \xi)\} f_Y(s_1 | X; \alpha, \theta, \beta) S_G(\tau_j | X, s_1; A, \gamma) ds_1. \end{aligned}$$

We use Gauss–Legendre quadrature to approximate the above integrals. In our experience, the Gauss–Legendre quadrature is reasonably accurate with ten abscissa points. To maximize $L_n(\psi)$, we use the quasi-Newton algorithm (Press et al., 1992). The resulting nonparametric maximum likelihood estimators of ψ are denoted by $\hat{\psi}_n \equiv (\hat{\eta}_n, \hat{A}_n)$.

3. ASYMPTOTIC PROPERTIES

In this section, we establish the asymptotic properties of the nonparametric maximum likelihood estimators $\hat{\psi}_n$. Besides the conditional independent censoring assumption described in the previous section, we make the following assumptions.

Assumption 1. The covariate X has bounded support. If $c_1^\top X = c_0$ with probability 1, then $c_0 = 0$ and $c_1 = 0$.

Assumption 2. There exists some positive constant δ_1 such that $\text{pr}(C_s \geq \tau \mid X) = \text{pr}(C_s = \tau \mid X) \geq \delta_1$ almost surely, where τ is a constant denoting the end of the study. Additionally, with probability 1, $C_s \leq C_d$.

Assumption 3. The true parameter values of $(\alpha, \theta, \beta, \xi, \gamma)$, denoted by $\eta_0 \equiv (\alpha_0, \theta_0, \beta_0, \xi_0, \gamma_0)$, lie in the interior of a known compact set in the domain of η .

Assumption 4. The true baseline cumulative hazard function A_0 is strictly increasing and continuously differentiable in $[0, \tau]$.

Assumption 1 is the usual condition for a design matrix in regression settings to ensure model identifiability. Assumptions 3 and 4 and the first part of Assumption 2 are standard regularity and technical conditions for a regression model with right-censored data. Assumption 2 also implies $\text{pr}(C_d \geq \tau \mid X) = \text{pr}(C_d = \tau \mid X) \geq \delta_1$. Assumptions 1 and 4 imply that $\text{pr}(T_d > T_s \mid X) \geq \delta_2$ for some positive constant δ_2 ; hence some patients experience disease progression before death and so we can observe the gap time G with a positive probability.

We show in the [Supplementary Material](#) that the unknown parameters η and $A(t)$ ($t \in [0, \tau]$) are identifiable. We next establish consistency and asymptotic normality of the proposed nonparametric maximum likelihood estimators in the following two theorems.

THEOREM 1. *Under the conditional independent censoring assumption and Assumptions 1–4, with probability tending to 1, $\|\hat{\eta}_n - \eta_0\| + \sup_{t \in [0, \tau]} |\hat{A}_n(t) - A_0(t)| \rightarrow 0$, where $\|\cdot\|$ is the Euclidean norm.*

THEOREM 2. *Under the conditional independent censoring assumption and Assumptions 1–4, $n^{1/2}(\hat{\eta}_n - \eta_0, \hat{A}_n - A_0) \rightarrow \mathcal{G}$ in distribution, where \mathcal{G} is a continuous zero-mean Gaussian process in $l^\infty(\mathcal{H})$ and $\mathcal{H} = \{(h_1, h_2) : h_1 \in \mathbb{R}^{3p+5}, h_2(\cdot) \text{ is a function on } [0, \tau]; \|h_1\| \leq 1, |h_2(t)|_{V[0, \tau]} \leq 1\}$. Here $l^\infty(\mathcal{H})$ denotes the space of all bounded linear functionals on \mathcal{H} and $|h|_{V[0, \tau]}$ denotes the total variation of h in $[0, \tau]$. Furthermore, the limiting covariance matrix of $n^{1/2}(\hat{\eta}_n - \eta_0)$ attains the semiparametric efficiency bound.*

Theorem 1 states the consistency of the nonparametric maximum likelihood estimators. The basic idea in the proof of Theorem 1 is as follows. We first show by contradiction that $\hat{A}_n(\tau)$ cannot diverge. Once the boundedness of $\hat{A}_n(\tau)$ is established, a subsequence of \hat{A}_n can be found that converges pointwise to a bounded monotone function A^* in $[0, \tau]$, and the same sequence of $\hat{\eta}_n$ converges to some η^* . We construct a step function \bar{A}_n with jumps at the observed gap times converging to A_0 . Finally, we prove that the Kullback–Leibler information between the true density and the density indexed by (η^*, A^*) is nonpositive. Consistency will then follow from the identifiability result. The details are given in the Appendix.

Theorem 2 implies that for any $(h_1, h_2) \in \mathbb{R}^{3p+5} \times l^\infty(\mathcal{H})$, $n^{1/2}(\hat{\eta}_n - \eta_0)^\top h_1 + n^{1/2} \int_0^\tau h_2(t) (d\hat{A}_n - dA_0)$ is asymptotically normal with mean zero. This result can be derived by verifying the four conditions in Theorem 3.3.1 of [van der Vaart & Wellner \(1996\)](#). To estimate the covariance matrix of $(\hat{\eta}_n, \hat{A}_n)$, we view the nonparametric likelihood as a parametric likelihood with η and the jump sizes of A at the observed gap times as parameters. Following the arguments in Theorem 2 of [Parner \(1998\)](#), we can then estimate the asymptotic covariance matrix by inverting the observed information matrix according to parametric likelihood theory. The proof of Theorem 2 is given in the Appendix.

Table 1. *Summary statistics for nonparametric maximum likelihood estimates based on 1000 replicates with $n = 300$*

Parameter	True	Bias ($\times 100$)	SE ($\times 100$)	SEE ($\times 100$)	CP ($\times 100$)
α	1	2.7	12.8	12.6	95
θ	1	1.7	6.4	6.5	94
β_1	1	2.3	10.7	10.3	94
β_2	-1	0.1	17.6	17.8	96
ξ_0	0	6.7	30.5	29.1	94
ξ_1	0.5	0.4	26.5	25.6	95
ξ_2	0.5	7.3	54.0	49.6	95
γ_1	1	-1.1	28.3	27.8	95
γ_2	0.5	0.0	48.4	46.5	95
$A(0.20)$	0.2	0.9	8.7	8.5	96
$A(0.25)$	0.25	1.4	10.3	10.1	95
$A(0.30)$	0.3	1.9	12.0	11.6	94
$A(0.35)$	0.35	2.1	13.2	13.0	95

SE, sample standard deviation of the estimates; SEE, average of the standard error estimates; CP, coverage probability of the 95% confidence interval based on a normal approximation.

4. SIMULATION STUDIES

We conducted extensive simulation studies to examine the performance of the proposed non-parametric maximum likelihood estimators under joint modelling of the composite endpoint Y and the gap time G between Y and T_d . We first generate two covariates, $X_1 \sim N(0, 1)$ and $X_2 \sim \text{Ber}(1/2)$, and then generate the composite endpoint Y from the model $\Lambda(t | X) = t \exp(\beta_1 X_1 + \beta_2 X_2)$. This implies that the true values of α and θ in the Weibull distribution both equal 1. We next determine whether the composite endpoint is T_s or T_d based on a logistic regression model $\text{pr}(G = 0 | X_1, X_2; \xi) = \exp(\xi_0 + \xi_1 X_1 + \xi_2 X_2) / \{1 + \exp(\xi_0 + \xi_1 X_1 + \xi_2 X_2)\}$. If $G = 0$, then the composite endpoint is T_d . Otherwise, we generate G based on the model $A(t | X_1, X_2; \gamma) = t \exp(\gamma_1 X_1 + \gamma_2 X_2)$. The true value of the cumulative hazard function for G given $G > 0$ is $A(t) = t$. Then $T_d = T_s + G$. Finally, we generate the censoring time for T_s , denoted by C_s , from an exponential distribution with mean 0.5; then $C_d = C_s + u$, where u follows an exponential distribution with mean 2.

We set the true regression parameter values to be $(\beta_1, \beta_2, \xi_0, \xi_1, \xi_2, \gamma_1, \gamma_2) = (1, -1, 0, 0.5, 0.5, 1, 0.5)$. The relative frequencies of the five scenarios (a)–(e) are 8.84%, 16.87%, 29.80%, 1.91% and 42.59%, respectively. We estimate the covariance matrix of $(\hat{\eta}_n, \hat{A}_n)$ by inverting the observed Fisher information matrix, i.e., the negative second derivatives of the loglikelihood with respect to η and the jump sizes of $A(\cdot)$, evaluated at $(\hat{\eta}_n, \hat{A}_n)$. Table 1 shows that the estimators have low bias and that the standard errors reflect the actual variation of the estimates. The coverages of the 95% confidence intervals are close to the nominal level.

We also used four naive methods to analyse the composite endpoint. All treat Y as the survival endpoint subject to right censoring. Denote the censoring time by C and the censoring indicator by $\Delta = I(Y \leq C)$. In scenarios (a), (b) and (d), we can directly observe $Y = \tilde{Y}$ and therefore $\Delta = 1$. In scenario (c) we do not observe the composite endpoint but we know that Y belongs to $(\tilde{Y}, \tilde{Z}]$. In scenario (e), the composite endpoint is censored but it is not clear whether the censoring time is \tilde{Y} or \tilde{Z} or some other time-points between \tilde{Y} and \tilde{Z} . Table 2 provides the definitions of $\min(Y, C)$ and Δ for the naive methods in scenarios (c) and (e).

Naive methods (i) and (iii) assume that under scenario (c) Y is censored at \tilde{Y} , but they ignore the information that Y cannot exceed \tilde{Z} . Naive methods (ii) and (iv), under scenario (c), assume

Table 2. Definition of $\min(Y, C)$ and Δ for the four naive methods

Naive method	Scenario (c)	Scenario (e)
(i)	$C = \tilde{Y}, \Delta = 0$	$C = \tilde{Y}, \Delta = 0$
(ii)	$Y = \tilde{Z}, \Delta = 1$	$C = \tilde{Y}, \Delta = 0$
(iii)	$C = \tilde{Y}, \Delta = 0$	$C = \tilde{Z}, \Delta = 0$
(iv)	$Y = \tilde{Z}, \Delta = 1$	$C = \tilde{Z}, \Delta = 0$

that time to death is the composite endpoint and ignore the possibility that disease progression occurs in the interval (\tilde{Y}, \tilde{Z}) . In scenario (e), the censoring time $C \in [\tilde{Y}, \tilde{Z}]$ is unknown and is imputed with either \tilde{Y} or \tilde{Z} . FDA guidance recommends censoring patients without an event at the last adequate disease assessment date and to censor at the last disease assessment if death occurs after more than one missed disease assessment. This is similar to naive method (iii).

For all four methods, we fit the parametric Weibull proportional hazards model in order to have a fair comparison, since the true composite endpoint was generated from a Weibull distribution. We consider simulation settings similar to the above, but also incorporate Y in the model of $\text{pr}(G = 0 | X_1, X_2, Y; \xi)$ and $A(t | X_1, X_2, Y; \gamma)$. The corresponding regression parameters for Y are $\xi_3 = \gamma_3 = 0.5$. Table 3 summarizes the results. The proposed method essentially outperforms all naive methods in terms of mean squared error efficiency. Although the parameter estimates from naive method (i) have low biases, they are less efficient than the proposed estimators, because naive method (i) ignores the possibility that the composite endpoint is the observed death time under scenario (c). Naive method (ii) treats the observed death time as the composite endpoint under scenario (c) and therefore led to an excessive number of events. Consequently, the estimates from naive method (ii) have smaller variation but large biases. Similarly, naive method (iv) also yielded biased parameter estimates. Estimates from naive method (iii) have both large biases and large variations.

In the next set of simulation studies, we compared the performances of the Wald tests of the effect of X_2 based on the proposed method and the four naive methods. We consider the same simulation setting as above except that β_2 varies from 0 to -0.6 . Table 4 presents the Type I error rates and powers at the 0.05 significance level based on 1000 replicates with $n = 300$. The proposed method controls the Type I error rate accurately and is substantially more powerful than all four naive methods. In particular, naive methods (ii) and (iv) tend to have inflated Type I error rates.

Naive methods (ii) and (iv) replace the composite endpoint with the death time under scenario (c). Intuitively, the covariate effect on the composite endpoint is a complicated function of the covariate effects on both time to disease progression and time to death, as well as other parameters. Consequently, the covariate effect on the death time tends to contribute more to the covariate effect on the composite endpoint in naive methods (ii) and (iv). We consider a simple example to demonstrate this observation. Suppose that T_s and T_d are independent. Then the hazard function for the composite endpoint is $\lambda(t | X) = \exp(\beta^T X) + \exp(\gamma^T X)$. If $\beta_1 = \gamma_1$, then $\lambda(t | X) = \exp(\beta_1 X_1) \{ \exp(\beta_2 X_2) + \exp(\gamma_2 X_2) \}$. The log hazard ratio of the composite endpoint for X_2 is then $\beta_2^* = \log[\{ \exp(\beta_2 X_2) + \exp(\gamma_2 X_2) \} / 2]$. Therefore, it is possible that $\beta_2^* = 0$ but $\beta_2 \neq 0$ and $\gamma_2 \neq 0$. For example, we let $\beta_2 = 0.2$ and $\gamma_2 = -0.25$, yielding $\beta_2^* = 0$. We conducted a simulation study using this setting, where under the null hypothesis there is no treatment effect on the composite endpoint. Based on 1000 replicates with sample size $n = 300$, naive methods (ii) and (iv) led to inflated Type I errors at a significance level of 0.05, with empirical Type I error rates of 0.09 and 0.12, respectively. The proposed method has an empirical Type I error rate of 0.06.

Table 3. Comparison of the proposed method and four naive methods based on 1000 replicates with $n = 300$

Parameter	Bias ($\times 100$)	SD ($\times 100$)	MSE ($\times 100$)	RE
		Proposed method		
α	2.1	12.2	1.5	
θ	2.1	6.6	0.5	
β_1	2.2	10.0	1.1	
β_2	0.4	17.4	3.0	
		Naive method (i)		
α	1.5	17.8	3.2	2.10
θ	2.2	8.8	0.8	1.71
β_1	2.3	13.1	1.8	1.66
β_2	-1.3	23.7	5.7	1.87
		Naive method (ii)		
α	-12.1	8.0	2.1	1.38
θ	12.6	6.6	2.0	4.20
β_1	-11.6	9.2	2.2	2.09
β_2	25.9	15.9	9.2	3.04
		Naive method (iii)		
α	112.8	48.4	150.7	99.04
θ	-20.0	6.5	4.4	9.16
β_1	19.0	13.1	5.3	5.03
β_2	-15.1	23.9	8.0	2.63
		Naive method (iv)		
α	18.2	11.6	4.6	3.05
θ	1.6	6.3	0.4	0.87
β_1	6.8	9.4	1.3	1.27
β_2	9.5	15.6	3.3	1.10

SD, empirical standard deviation of the estimates; MSE, mean squared error; RE, mean squared error relative efficiency of the proposed estimators compared to the estimators using the naive methods.

Table 4. Type I error rate ($\times 100$) and power ($\times 100$) for testing the effect of X_2 on composite endpoints at significance level 0.05 based on 1000 replicates with $n = 300$

β_2	Proposed	Naive (i)	Naive (ii)	Naive (iii)	Naive (iv)
0	5	5	6	5	8
-0.1	8	8	5	7	4
-0.2	23	18	11	18	12
-0.3	47	35	27	37	33
-0.4	71	50	45	54	56
-0.5	88	66	67	74	81
-0.6	96	80	83	88	94

To examine the performance of the proposed test under model misspecifications, we consider a variety of models for the composite endpoint and the gap time, including the semiparametric transformation model with a normal error (Diao & Lin, 2005) and copula models. The proposed test appears to be robust against model misspecification and is substantially more powerful than the naive methods. Finally, as suggested by a referee, we generated data by mimicking real applications. Specifically, we assess disease progression every six months or every three months,

Table 5. *Frequencies and relative frequencies of the five scenarios for the prostate cancer data*

Scenario	Standard treatment		New treatment	
	Frequency	Relative frequency (%)	Frequency	Relative frequency (%)
(a)	130	23.2	119	21.2
(b)	0	0.0	0	0.0
(c)	74	13.2	72	12.9
(d)	121	21.6	112	20.0
(e)	235	42.0	257	45.9

leading to interval-censored data. While the proposed method is not designed for such data, we impute the time to disease progression with the midpoint of the interval and then apply the proposed method, which seems to be reasonably robust in this situation. Details of these two sets of simulation studies are provided in the [Supplementary Material](#).

5. APPLICATION

We now apply the proposed method to a randomized placebo-controlled clinical trial for non-metastatic prostate cancer patients at high risk for bone metastasis. The primary endpoint was bone metastasis-free survival, defined as time from randomization to bone metastasis or death, whichever occurs first. Events of bone metastasis were determined by central review of images collected periodically at baseline and on study, and therefore were censored at the last image assessment date for patients who did not experience an event. All-cause death was assessed via study contact, and was censored at the last study contact date for patients who were alive. Some patients discontinued image assessments but stayed on study for overall survival assessment, so they were censored earlier for bone metastasis but could die or be censored for death later. The study was stratified by previous or current use of chemotherapy and by high risk for metastasis based on prostate-specific antigen. The treatment phase of the study ended when a targeted number of patients developed bone metastasis or died. In addition, the study had a long-term survival follow-up phase, where patients who had progressed or who did not wish to continue the scheduled study assessment were followed for survival status only.

We analysed a randomly selected subset of 1120 patients provided by the trial sponsor, with 560 in each arm. The proposed method was used to analyse the composite endpoint of bone metastasis-free survival, adjusting for the stratification factors, and was compared with some naive methods and two existing methods. For the proposed method, three covariates were included in the model for the composite endpoint, the model for $\text{pr}(G = 0)$ and the model for G when $G > 0$. These covariates are: *trt*, taking value 1 for new treatment and value 0 for standard treatment; *chemo*, taking value 1 for previous or current use of chemotherapy and value 0 otherwise; and *psa*, taking value 1 if the patient was assessed to be at high risk for metastasis based on prostate-specific antigen and value 0 otherwise. Among the 1120 patients, 80 had previous or current use of chemotherapy and 542 were at high risk based on prostate-specific antigen. We included Y in both the model for $\text{pr}(G = 0)$ and the model for G when $G > 0$.

We used data in both the treatment and the long-term survival follow-up phases. Table 5 shows the breakdown of the five scenarios described in §2. Most patients did not experience bone metastasis and/or were still alive. The proposed method was used to evaluate the treatment effect on bone metastasis-free survival. Table 6 shows the results for the model on the bone metastasis-free survival based on the proposed method, the four naive methods, and the method

Table 6. Results from the model on bone metastasis-free survival for the prostate cancer data

Parameter	Est	SE	Stat	<i>p</i> -value
Proposed method				
α	1.1647	0.0674		
θ	1.2539	0.0428		
trt	-0.1895	0.0805	-2.3543	0.0186
chemo	0.2760	0.1414	1.9520	0.0509
psa	0.6176	0.0823	7.5070	<0.0001
Naive method (i)				
α	1.3322	0.0880		
θ	1.2460	0.0474		
trt	-0.1973	0.0912	-2.1623	0.0306
chemo	0.2994	0.1621	1.8475	0.0647
psa	0.5486	0.0930	5.8999	<0.0001
Naive method (ii)				
α	1.1406	0.0631		
θ	1.3040	0.0436		
trt	-0.1837	0.0799	-2.2986	0.0215
chemo	0.2439	0.1397	1.7461	0.0808
psa	0.6219	0.0817	7.6119	<0.0001
Naive method (iii)				
α	1.5015	0.1072		
θ	1.1617	0.0446		
trt	-0.2010	0.0912	-2.2026	0.0276
chemo	0.2955	0.1617	1.8271	0.0677
psa	0.5142	0.0927	5.5459	<0.0001
Naive method (iv)				
α	1.2653	0.0754		
θ	1.2166	0.0410		
trt	-0.1878	0.0799	-2.3496	0.0188
chemo	0.2493	0.1395	1.7873	0.0739
psa	0.5916	0.0815	7.2588	0.0001
Chen et al. (2015)				
trt	-0.1715	0.0794	-2.1600	0.0311
chemo	0.2453	0.1414	1.7345	0.0826
psa	0.5593	0.0812	6.8845	<0.0001

Est, parameter estimate; SE, standard error; Stat, Wald test statistic; trt, treatment indicator; psa, prostate-specific antigen.

proposed in the unpublished paper by Chen et al. For the proposed method, the prostate-specific antigen risk level was highly associated with the bone metastasis-free survival. Patients at high risk based on prostate-specific antigen, psa, had a 85.4% increase in hazard relative to patients who were not at high risk. The corresponding 95% confidence interval was (57.8%, 117.9%) and the *p*-value was less than 0.0001. The hazard ratio of treatment versus control for the composite endpoint was estimated at 0.827 with a 95% confidence interval of (0.707, 0.969) and a *p*-value of 0.0186. In this example, all methods gave consistent conclusions that the treatment significantly reduced the risk for bone metastasis or death. However, as expected, naive methods (i) and (iii) had large standard errors and there were also notable differences in some parameter estimates between the proposed method and the four naive methods. According to the proposed method, no covariates have significant effects in the zero-inflated hazard model (2), except that psa has a significant effect on the probability that a composite endpoint is death.

With other covariates fixed, the odds that the composite endpoint is death for patients at high risk based on prostate-specific antigen was estimated at 1.939 times the odds for patients at low risk. The corresponding 95% confidence interval was (1.00, 3.76). We can draw similar conclusions from the results based on the proposed method and the method of Chen et al. However, there were notable differences in the parameter estimates, especially for the effect of psa. In addition, the inference procedure of Chen et al. requires simulation and thus is computationally more intensive.

6. DISCUSSION

Other parametric and semiparametric models may be used for the time to the composite endpoint and the gap time, respectively. The parametric assumption on the distribution of the composite endpoint may be relaxed through the use of B-splines over a sieve space. Furthermore, it would be desirable to develop diagnostic tools to check the goodness-of-fit of the proposed model. Future research along this direction is warranted.

In some applications, disease progression status is assessed periodically, leading to interval censoring. Zeng et al. (2015) showed that the right-endpoint imputation method would lead to biased estimation and reduced power when progression status is known only at periodic assessment times. While it is challenging to analyse interval-censored data by using semiparametric or nonparametric methods, we can extend the proposed models to this setting. Specifically, under models (1)–(3), we can derive the likelihood for each possible scenario of the observed data and develop inference procedures. This extension is currently under investigation. On the other hand, our simulation studies have demonstrated that the proposed method with midpoint imputation is reasonably robust with respect to interval censoring.

We considered a specific case where one component is terminal. In the case where no event terminates the other component event, one can choose one component event as the terminal event and apply the proposed approach. If one component event under monitoring is longer than the other component, the component event with the longer censoring time can be chosen as the terminal event. Otherwise, one can choose the lower-frequency component. Our method can also be extended without assuming this terminating structure.

Often a composite endpoint has more than two components. For example, in evaluating bone-target agents for advanced cancer patients, a composite endpoint of skeletal-related events includes four components: pathologic fracture, spinal cord compression, radiation to bone, and surgery to bone. In general, one can group components into two clusters and then apply the proposed method. For example, for the skeletal-related events, one group could consist of pathologic fracture, which is identified through regular skeletal survey, and the other three events could make up the second cluster, which is monitored through clinical visits.

We assume conditional independent censoring for both endpoints. However, this assumption may be violated in practice. For example, in the case of progression-free survival, patients may drop out of disease assessment due to worsening disease status. It would be interesting to extend the proposed method to account for informative censoring. In the prostate cancer trial, a large percentage of patients were event-free at the end of the study. We may use public records post trial to retrieve the information about survival status. However, caution is needed when including such information in the analysis because the sample post trial can be biased; for example, patients may die from other causes or take different treatments post trial. Finally, it remains the best approach to minimize the possibility of missing data during the design and conduct of a clinical trial and so minimize the impact of missing data on the analysis.

ACKNOWLEDGEMENT

The authors are grateful to the referees and the associate editor for insightful comments.

SUPPLEMENTARY MATERIAL

[Supplementary material](#) available at *Biometrika* online includes additional simulation studies, the proof of identifiability of the proposed model, and further details of the proofs of Theorems 1 and 2.

APPENDIX

Proof of Theorem 1

We introduce notation that will be used throughout the proofs of Theorems 1 and 2. Let O_i ($i = 1, \dots, n$) denote the observations for the i th subject. Define $\mathbb{P}_n\{g(O)\} = n^{-1} \sum_{i=1}^n g(O_i)$ and $\mathbb{P}\{g(O)\} = E\{g(O)\}$.

The proof of consistency consists of two main steps. In the first, we prove that $\limsup_n \hat{A}_n(\tau)$ has an upper bound with probability 1. Therefore, for any subsequence, there exists a further subsequence of $(\hat{\eta}_n, \hat{A}_n)$ that converges to (η^*, A^*) weakly by the Helly selection theorem. In the second step, we prove that $(\eta^*, A^*) = (\eta_0, A_0)$.

Step 1. We prove the boundedness of $\hat{A}_n(\tau)$ by contradiction. Define $R_j(O; \psi) = \log W_j(O; \psi)$ ($j = 1, \dots, 5$), where $W_j(O; \psi)$ ($j = 1, \dots, 5$) are as defined in § 2. Using the same notation for simplicity, we replace $a(t)$ with $A\{t\}$, the jump size of $A(\cdot)$ at t , in $W_j(O; \psi)$ and $R_j(O; \psi)$ ($j = 1, \dots, 5$).

Define $\hat{v}_n = \hat{A}_n(\tau)$ and $\tilde{A}_n(t) = \hat{A}_n(t)/\hat{v}_n$ for $t \in [0, \tau]$. To prove that \hat{A}_n in $[0, \tau]$ is bounded, it is sufficient to show that \hat{v}_n is bounded. By Assumptions 1 and 3 and the definition of $\tilde{A}_n(t)$, we can show that $R_1(O; \hat{\eta}_n, \hat{A}_n) - R_1(O; \hat{\eta}_n, \tilde{A}_n) \leq c_1 + \log \hat{v}_n - c_2 \tilde{A}_n(\tilde{G}) \hat{v}_n$ for some constant c_1 and a positive constant c_2 . Furthermore, we can show that $R_2(O; \hat{\eta}_n, \hat{A}_n) - R_2(O; \hat{\eta}_n, \tilde{A}_n) = 0$, $R_3(O; \hat{\eta}_n, \hat{A}_n) - R_3(O; \hat{\eta}_n, \tilde{A}_n) \leq c_3$, $R_4(O; \hat{\eta}_n, \hat{A}_n) - R_4(O; \hat{\eta}_n, \tilde{A}_n) \leq c_3 - c_4 \tilde{A}_n(\tilde{G}) \hat{v}_n$ and $R_5(O; \hat{\eta}_n, \hat{A}_n) - R_5(O; \hat{\eta}_n, \tilde{A}_n) \leq c_3$ for some constant c_3 and a positive constant c_4 .

Suppose that $\hat{v}_n \rightarrow \infty$. It follows that, for some constant c_5 and a positive constant c_6 ,

$$\begin{aligned} 0 &\leq \frac{1}{n} l_n(\hat{\eta}_n, \hat{A}_n) - \frac{1}{n} l_n(\hat{\eta}_n, \tilde{A}_n) \\ &\leq c_5 + \log \hat{v}_n - c_2 \mathbb{P}_n\{I_1(O) \tilde{A}_n(\tilde{G})\} \hat{v}_n - c_4 \mathbb{P}_n\{I_4(O) \tilde{A}_n(\tilde{G})\} \hat{v}_n \\ &\leq c_5 + \log \hat{v}_n - c_6 \hat{v}_n \rightarrow -\infty, \end{aligned}$$

where the penultimate inequality is obtained from the conditionally independent censoring assumption, Assumptions 1, 3 and 4, and the Glivenko–Cantelli theorem. This result contradicts the definition of $(\hat{\eta}_n, \hat{A}_n)$. The above argument holds for every sample in the probability space except on a set with zero probability. Thus we have shown that, with probability 1, $\hat{A}_n(\tau)$ is bounded for any sample of size n . Therefore, by Helly’s selection theorem, we can choose a further subsequence, still indexed by $\{n\}$, such that $(\hat{\eta}_n, \hat{A}_n)$ converges to (η^*, A^*) with probability 1.

Step 2. In this step, we show that $(\eta^*, A^*) = (\eta_0, A_0)$. By differentiating $l_n(\eta, A)$ with respect to $A\{\tilde{G}_i\}$ for $\Delta_{1i} = \Delta_{3i} = 1$ and setting the derivative to zero, we can see that $\hat{A}_n\{\tilde{G}_i\}$ satisfies the equation

$$\begin{aligned} \hat{A}_n\{\tilde{G}_i\} &= \frac{I_1(O_i)}{n \mathbb{P}_n \left\{ \sum_{k=1}^5 I_k(O) Q_k(t, O; \hat{\eta}_n, \hat{A}_n) \right\}} \Big|_{t=\tilde{G}_i} \\ &= \frac{I_1(O_i) + n \mathbb{P}_n I_3(O) Q_{31}(t, O; \hat{\psi}_n)}{n \mathbb{P}_n \left\{ \sum_{k \in \{1, 2, 4, 5\}} I_k(O) Q_k(t, O; \hat{\psi}_n) + I_3(O) Q_{32}(t, O; \hat{\psi}_n) \right\}} \Big|_{t=\tilde{G}_i}, \end{aligned} \tag{A1}$$

where the expressions for $Q_k(t, O; \psi)$ ($k = 1, \dots, 5$), $Q_{31}(t, O; \psi)$ and $Q_{32}(t, O; \psi)$ are given in the [Supplementary Material](#). We can easily verify that, with probability 1, $Q_k(t, O; \hat{\psi}_n)$ ($k = 1, 4, 5$), $Q_{31}(t, O; \hat{\psi}_n)$ and $Q_{32}(t, O; \hat{\psi}_n)$ are nonnegative. Furthermore, for any \tilde{G}_i with $\Delta_{1i} = \Delta_{3i} = 1$, $\mathbb{P}_n\{I_1(O)Q_1(\tilde{G}_i, O; \hat{\psi}_n)\}$ is positive. Therefore, the denominator in (A1) is bounded away from zero, and for a subject i with $\Delta_{1i} = \Delta_{3i} = 1$, $\hat{A}_n\{\tilde{G}_i\}$ is positive and bounded.

We next construct another step function $\bar{A}_n(t)$ with jumps only at the observed gap time \tilde{G}_i by replacing $\hat{\psi}_n$ with ψ_0 in (A1). We verify that $\bar{A}_n(t)$ converges to A_0 uniformly in $t \in [0, \tau]$ with probability 1. As is shown in the [Supplementary Material](#), the class $\mathcal{F}_1 = \{\sum_{k=1}^5 I_k(O)Q_k(t, O; \eta, A) : t \in [0, \tau], \eta \in \mathcal{B}_0, A \in \mathcal{A}, A(0) = 0\}$ is bounded and P-Donsker, where $\mathcal{A} = \{g : g \text{ is a nondecreasing function in } [0, \tau], g(\tau) \leq B_0\}$ and B_0 is a positive constant such that $\hat{A}_n(\tau) \leq B_0$ with probability 1. Since a P-Donsker class is also Glivenko–Cantelli, by the Glivenko–Cantelli theorem ([van der Vaart & Wellner, 1996](#)), $\bar{A}_n(t)$ converges uniformly to $E\{I_1(O)/\mu(t)\}$, where $\mu(t) = E\{\sum_{k=1}^5 I_k(O)Q_k(t, O; \eta_0, A_0)\}$. By the conditional independent censoring assumption, we can prove that $E\{I_1(O)/\mu(t)\} = A_0(t)$. Consequently, we conclude that \bar{A}_n converges uniformly to A_0 in $[0, \tau]$ with probability 1.

By the construction of $\hat{A}_n(t)$ and $\bar{A}_n(t)$, we can see that $\hat{A}_n(t)$ is absolutely continuous with respect to $\bar{A}_n(t)$; furthermore, by letting n go to infinity, $A^*(t)$ is differentiable with respect to $A_0(t)$ so that $A^*(t)$ is differentiable with respect to t . It follows that $d\hat{A}_n(t)/d\bar{A}_n(t)$ converges to $dA^*(t)/dA_0(t)$ uniformly in $t \in [0, \tau]$.

By the definition of $(\hat{\eta}_n, \hat{A}_n)$, $n^{-1}l_n(\hat{\eta}_n, \hat{A}_n) - n^{-1}l_n(\eta_0, \bar{A}_n) \geq 0$. Since $\mathcal{B}_0 \times \mathcal{A}$ is a Donsker class and the functionals $R_k(O; \eta, A)$ ($k = 1, \dots, 5$) are bounded Lipschitz functionals with respect to $\mathcal{B}_0 \times \mathcal{A}$, the class $\mathcal{F}_2 = \{\sum_{k=1}^5 I_k(O)R_k(O; \eta, A) : \eta \in \mathcal{B}_0, A \in \mathcal{A}, A(0) = 0, A(\tau) \leq B_0\}$ is P-Donsker and hence a Glivenko–Cantelli class. Therefore, by taking $n \rightarrow \infty$, the left-hand side in the above inequality converges to the negative Kullback–Leibler information. It then follows that, with probability 1, $\sum_{k=1}^5 I_k(O)R_k(O; \eta^*, A^*) = \sum_{k=1}^5 I_k(O)R_k(O; \eta_0, A_0)$. Therefore, from the identifiability result, we obtain $(\eta^*, A^*) = (\eta_0, A_0)$. This completes the proof of Theorem 1.

Proof of Theorem 2

To prove Theorem 2, we verify the four conditions (P1), \dots , (P4) in Theorem 3.3.1 of [van der Vaart & Wellner \(1996\)](#), which are listed in the [Supplementary Material](#). We first define a neighbourhood of the true parameters (η_0, A_0) by $\mathcal{U} = \{(\eta, A) : \|\eta - \eta_0\| + \sup_{t \in [0, \tau]} |A(t) - A_0(t)| < \epsilon_0\}$ for a very small constant ϵ_0 . Based on the consistency theorem, $(\hat{\eta}_n, \hat{A}_n)$ belongs to \mathcal{U} with probability close to 1 when the sample size n is large enough.

For any one-dimensional submodel given as $\{\eta + \epsilon h_1, A + \epsilon \int h_2 dA\}$ for $(\eta, A) \in \mathcal{U}$ and $H \equiv (h_1, h_2) \in \mathcal{H}$, we can derive the score function for a single observation O as $V(O; \psi)[H] = l_\eta(O; \psi)^\top h_1 + l_A(O; \psi)[\int h_2 dA]$, where $l_\eta(O; \psi) = \sum_{k=1}^5 I_k(O)dR_k(O; \psi)/d\eta$ and

$$\begin{aligned} l_A(O; \psi) \left[\int h_2 dA \right] &= I_1(O) \left\{ h_2(\tilde{G}) - \exp\{(X, \tilde{Y})^\top \gamma\} \int_0^{\tilde{G}} h_2 dA \right\} \\ &+ \frac{I_3(O)}{\exp\{R_3(O; \psi)\}} \int_0^{\tilde{G}} \{1 - q(X, \tilde{Z} - s, X; \xi)\} \exp\{(X, \tilde{Z} - s)^\top \gamma\} \\ &\quad \times \exp[-A(s) \exp\{(X, \tilde{Z} - s)^\top \gamma\}] \left[h_2(s) - \exp\{(X, \tilde{Z} - s)^\top \gamma\} \int_0^s h_2 dA \right] \\ &\quad \times \lambda(\tilde{Z} - s; \alpha, \theta) \exp\{X^\top \beta\} \exp[-\Lambda(\tilde{Z} - s; \alpha, \theta) \exp\{X^\top \beta\}] dA(s) \\ &- I_4(O) \exp\{(X, \tilde{Y})^\top \gamma\} \int_0^{\tilde{G}} h_2 dA \end{aligned}$$

$$\begin{aligned}
& - \frac{I_5(O)}{\exp\{R_5(O; \psi)\}} \int_{\tilde{Y}}^{\tilde{Z}} \{1 - q(X, s_1; \xi)\} f_Y(s_1 | X; \alpha, \theta, \beta) \\
& \quad \times S_G(\tilde{Z} - s_1 | X, s_1; A, \gamma) \exp\{(X, s_1)^\top \gamma\} \int_0^{\tilde{Z}-s_1} h_2 \, dA \, ds_1.
\end{aligned}$$

For ease of notation, we omit $[H]$ from $V(O; \psi)[H]$. We define $U_n(\psi) = \mathbb{P}_n\{V(O; \psi)\}$ and $U(\psi) = \mathbb{P}\{V(O; \psi)\}$. Thus it is easy to see that both $U_n(\psi)$ and $U(\psi)$ are maps from \mathcal{U} to $l^\infty(\mathcal{H})$ and that $n^{1/2}\{U_n(\psi) - U(\psi)\}$ is an empirical process in the space $l^\infty(\mathcal{H})$. It then follows that $U_n(\hat{\psi}_n) = 0$ and $U(\psi_0) = 0$.

To prove property (P1), we make use of Lemma 3.3.5 of [van der Vaart & Wellner \(1996\)](#). Based on the explicit expression for the score function, $V(O; \eta, A)[H]$ is continuously differentiable with respect to η and $\|dV(O; \eta, A)/d\eta\| \leq c_9$, where c_9 is a positive constant. Furthermore,

$$\begin{aligned}
|V(O; \eta, A) - V(O; \eta, \tilde{A})| & \leq c_{10} \left[I_1(O) \left\{ |A(\tilde{G}) - \tilde{A}(\tilde{G})| + \int_0^{\tilde{G}} |A(t) - \tilde{A}(t)| \, dt \right\} \right. \\
& \quad + I_3(O) \left\{ \int_0^{\tilde{G}} |A(s) - \tilde{A}(s)| \, dA(s) + \int_0^{\tilde{G}} |A(s) - \tilde{A}(s)| \, d\tilde{A}(s) \right. \\
& \quad \left. \left. + \int_0^{\tilde{G}} |dA(s) - d\tilde{A}(s)| \right\} \right. \\
& \quad \left. + I_4(O) |A(\tilde{G}) - \tilde{A}(\tilde{G})| + I_5(O) \int_{\tilde{Y}}^{\tilde{Z}} |A(s) - \tilde{A}(s)| \, ds \right]
\end{aligned}$$

for some positive constant c_{10} . Therefore $\sup_{H \in \mathcal{H}} E[\{V(O; \eta, A) - V(O; \eta, A)\}^2]$ converges to zero if $\|\eta - \eta_0\| + \sup_{t \in [0, \tau]} \{|A(t) - A_0(t)|\} \rightarrow 0$. Additionally, by using a similar argument to that in the proof of Theorem 1, we can show that the class $\mathcal{F}_3 = \{V(O; \eta, A)[H] - V(O; \eta_0, A_0)[H] : (\eta, A) \in \mathcal{U}, H \in \mathcal{H}\}$ is P-Donsker. Therefore, according to Lemma 3.3.5 of [van der Vaart & Wellner \(1996\)](#), property (P1) holds.

Property (P2) holds because of the P-Donsker property of the class $\{V(O; \eta, A)[H] : H \in \mathcal{H}\}$. Furthermore, the limit random elements ζ constitute a Gaussian process indexed by $H \in \mathcal{H}$ and the covariance between $\zeta(H_1)$ and $\zeta(H_2)$ is equal to $E\{V(O; \eta_0, A_0)[H_1] \times V(O; \eta_0, A_0)[H_2]\}$.

The Fréchet differentiability in (P3) can be directly verified using the smoothness of $U(\eta, A)$. The derivative of $U(\eta, A)$ at (η_0, A_0) , denoted by $U'(\eta_0, A_0)$, is a map from the space $\{(\eta - \eta_0, A - A_0) : (\eta, A) \in \mathcal{U}\}$ to $l^\infty(\mathcal{H})$.

It remains to show that U' is continuously invertible at (η_0, A_0) . Following the argument in the appendix of [Zeng & Lin \(2007\)](#), it suffices to prove that for any one-dimensional submodel given as $\{\eta + \epsilon h_1, A + \epsilon \int h_2 \, dA\}$ for $H \in \mathcal{H}$, the Fisher information along this submodel is nonsingular. If the Fisher information along this submodel is singular, the score function along this submodel is zero with probability 1. Using similar arguments to those in the proof of identifiability of the model, we can show that $V(O; \eta, A_0)[H] = 0$ yields $h_1 = 0$ and $h_2(t) = 0$ for any $t \in [0, \tau]$.

We now have verified properties (P1)–(P4), so Theorem 3.3.1 of [van der Vaart & Wellner \(1996\)](#) allows us to conclude that $n^{1/2}(\hat{\eta}_n - \eta_0, \hat{A}_n - A_0)$ converges weakly to a tight Gaussian random element $-U'^{-1}\zeta$ in $l^\infty(\mathcal{H})$. Moreover, it can be shown that $\hat{\eta}_n$ is an asymptotic linear estimator for η_0 and that the corresponding influence functions are in the space spanned by the score functions. This implies that $\hat{\eta}_n$ is semiparametrically efficient by semiparametric efficiency theory.

REFERENCES

- BORUVKA, A. & COOK, R. J. (2016). Sieve estimation in a Markov illness-death process under dual censoring. *Biostatistics* **17**, 350–63.
- DIAO, G. & LIN, D. Y. (2005). A powerful and robust method for mapping quantitative trait loci in general pedigrees. *Am. J. Hum. Genet.* **77**, 97–111.

- EMA (2008). *Appendix 1 to the Guideline on the Evaluation of Anticancer Medicinal Products in Man* (chmp/ewp/205/95 rev. 3). London: European Medicines Agency.
- FDA (2007). *Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*. Rockville, Maryland: U.S. Food and Drug Administration.
- FDA (2008). *Guidance for Sponsors, Clinical Investigators, and IRBs: Data Retention When Subjects Withdraw from FDA-Regulated Clinical Trials*. Rockville, Maryland: U.S. Food and Drug Administration.
- LI, Y. & ZHANG, Q. (2015). A Weibull multi-state model for the dependence of progression-free survival and overall survival. *Statist. Med.* **34**, 2497–513.
- PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press, 2nd ed.
- QUAN, H., ZHANG, D., ZHANG, J. & DEVLAMYNCK, L. (2007). Analysis of a binary composite endpoint with missing data in components. *Statist. Med.* **26**, 4703–18.
- VAN DER VAART, A. & WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.
- WEI, L.-J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.
- WU, L. & COOK, R. J. (2012). Misspecification of Cox regression models with composite endpoints. *Statist. Med.* **31**, 3545–62.
- ZENG, D. & LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with Discussion). *J. R. Statist. Soc. B* **69**, 507–64.
- ZENG, L., COOK, R. J., WEN, L. & BORUVKA, A. (2015). Bias in progression-free survival analysis due to intermittent assessment of progression. *Statist. Med.* **34**, 3181–93.

[Received on 4 July 2017. Editorial decision on 19 January 2018]