

Design of cancer trials based on progression-free survival with intermittent assessment

LEILEI ZENG

*Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: lzeng@uwaterloo.ca*

RICHARD J. COOK

*Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada*

KER-AI LEE

*Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada*

Summary

Therapeutic advances in cancer mean that it is now impractical to performed phase III randomized trials evaluating experimental treatments on the basis of overall survival. As a result, the composite endpoint of progression-free survival has been routinely adopted in recent years as it is viewed as enabling a more timely and cost-effective approach to assessing the clinical benefit of novel interventions. This article considers design of cancer trials directed at the evaluation of treatment effects on progression-free survival. In particular, we derive sample size criteria based on an illness-death model that considers cancer progression and death jointly while accounting for the fact that progression is assessed only intermittently. An alternative approach to design is also considered in which the sample size is derived based on a misspecified Cox model, which uses the documented time of progression as the progression time rather than dealing with the interval censoring. Simulation studies show the validity of the proposed methods.

Keywords: cancer trials, intermittent assessment of progression, multistate model, progression-free survival, sample size estimation

This is the peer reviewed version of the following article: Leilei Zeng, Richard J. Cook and Ker-Ai Lee, Design of cancer trials based on progression-free survival with intermittent assessment. *Statistics in Medicine* (2018), 37(12): 1947–2066 which has been published in final form at <https://doi.org/10.1002/sim.7641>.

1 INTRODUCTION

Cancer therapies are routinely evaluated in phase III clinical trials based on overall survival, progression-free survival, and progression times. While interest lies primarily in prolonging survival, sample size

requirements are often determined based on a Cox regression (Cox, 1972) analysis of the progression-free survival time. There has been much discussion in the oncology literature in recent years on the suitability of this endpoint and the interpretation of associated treatment effects (Fleming et al., 2009; Booth and Eisenhauer, 2012; Saaed et al., 2010). Miksad et al. (2009) sought insight into the relationship between treatment effects on progression-free survival and overall survival by viewing progression-free survival time as a surrogate for overall survival. Ballman et al. (2007) examined the validity of this assumption through assessing the association between treatment effects on these two endpoints. While these are reasonable approaches when only summary data are available, provided individual patient data are available an appealing alternative is to consider the associated times (progression and death) as arising from a three-state illness-death model (Broglia and Berry, 2009; Xu et al., 2010) as depicted in Figure 1. This model is highly useful when modelling the incidence of non-fatal events when mortality rates are appreciable and gives the most suitable framework for joint modelling of non-terminal and terminal events.

While deaths are observed subject to right-censoring, the precise times at which the criteria for progression are satisfied are never known since progression status is only determined at periodic assessment times. A further complication is that even progression status is indeterminate for individuals who were progression-free at their last assessment before their death. Zeng et al. (2015) study the asymptotic bias of the Kaplan-Meier estimator of the progression-free survival distribution and the estimator of treatment effect from a Cox regression model when the time that progression is detected is taken as the actual progression time; the power implications of such analyses are also examined. Frydman and Szarek (2009) consider this problem and discuss non-parametric estimation of the progression-free survival time distribution. Boruvka and Cook (2016) consider semiparametric methods for the incorporation of treatment effects, which are most appealing when the assessment times are highly irregular as in uncontrolled (e.g. registry-based) observational studies. When assessment times are regularly scheduled and common across individuals, identifiability problems can arise with general semiparametric models. We focus on parametric models for fitting the three-state process in Figure 1.

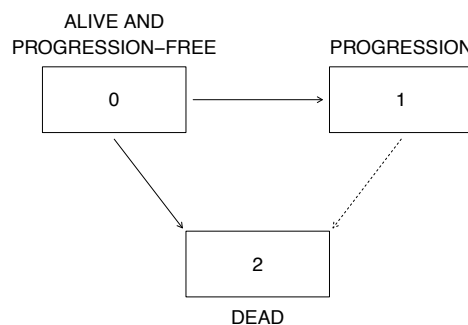


Figure 1: A three-state model for joint consideration of progression and death.

We derive sample size criteria for randomized controlled trials aiming to model treatment effects on progression-free survival in the context of this three-state model in Figure 1. We adopt Markov intensities with multiplicative effects of treatment. When the treatment effect for the transition intensity to the progression state is the same as the treatment effect on the transition intensity from the initial state to the death state, a Cox model for the progression-free survival time can in principle yield a consistent estimate for this common effect. In this framework, we construct the likelihood for the setting in which progression status is under an intermittent observation scheme while survival times are subject to random right censoring over a period of interest following Jackson (2011). Sample size requirements are then derived to ensure power requirements are met for detecting a minimal clinically important effect. The required sample size is shown to vary according to the frequency of assess-

ments thus allowing one to weigh the options of more frequent assessments against the recruitment of additional patients.

The remainder of this paper is organized as follows. In Section 2 notation is defined, the three-state model is reviewed, the nature of the data is described, and the likelihood is constructed. Section 3 reviews the standard approach to analyze the composite progression-free survival time and the associated sample size formula and derives the formula based on the proposed model and likelihood. A more *ad hoc* approach to sample size adjustment is also given based on Cox model typically used in this setting, which is misspecified in this framework (Zeng et al., 2015). Section 4 discusses related design criteria for analyses geared simply towards progression. Simulation studies validating the derivations are reported in Section 5, and concluding remarks are made in Section 6.

2 MULTI-STATE MARKOV MODELS FOR DISEASE PROGRESSION AND DEATH

2.1 THE BASIC FORMULATION

Suppose individuals are in state 0 (alive and progression-free) at the time of randomization and may enter either state 1 upon progression or state 2 upon death. The possibility of death after progression is accommodated by allowing transition from state 1 to state 2. Let $T_{j\ell}$ denote the potential $j \rightarrow \ell$ transition time (i.e. time from leaving state j to entering state ℓ), so $T_1 = T_{01}$ is the time to progression, and $T_2 = I(T_{02} < T_{01})T_{02} + I(T_{01} \leq T_{02})(T_{01} + T_{12})$ is the overall survival time that can be heavily influenced by the intermediate event of cancer progression, where the function $I(\cdot)$ takes a value of 1 if the condition is true. The progression-free survival time is $T = \min(T_1, T_2)$.

We begin the discussion by considering a homogeneous population and omit consideration of covariates. The multistate models are defined by the transition intensities between the states $\mathcal{S} = \{0, 1, 2\}$ at any time t

$$\lim_{\Delta t \rightarrow 0} \frac{P(Z(t + \Delta t) = \ell \mid Z(t) = j, H(t))}{\Delta t} = \lambda_{j\ell}(t \mid H(t)), \quad j < \ell \in \mathcal{S}, \quad (1)$$

where $Z(t)$ denotes the state occupancy of the stochastic process at time t , $H(t) = \{Z(u), 0 < u < t\}$ indicates the history of the process containing information on the timing and nature of any transitions over $(0, t)$. The Markov assumption is perhaps most commonly adopted in multistate models where the transition intensities depend on the process history only through the current state and time, $\lambda_{j\ell}(t \mid H(t)) = \lambda_{j\ell}(t)$. For general Markov models, the transition probabilities $p_{j\ell}(s, t) = P(Z(t) = \ell \mid Z(s) = j)$ can be expressed as a complicated function of transition intensities by solving the Kolmogorov forward differential equation

$$\frac{d}{dt} \mathbb{P}(s, t) = \mathbb{P}(s, t) \mathbb{A}(t), \quad t > s \quad (2)$$

where $\mathbb{P}(s, t) = [p_{j\ell}(s, t)]$ is the transition probability matrix, $\mathbb{A}(t) = [\mathbb{A}_{j\ell}(t)]$ is the transition intensity matrix with entries $\mathbb{A}_{j\ell}(t) = \lambda_{j\ell}(t)$ for $j \neq \ell \in \mathcal{S}$ and $\mathbb{A}_{jj}(t) = -\sum_{\ell \neq j} \lambda_{j\ell}(t)$ (Cox and Miller, 1965). If the process is strictly progressive, $\lambda_{j\ell}(t) = 0$ for $j > \ell$ and the transition probabilities can be explicitly written as an integral of the transition intensities with respect to the possible times of intermediate events.

Under the three-state illness-death model, the transition intensities $\lambda_{01}(t)$ and $\lambda_{02}(t)$ are the cause-specific hazard rates for progression and death before progression, and $\lambda_{12}(t)$ is the intensity function for moving from the ‘‘progressed’’ state to the ‘‘dead’’ state among those who progress prior to death. For the composite progression-free survival endpoint, the survival function is thus

$$S(t) = P(Z(t) = 0 \mid Z(0) = 0) = \exp\left(-\int_0^t \lambda_{01}(s) + \lambda_{02}(s) ds\right),$$

and its hazard function is $h(t) = \lambda_{01}(t) + \lambda_{02}(t)$, $t > 0$.

2.2 LIKELIHOOD UNDER INTERMITTENT OBSERVATION OF PROGRESSION

Consider a clinical trial with a common administrative censoring time τ defining the planned period of observation $[0, \tau]$. Individuals may withdraw from the study prematurely at a random censoring time $C < \tau$ assumed to be independent of the disease process; we assume that the vital status is known over the resulting period of observation. Let $G(u)$ denote the distribution function for the random censoring time. Progression status is only assessed at a maximum of K fixed times $0 = a_0 < a_1 < \dots < a_K \leq \tau$. The assessment process for progression can of course be terminated by death or withdrawal from the study. For a continuous-time multistate process, this leads to a mixed observation scheme, that is the time of death is subject to right-censoring while the time of progression is subject to interval censoring.

Figure 2 illustrates the different types of data that may arise from such an observation scheme. Note that death can be observed with or without prior confirmation of progression. For an individual whose disease progression has been detected at the previous assessment (as the top two cases in Figure 2), we know that time to progression is interval-censored and it occurred before death. However, if the last assessment of an individual is negative (as the third and fourth cases in Figure 2, respectively), it is unknown whether or not progression has occurred after that assessment and prior to death or censoring; such an uncertainty, as reflected by the possible paths within the box in the figure, must be taken into account for likelihood construction. The last two cases in Figure 2 are for those who remained until the administrative end of the study and had their last assessment of progression at that time.

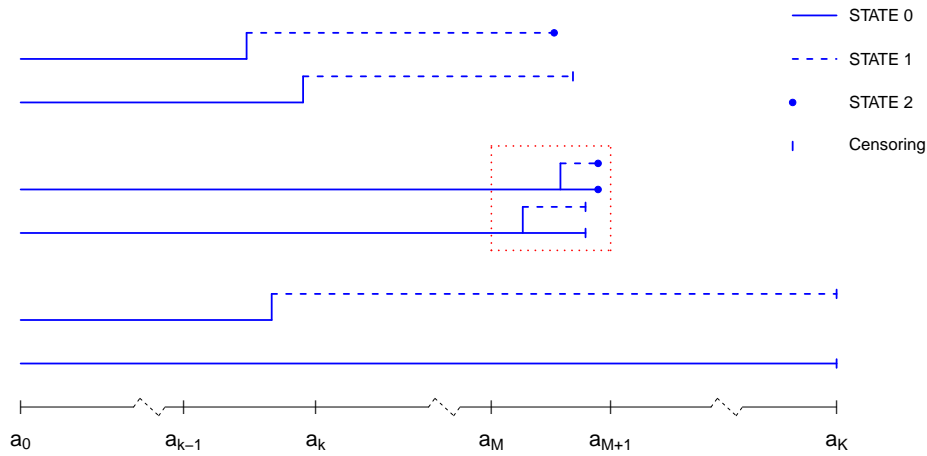


Figure 2: A three-state model with intermittent observations on individuals.

Let $T^\dagger = \min(T_2, C)$ denote the minimum of the time to death and censoring, resulting in $M = \max\{k; a_k < T^\dagger, k = 0, \dots, K\}$ as the random number of progression assessments for an individual. Let $0 = z_0 \leq z_1 \leq \dots \leq z_M \leq 1$ denote the observed disease status at the assessment times $0 = a_0 < a_1 < \dots < a_M$, and $\delta = I(T^\dagger = T_2)$ indicate that death is observed. If $\lambda_{j\ell}(t) = \lambda_{j\ell}(\boldsymbol{\theta})$ depends on q functionally independent parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ for each $j, \ell \in \mathcal{S}$, the likelihood contribution from a single individual can be written as

$$\begin{aligned}
 L(\boldsymbol{\theta}) &= \prod_{k=0}^{M-1} P(Z(a_{k+1}) = z_{k+1} \mid Z(a_k) = z_k) \sum_{\ell \in \{0,1\}} P(Z(t^\dagger) = \ell \mid Z(a_M) = z_M) \lambda_{\ell 2}^\delta(t^\dagger) \\
 &= \prod_{k=0}^{M-1} p_{z_k, z_{k+1}}(a_k, a_{k+1}) \sum_{\ell \in \{0,1\}} p_{z_M, \ell}(a_M, t^\dagger) \lambda_{\ell 2}^\delta(t^\dagger); \tag{3}
 \end{aligned}$$

where the summation in the second term accounts for the fact that the progression status right before death or censoring may not be known due to the intermittent observation. With a time-homogeneous process, the transition probability $p_{j\ell}(a_k, a_{k+1})$, $j, \ell \in \mathcal{S}$, relates to the constant intensity matrix $\mathbb{A}(t) = \mathbb{A}_0 = [\lambda_{j\ell}]$ through the matrix exponential

$$\mathbb{P}(s, t) = \exp\{(t - s)\mathbb{A}_0\}, \quad t > s.$$

The likelihood function for transition intensity parameter θ is simply a product of (3) over all individuals, and the estimates of θ can be obtained by maximizing such a function. These methods can be extended to fit certain non-homogeneous Markov models as discussed in Kalbfleisch and Lawless (1985). For example, one can assume that the time-dependent transition intensity matrix can be parameterized as $\mathbb{A}(t) = \mathbb{A}_0 \cdot g(t; \theta)$, a product of a fixed intensity matrix \mathbb{A}_0 and a known function of time $g(t; \theta)$ up to a parameter θ . Alternatively, a piecewise constant model can be adopted such that the transition matrix is constant between certain observation times but is allowed to change at observation times, for example, $\mathbb{A}(t) = \mathbb{A}_k$ if $t \in [a_k, a_{k+1})$, $k = 0, \dots, K$.

3 DESIGN METHODS FOR THE ENDPOINT OF PROGRESSION-FREE SURVIVAL

3.1 THE CONVENTIONAL APPROACH TO DESIGN

Consider randomized controlled cancer trials designed to compare a standard ($X = 0$) and a new experimental treatment ($X = 1$) with respect to a composite endpoint, progression-free survival. Suppose the underlying disease process can be described by the three-state Markov model described in Section 2, and the treatment is associated with the transition intensities in the multiplicative form such that

$$\lambda_{j\ell}(t; X) = \lambda_{j\ell}(t) \exp(\beta_{j\ell} X), \quad j, \ell \in \mathcal{S}, \quad (4)$$

where $(\beta_{01}, \beta_{02}, \beta_{12})$ is the vector of regression coefficients. The hazard for progression-free survival also has a multiplicative form

$$h(t; X) = h_0(t) \exp(\beta X), \quad (5)$$

if one of the following two conditions is satisfied: (i) $\beta_{01} = \beta_{02}$ or (ii) $\lambda_{01}(t) = \pi \lambda_{02}(t)$ where π is a proportionality constant (Wu and Cook, 2012). In the first scenario, the treatment effect of progression-free survival is the same on progression and death without progression, $\beta = \beta_{01} = \beta_{02}$, and the baseline hazard is simply the summation of the baseline cause-specific hazards of the two components (i.e. $h_0(t) = \lambda_{01}(t) + \lambda_{02}(t)$). In the second scenario of a common cause-specific baseline hazard for progression and death without progression, the multiplicative treatment effect on progression-free survival is $\beta = \log[\pi \exp(\beta_{01}) + \exp(\beta_{02})]$. Neither of these conditions are particularly plausible, but they are implicitly made if one accepts the illness-death model as a description of the disease process with multiplicative intensities.

This paper focuses on the design of cancer clinical trials based on progression-free survival. When the proportionality assumption holds for this endpoint, the usual approach for testing the treatment effect with null hypothesis $H_0: \beta = \beta_0$ is based on the Cox model (5) using the Wald test statistic

$$\frac{\hat{\beta} - \beta_0}{\text{se}(\hat{\beta})} \stackrel{H_0}{\rightsquigarrow} N(0, 1)$$

where $\hat{\beta}$ and $\text{se}(\hat{\beta})$ are the MLE and its standard error correspondingly. The resulting sample size formula is

$$n = \frac{(Z_{\alpha_1/2} + Z_{\alpha_2})^2}{\beta_A^2} \text{asvar}_A(\hat{\beta}) \quad (6)$$

for a two-sided Wald test of $H_0: \beta = 0$ at a significance level of α_1 and a power of $1 - \alpha_2$ for detecting an alternative treatment effect $\beta = \beta_A$. In (6) $\text{asvar}_A(\hat{\beta})$ is the asymptotic variance of the maximum likelihood estimate $\hat{\beta}$ evaluated at the alternative value β_A ; see Demidenko (2007). A central problem of this approach in this setting is that progression status is only intermittently assessed so the composite progression-free survival time is strictly subject to a hybrid censoring scheme. That is, progression times are interval-censored but the time of progression is taken as the assessment at which progression was detected or treated as right-censored if it is not observed. Convention also implicitly treats individuals as progression-free at the time of death if they have not been observed to have satisfied the condition for progression. The sample size formula (6) based on the Cox models is therefore not valid as it assumes a continuous follow up and the event times being exactly observed or right censored.

We develop two approaches for addressing this issue for a more rigorous treatment of the progression-free survival endpoint that accounts for the intermittent assessment on progression status. For the first, a sample size formula is derived based on the multistate Markov model that incorporates the mixed observation schemes for the component events of progression and death. The key step is an efficient procedure for calculating the Fisher-information matrix that only relies on the first derivatives of the log-likelihood function. We also recognize that Cox models are routinely used for the analysis of progression-free survival and the intermittent assessment of progression is routinely ignored through use of the surrogate progression time defined as the time of the first positive assessment. Zeng et al. (2015) showed that such an approach results in biases in the estimates of treatment effect and a loss of power for the associated test. For the second approach we propose a sample size adjustment based on misspecified Cox models by deriving the limiting behaviour of the naive estimator as in Zeng et al. (2015) and accommodating the bias and robust large sample variance in the calculations to ensure the power is maintained at the nominal level despite the model misspecification.

3.2 DESIGN APPROACH BASED ON MARKOV MODELS

We consider a simple time homogeneous three-state Markov model allowing progression, death prior to and post progression, and assume proportional treatment effects on each transition intensity as well as common effects $\beta = \beta_{01} = \beta_{02}$ so that the proportional hazard holds for the composite progression-free survival endpoint. We apply this model to the design of clinical trials for comparing the progression-free survival of two treatments when the progression status is only intermittently assessed. The model assumptions considered here are deliberately simple to illustrate the approach without introducing unnecessary complications and to provide an approach can be easily implemented in practice. Extensions to deal with time non-homogeneous Markov models are of course possible.

Under the Markov model, obtaining the maximum likelihood estimates and the associated asymptotic covariances matrix would require the evaluation of both first and second derivatives of the log-likelihood function given in Section 2.2, which can be quite cumbersome. Kalbfleisch and Lawless (1985) proposed an efficient estimation procedure for panel data that only requires the first derivatives. We can show that this is also the case with a mixed observation scheme with intermittent assessment of progression status and deaths observed subject to right-censoring. Let $\boldsymbol{\lambda} = (\lambda_{01}, \lambda_{02}, \lambda_{12})'$ denote the baseline transition intensities. At any given time t , let $Y^\dagger(t) = I(t < T^\dagger)$ indicate that an individual is alive and under observation, $Y_j(t) = I(Z(t) = j)$ indicate occupancy of state j , and $Y_j^\dagger(t) = Y(t)Y_j(t)$. The likelihood function for $\boldsymbol{\theta} = (\beta, \beta_{12}, \boldsymbol{\lambda})'$ can be written for the time-homogeneous illness-death model as

$$L(\boldsymbol{\theta}) = \prod_{k=0}^{K-1} \prod_{j=0}^1 L_{jk}^{Y_j^\dagger(a_k)}, \quad (7)$$

where L_{jk} denotes the likelihood contribution from the observation interval $[a_k, a_{k+1})$ of a subject

who is alive and in state j at a_k , which is of the form

$$L_{jk} = \prod_{\ell=0}^1 p_{j\ell}(a_k, a_{k+1})^{Y_{\ell}^{\dagger}(a_{k+1})} \left[\sum_{\ell=0}^1 p_{j\ell}(a_k, t^{\dagger}) (\lambda_{\ell 2} e^{\beta_{\ell 2} X})^{\delta} \right]^{1 - Y^{\dagger}(a_{k+1})}.$$

The `msm` package (Jackson, 2011) in R can be used to maximize this likelihood. The score function for parameter θ_r is then

$$S_r(\theta) = \frac{\partial \log L(\theta)}{\partial \theta_r} = \sum_{k=0}^{K-1} \sum_{j=0}^1 Y_j^{\dagger}(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r}, \quad (8)$$

where $\ell_{jk} = \log L_{jk}$. Since $\sum_{\ell=0}^2 p_{j\ell}(a_k, a_{k+1}) = 1$ it can be shown that

$$E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \mid Y_j^{\dagger}(a_k) = 1 \right] = 0 \quad (9)$$

for $j = 0, 1$ and $k = 0, \dots, K-1$; see Appendix A. This result implies that only the first derivatives are needed for the calculation of the Fisher information matrix $\mathcal{I}(\theta)$ with (r, q) element

$$\mathcal{I}_{rq} = E \left[- \frac{\partial^2 \log L(\theta)}{\partial \theta_r \partial \theta_q} \right] = \sum_{k=0}^{K-1} \sum_{j=0}^1 E \left[Y_j^{\dagger}(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \right]. \quad (10)$$

The expectation in (10) can be taken in a stepwise manner by conditioning first on random censoring time C , and then on $I(Z(a_k) = j)$ and finally death time T_2 . The final step requires averaging over the random censoring time C . We know that if $C \in A_k = [a_k, a_{k+1})$, there will be k possible follow up assessments for progression $0 < a_1 < \dots < a_k < C$ as long as the individual survives to a_k , hence the integration over random censoring time C is taken as a summation of integrals over these inter-assessment intervals. More specifically, we write

$$\mathcal{I}_{rq} = \sum_{m=0}^{K-1} \int_{a_m}^{a_{m+1}} \sum_{k=0}^m \sum_{j=0}^1 E \left[Y_j^{\dagger}(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \mid C = c \right] dG(c) + \sum_{k=0}^{K-1} \sum_{j=0}^1 E \left[Y_j^{\dagger}(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \right] \mathcal{G}(\tau), \quad (11)$$

where $\mathcal{G}(t) = 1 - G(t)$. The formulas for the conditional expectations are given in Appendix B.

With a sample of n independent individuals, the overall likelihood is a product of terms like (7) from each individual and the resulting estimation asymptotically follow $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \mathcal{I}^{-1}(\theta))$. So $\sqrt{n}(\hat{\beta} - \beta) \sim N(0, [\mathcal{I}^{-1}(\theta)]_{11})$. This can be used in the standard sample size formula given in (6).

3.3 SAMPLE SIZE ADJUSTMENT BASED ON MISSPECIFIED COX MODELS

As pointed out earlier, Cox models are commonly adopted for progression-free analysis and the right-endpoint imputation is routinely used to deal with intermittent assessment of progression. Specifically, the surrogate for the progression-free survival time T using right endpoint imputation is given by $S = \min\{a_k I(Z(a_k) = 1), k = 1, \dots, K\} \cdot I(Z(a_M) = 1) + T_2 \cdot I(Z(a_M) = 0)$, which takes the value of the assessment time if progression is detected, and the time of death otherwise. The progression-free survival is considered right-censored at the last negative assessment if the final assessment is negative and death has not been observed during the course of follow-up. The proportional assumption, although true for progression-free survival time, no longer holds for the surrogate event time S , and the working Cox model for the surrogate time S

$$h_S(s; x) = h_{S0}(s) \exp(\gamma X), \quad s > 0,$$

is thus a misspecified model with baseline hazard $h_{S0}(s)$ and treatment effect γ . The estimator $\hat{\gamma}$ is inconsistent, and its limiting value γ^* is the solution to the expectation of the partial score equation

$$E(U(\gamma)) = \int_0^\infty E [Y^\dagger(s)X dN(s)] - \frac{r^{(1)}(\gamma; s)}{r^{(0)}(\gamma; s)} E [Y^\dagger(s)dN(s)] = 0, \quad (12)$$

where $Y^\dagger(s) = I[s < \min(S, C)]$, $dN(s) = I(S = s)$ is the event indicator, and $r^{(\ell)}(\gamma, s) = E [Y^\dagger(s)X^\ell \exp(\gamma X)]$, $\ell = 0, 1$. Following Lin and Wei (1989), its asymptotic variance takes the form

$$\text{asvar}(\sqrt{n}(\hat{\gamma} - \gamma^*)) = [A^{-1}(\gamma^*)][B(\gamma^*)][A^{-1}(\gamma^*)]' \quad (13)$$

where

$$A(\gamma) = \int_0^\infty E \left\{ Y^\dagger(s) \left[\frac{r^{(1)}(\gamma, s)}{r^{(0)}(\gamma, s)} - \left(\frac{r^{(1)}(\gamma, s)}{r^{(0)}(\gamma, s)} \right)^2 \right] dN(s) \right\}$$

and

$$B(\gamma) = \int_0^\infty E \left\{ Y^\dagger(s) \left[X - \frac{r^{(1)}(\gamma, s)}{r^{(0)}(\gamma, s)} \right]^2 dN(s) \right\}.$$

Note that all the indicator variables are defined based on the imputed event time S , and the expectations can be taken with respect to its distribution either analytically or numerically.

The sample size to ensure $100(1 - \alpha_2)\%$ power for a two-sided test at the $100\alpha_1\%$ significance level under the misspecified Cox model is therefore

$$n = \frac{(Z_{\alpha_1/2} + Z_{\alpha_2})^2}{\gamma_A^2} \text{asvar}_{\gamma_A}(\hat{\gamma}), \quad (14)$$

where γ_A is the limiting value γ^* of the treatment effect estimator $\hat{\gamma}$ under the alternative hypothesis.

4 DESIGN METHODS FOR TIME-TO-PROGRESSION ENDPOINT

Time-to-progression (TTP), defined as the time from randomization until objective tumour progression, also serves as an endpoint to support cancer drugs and biologics approval.

From the pure definition point of view, the treatment effect on progression has a more direct and clearer interpretation than the one on the composite progression-free survival endpoint. This is especially the case when the treatment effects on progression and death without progression are not common so that proportional hazard is no longer true for the combined event of the two. From analytical point of view, there are important issues with TTP analyses. Death is always a competing risk, that is an event that either hinders the observation of the event of interest (i.e. progression) or modifies the chance that this event occurs. The traditional survival methods, commonly used for the analysis of TTP endpoint, do not consider death and the patients who die without progression are typically right censored either at the time of death or at a visit time when progression is intermittently assessed. Such an approach estimates the TTP in a hypothetical setting where patients could be at risk of progression after death. More importantly, it induces biases due to informative censoring (nonrandom pattern of loss from the study) when death and progression are related, which is often the case. TTP is hence considered a less preferred regulatory endpoint compared to progression-free survival (FDA, 2007, Fleming et al., 2009). Alternatively, competing risk methods have been used for time to progression analysis, which involve modelling both the rates of progression and its competing event death. The issue of this approach arises with an intermittent assessment scheme. For individuals who die without prior detected progression, simply recording them as experienced the competing event of death would obviously result in biased estimates unless one makes a hypothetical and unrealistic assumption that progression is not the cause for death.

The three-state illness-death models described in Section 2 serve as an appealing framework for the design and analysis of TTP endpoint. It considers progression, death with and without progression jointly, and it accounts for intermittent assessment of progression appropriately with flexible model specifications. Under the multiplicative transition intensity models (4), the focus lies in β_{01} which is the treatment effect on TTP. The asymptotic variance of the estimate $\hat{\beta}_{01}$ is simply the (β_{01}, β_{01}) element of the inverse of the Fisher information $\mathcal{I}(\theta)$, which again can be used in formula (6) to obtain the sample size required for testing the treatment effect on time to progression. The way to obtain the Fisher information $\mathcal{I}(\theta)$ has been described in Section 3.2.

5 FINITE SAMPLE PERFORMANCE OF ASYMPTOTIC RESULTS

5.1 EMPIRICAL STUDIES

In this section, we show some examples of sample size calculations for cancer trials based on progression-free survival with intermittent assessments for progression. We also evaluate the performance of the proposed method based on multi-state models via simulation studies.

For simplicity, we assume the disease process to be time-homogeneous with constant baseline intensities λ_{01} , λ_{02} and λ_{12} and treatment effects $\beta_{01} = \beta_{02} = \beta$ and β_{12} . Consider the design of a clinical trial where (1) patients are randomized to the treatment or control arm with equal probability; (2) disease progression status is to be assessed at K evenly spaced assessment times, i.e. $a_k = k\tau/K$, $k = 0, \dots, K$; (3) survival status is monitored continuously during the administration follow-up period $(0, \tau)$; and (4) patients may drop out of the study at a random time according to an exponential distribution with rate ρ . The sample size for the trial is based on testing the treatment effect on progression-free survival (which is parameterized in terms of coefficient β) using a two-sided Wald-test with type I error rate of $\alpha_1 = 0.05$ and power of $\{80\%, 90\%\}$.

For illustration purposes, we set the values for the required input design parameters λ_{01} , λ_{02} , λ_{12} and ρ to satisfy the following constraints: (i) $P(T > \tau, C > \tau \mid X = 0) = 2\%$ (percentage PFS right censored at the end of the study if untreated); (ii) $P(T > C, C < \tau \mid X = 0) = 38\%$ (percentage PFS right censored before the end of study if untreated); (iii) $P_1 = P(T_{01} < T_{02} \mid X = 0) = \lambda_{01}/(\lambda_{01} + \lambda_{02}) = \{0.6, 0.8\}$ (percentage progression among untreated); and (iv) $\lambda_{12}/\lambda_{02} = 1.5$ (relative risk of death with versus without progression among untreated). Note that conditions (i) and (ii) together imply a net right censoring rate of $\pi_N = 40\%$ for progression-free survival time T among untreated. We set $\beta_{01} = \beta_{02} = \beta = \log 0.75$ for a moderate treatment effect on progression-free survival, and $\beta_{12} = \log 1 = 0$ for null effect on death post progression. Without loss of generality, we assume equal probability for treatment, set $\tau = 1$ and use interval $[0, 1]$ to indicate the administrative study period. Given the treatment assignment, the times to progression and death are simulated based on the three-state illness-death model with multiplicative intensities. More specifically, T_{01} and T_{02} were simulated based on exponential distributions with rates $\lambda_{01} \exp(\beta X)$ and $\lambda_{02} \exp(\beta X)$, respectively, and if $T_{01} < T_{02}$, we simulate T_{12} according to an exponential distribution with rate $\lambda_{12} \exp(\beta_{12} X)$. We record the actual overall survival time as $T_2 = T_{02}I(T_{02} < T_{01}) + (T_{01} + T_{12})I(T_{02} > T_{01})$, and the progression-free survival time as $T = \min\{T_{01}, T_{02}\}$. We then consider progression being intermittently assessed and set the assessment times to be $a_k = k\tau/K + \epsilon_k$, $k = 0, \dots, K$, where minor variations are allowed by adding a normal noise $\epsilon_k \sim N(0, \sigma_e)$ with $\sigma_e = \tau/(20K)$. The random right censoring time C is simulated by an exponential distribution with the rate ρ . The progression status $Z(a_0), \dots, Z(a_M)$ is then obtained at the assessments $0 = a_0 < a_1 < \dots < a_M < \min\{C, T_2, \tau\}$. The surrogate progression-free survival time S is recorded in the conventional way in practice, that is when $Z(a_M) = 1$ then $S = \min\{a_k; I(Z(a_k) = 1), k = 1, \dots, M\}$, and when $Z(a_M) = 0$ then $S = T_2$ if $T_2 < C$, or censored at a_M otherwise.

We consider three different approaches for sample size calculation. First, we follow the conven-

tional way by using standard Cox models assuming progression-free survival will be continuously monitored and event times will be exactly observed, which gives required sample size of $n^* = 683$ and 914 to detect the treatment effect of $\beta = \log 0.75$ with a power of 80% and 90%, respectively. We then simulate event times for the given sample sizes but consider intermittent assessments for progression and estimate the treatment effect on progression-free survival using three different analysis approaches: a Cox model using the actual progression-free survival time (PFS_T); a Cox model on the surrogate progression-free survival time (PFS_S); and a time-homogenous illness-death model based on the actual mixed type observation scheme (MSM). Empirical percentage power (EP) for testing the treatment effect on progression-free survival based on 1000 simulations are reported under the ‘‘Design Based on PFS_T ’’ in Table 1. The Cox regression analysis of the actual event time yield estimates with negligible empirical biases and power at the nominal level as expected, although it is a purely hypothetical analysis and impossible to do under the intermittent assessment scheme in practice. The naive Cox analysis using surrogate PFS times resulted in biased estimates of treatment effect and lower power than one would wish for. The correct multistate analysis gave unbiased estimates with slightly larger standard errors and lower empirical power than the nominal level, reflecting the loss of information when event times are interval-censored. We then calculate sample sizes based on the time-homogeneous illness-death model and account for the mixed types of observations. As reported under ‘‘Design Based on MSM’’ in Table 1, the sample sizes (denoted by n) are much bigger than those based on standard Cox model assuming exact event times to be observed (denoted by n^*). The empirical power from fitting the correct multistate model on the observed data is very close to the desired nominal level, which validates the proposed calculation. The naive analysis using Cox model on surrogate event times results in a mild loss of power for the test of treatment effect, although one should keep in mind that the estimate itself is biased. Last but not least, we use the approach of adjusting the sample size based on misspecified Cox model. This gives the largest sample sizes compared to the other 2 approaches, and the simulation results imply that it can correct for the desired power for the test, but again the estimate remains biased.

5.2 APPLICATION TO THE DESIGN OF A METASTATIC BREAST CANCER TRIAL

Hortobagyi et al. (1996) report on a trial of women with stage IV breast cancer in which the primary objective was to evaluate the effect of a bisphosphonate called pamidronate on the incidence of skeletal complications due to metastatic bone disease. A secondary outcome was progression of bone disease, which was assessed using radiographic surveys scheduled at 3- to 6-month intervals over 30 months of follow-up. This observation scheme resulted in an interval-censored time of bone progression and right censored survival times. Mortality was appreciable in this population of patients with metastatic cancer, so the composite endpoint of progression-free survival was adopted. The sample involves 378 breast cancer patients with bone metastases. Ultimately, 133 patients were documented to have progressed and among these, 108 died during the observation period. Among the remaining 245 patients, 181 died with an unknown progression status.

We consider data from this trial to estimate the parameters of an illness-death model where the intermediate state represents bone progression and the absorbing state is entered upon death. The resulting parameter estimates can then be used for the design of a future study. In fitting this model, we assume a common treatment effect on the intensity for bone progression, as is implicit in the composite endpoint analysis. Using the likelihood in (3) to appropriately deal with the dual censoring scheme for progression and death, we fit a model with time-homogeneous intensities to the data from Hortobagyi et al. (1996). The estimates obtained for the baseline intensities are $\hat{\lambda}_{01} = 2.19 \times 10^{-3}$, $\hat{\lambda}_{02} = 1.45 \times 10^{-3}$ and $\hat{\lambda}_{12} = 2.33 \times 10^{-3}$; the coefficient for the effect of pamidronate on the composite endpoint of progression-free survival is $\hat{\beta} = -0.261$ (SE = 0.112, p -value=0.020) and the corresponding effect of treatment on death following bone progression is $\hat{\beta}_{12} = 0.009$ (SE = 0.177, p -value = 0.959). While there is evidence that pamidronate reduces the risk of the composite endpoint

Table 1: Sample sizes and empirical power for testing treatment effect on progression-free survival under intermittent assessment scheme based on different settings and approaches to analysis and design.

Power (%)	K	P_1^d	Design Based on PFS $_T^a$				Design Based on MSM b			Design Based on PFS $_S^c$	
			n^*	EP (%)			n	EP (%)		n	EP (%)
				PFS $_T$	PFS $_S$	MSM		PFS $_S$	MSM		PFS $_S$
80	4	0.6	683	80.3	72.1	74.5	780	76.7	79.2	853	79.8
		0.8	683	80.8	68.4	72.8	818	78.7	81.0	901	80.5
	8	0.6	683	81.1	77.7	78.5	724	81.2	81.2	744	81.7
		0.8	683	81.4	76.7	78.6	740	79.5	80.2	765	81.6
90	4	0.6	914	90.4	81.4	85.4	1044	87.4	89.6	1142	90.5
		0.8	914	90.8	81.1	84.8	1095	87.6	89.9	1206	91.1
	8	0.6	914	90.8	89.0	90.4	969	90.5	90.7	995	91.2
		0.8	914	90.7	86.9	88.0	990	88.8	89.5	1024	90.7

a PFS $_T$: Cox model using the actual progression-free survival time T ;

b MSM: Time-homogeneous illness-death model with intermittent observation of progression and exact observation on death with right censoring;

c PFS $_S$: Cox model on the surrogate progression-free survival time S with right-point imputation; and

$$^dP_1 = P(T_{01} < T_{02} \mid X = 0) = \lambda_{01}/(\lambda_{01} + \lambda_{02})$$

Table 2: Sample size requirement for a randomized trial for testing an effect on bone progression-free survival

Power (%)	Cens (%) a	K	n	Power (%)	Cens (%)	K	n
80	6	5	502	90	6	5	672
		10	495			10	663
	20	5	610		20	5	816
		10	590			10	790

a Cens is the net censoring rate (administrative and random) for the latent progression-free survival time.

of bone progression-free survival, we consider the hypothetical situation in which the goals was to design a future randomized trial to test this effect.

Using the estimates obtained for λ_{01} , λ_{02} , λ_{12} , β and β_{12} , we perform sample size calculations based on Section 3.2. We consider a study with administrative censoring planned at $\tau = 890$ days (~ 30 months), which leads to a 6% administrative censoring rate for the real progression-free survival time. We consider the possibility of loss to follow-up by introducing an exponential random censoring time with rate $\rho = 6.43 \times 10^{-4}$, which yields a net right-censoring rate of 20% ($\pi_N = 0.20$). We consider $K = 5$ or 10 equally spaced assessments for bone progression, which over 30 months makes

them every 3 months or every 6 months respectively. Table 2 reports the sample sizes required for the corresponding randomized trial to ensure a power of 80% and 90% to detect the effect of pamidronate on bone progression-free survival reflected by $\beta = -0.261$ when under a two-sided test at the 5% significance level. The findings suggest that for a given power the net censoring rate for progression-free survival is more influential on the sample size than the number of assessments.

6 DISCUSSION

Standard analyses of progression-free survival yield biased estimators of the cumulative distribution function when the first documented time of progression is treated as the progression time. Kaplan-Meier estimates often exhibit a clustering of progression times around the scheduled assessments, reflecting the fact that the times are in part driven by the imaging schedule. While the bias is apparent in these plots, naive Kaplan-Meier estimates are ubiquitous in articles reporting on cancer clinical trials. Treating the progression time as interval-censored is more appropriate, which yields a composite endpoint with one component (progression) subject to interval censoring and another (death) to right censoring. The bias in the regression coefficient estimator of a semiparametric Cox model is examined in Zeng et al. (2015), and Boruvka and Cook (2016) discuss semiparametric estimation for this problem. The purpose of this article was to develop valid design criteria when analyses are appropriately based on an illness-death model with the progression status assessed intermittently and deaths are subject to right-censoring. We restricted attention to the setting of time-homogeneous transition intensities. If there is evidence of trend in the event intensities over the planned period of observation time non-homogeneous intensity functions can be adopted. At the design stage, the methods for sample size calculation, which we presented in Section 3.2 can be adapted to deal with non-homogeneous intensities by assuming either $\mathbb{A}(t) = \mathbb{A}_0 \cdot g(t; \theta)$ where \mathbb{A}_0 is a constant matrix and $g(t; \theta)$ is a known function of time up to a parameter θ , or piecewise-constant intensities for which $\mathbb{A}(t) = \mathbb{A}_k$ if $t \in [a_k, a_{k+1})$, $k = 1, \dots, K - 1$ where $0 = a_0 < a_1 < \dots < a_K = \tau$ define the break-points (assumed common here) over the planned period of observation. The likelihood function and the Fisher information can be constructed as in (7) and (10), respectively, but the calculation of the transition probabilities $p_{j\ell}(a_k, a_{k+1})$ and their first derivatives are slightly more complicated. With piecewise-constant intensities the closed-form expressions can still be obtained. While these more flexible intensities are appealing at the analysis stage, when planning a study much more information would be required to derive the sample size as the values for the piecewise-constant intensities would have to be specified for each transition and each interval of constant risk. We focussed on the setting with time homogenous intensities to avoid this at the planning stage but we note that at the analysis stage models with piecewise-constant intensities can be easily fitted using the `msm` package (Jackson, 2011). Software for computing sample size are available from the corresponding author upon request.

In some settings progression may be associated with symptoms, which might lead to clinical assessments at times other than the scheduled assessments; these are sometimes referred to as disease-driven assessments (Pullenayegum and Lim, 2016). The consequences of introducing additional disease process-related assessments depends on the frameworks for inference and the particular targets of inferences (Cook and Lawless, 2017). Zeng et al. (2015) found that the biases indeed for both the Kaplan-Meier estimator of the progressive-free survival distribution and the regression parameter in the Cox model were smaller when such visits were possible. This reduction in the bias arises because the lag between the actual and the documented progression time is smaller on average as a result of this.

The sample size derivations here illustrate that the power of a particular trial for detecting treatment effects on progression-free survival times can be increased both by increasing the number of participants and by increasing the frequency of the examination times for progression; the incremental gain in power from increasing the frequency of the assessments is comparatively low. In settings

where patients are easy to identify and recruit, the former is typically adopted. Our derivations show that power can be improved for a given sample size by increasing the frequency of examination for progression. The costs of increasing the frequency of assessments may be appreciable, however, in terms of both health service expenditures and inconvenience to patients and these trade-offs must be weighted, when planning a study. While we have restricted attention to evenly spaced assessment times, there may be merit to considering assessments scheduled more intensively around the times that events tend to occur. Use of easily available markers to dynamically schedule radiographic examinations in periods where biomarkers (e.g. circulating tumour cell counts) are suggestive of increased tumour activity can also be explored. Inverse intensity based weights (Lin et al., 2014) would be necessary to use to adjust for this observation scheme since the treatment effect would now be examined conditional on biomarker values and the inspection process would become dependent on the progression time.

ACKNOWLEDGEMENTS

This research was supported by Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to L. Zeng (RGPIN 115928) and R. J. Cook (RGPIN 155849) and the Canadian Institutes for Health Research to R. J. Cook (FRN 13887). R. J. Cook is a Canada Research Chair in Statistical Methods for Health Research.

REFERENCES

- Ballman, K. V., Buckner, J. C., Brown, P. D., Giannini, C., Flynn, P. J., LaPlant, B. R., and Jaeckle, K. A. (2007). The relationship between six-month progression-free survival and 12-month overall survival end points for phase II trials in patients with glioblastoma multiforme. *Neuro-Oncology*, 9(1):29–38.
- Booth, C. M. and Eisenhauer, E. A. (2012). Progression-free survival: meaningful or simply measurable? *Journal of Clinical Oncology*, 30(10):1030–1033.
- Boruvka, A. and Cook, R. J. (2016). Sieve estimation in a Markov illness-death process under dual censoring. *Biostatistics*, 17(2):350–363.
- Broglio, K. R. and Berry, D. A. (2009). Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute*, 101(23):1642–1649.
- Cook, R. and Lawless, J. (2017). Analysis of chronic disease processes based on cohort and registry data. In Adhikari, A., Adhikari, M., and Chaubey, Y., editors, *Mathematical and Statistical Applications in Life Sciences and Engineering*, chapter 15. Springer Nature Singapore Pte Ltd., Singapore.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34(2):187–220.
- Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Wiley, New York, first edition.
- Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in Medicine*, 26(18):3385–3397.
- FDA (2007). *Guidance for industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*. Clinical/Medical Guidances. U.S. Department of Health and Human Services.

- Fleming, T. R., Rothmann, M. D., and Lu, H. L. (2009). Issues in using progression-free survival when evaluating oncology products. *Journal of Clinical Oncology*, 27(17):2874–2880.
- Frydman, H. and Szarek, M. (2009). Nonparametric estimation in a Markov “illness-death” process from interval censored observations with missing intermediate transition status. *Biometrics*, 65(1):143–151.
- Hortobagyi, G., Theriault, D., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J., and Seaman J, Knight RD, Heffernan M and Reitsma DJ (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *The New England Journal of Medicine*, 335(24):1785–1791.
- Jackson, C. H. (2011). Multi-state models for panel data: The msm package for r. *Journal of Statistical Software*, 38(8):1–28.
- Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80(392):863–871.
- Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408):1074–1078.
- Lin, H., Scharfstein, D., and Rosenheck, R. (2014). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society. Series B*, 66(3):791–813.
- Miksad, R. A., Zietemann, V., Gothe, R., Schwarzer, R., Conrads-Frank, A., Schnell-Inderst, P., Stollenwerk, B., and Siebert, U. (2009). Progression-free survival as a surrogate endpoint in advanced breast cancer. *International Journal of Technology Assessment in Health Care*, 24(4):371–383.
- Pullenayegum, E. and Lim, L. (2016). Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statistical Methods in Medical Research*, 25(6):2992–3014.
- Saaed, E. D., Katz, A., and Buyse, M. (2010). Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. *Journal of Clinical Oncology*, 28(11):1958–1962.
- Wu, L. and Cook, R. J. (2012). Misspecification of cox regression models with composite endpoints. *Statistics in Medicine*, 31(28):3545–3562.
- Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*, 6(3):716–725.
- Zeng, L., J. D., Cook, R. J., Wen, L., and Boruvka, A. (2015). Bias in progression-free survival analysis due to intermittent assessment of progression. *Statistics in Medicine*, 34(24):3181–3193.

APPENDIX A: PROOF OF RESULTS IN SECTION 3.2

Recall $T^\dagger = \min(T_2, C)$ and $\delta = I(T_2 < C)$. In addition, we let $Y^\dagger(t) = I(t < T^\dagger)$ indicate an individual is alive and under observation, $Y_j(t) = I(Z(t) = j)$ indicate occupancy of state j , and $Y_j^\dagger(t) = Y(t)Y_j(t)$. For the three-state Markov model with the mixed types observation scheme of this paper, the score for β (the treatment effect on progression-free survival) is the form

$$S_1(\boldsymbol{\theta}) = \sum_{k=0}^{K-1} \sum_{j=0}^1 Y_j^\dagger(a_k) \frac{\partial \ell_{jk}}{\partial \beta},$$

where

$$\frac{\partial \ell_{jk}}{\partial \beta} = \sum_{\ell=0}^1 Y_\ell^\dagger(a_{k+1}) \frac{\partial \log p_{j\ell}(a_k, a_{k+1})}{\partial \beta} + \left[1 - Y^\dagger(a_{k+1})\right] \frac{\partial}{\partial \beta} \log \left[\sum_{\ell=0}^1 p_{j\ell}(a_k, t^\dagger) \lambda_{\ell 2}^\delta(t^\dagger) \right].$$

With a time-homogeneous model, the transition intensities take the form $\lambda_{01}(t) = \lambda_{01}e^{\beta X}$, $\lambda_{02}(t) = \lambda_{02}e^{\beta X}$ and $\lambda_{12}(t) = \lambda_{12}e^{\beta_{12}X}$. The transition probabilities $p_{j\ell}(t) = 0$ when $j > \ell$ and

$$\begin{aligned} p_{00}(t) &= \exp \left[-t(\lambda_{01}e^{\beta X} + \lambda_{02}e^{\beta X}) \right] \\ p_{11}(t) &= \exp \left[-t\lambda_{12}e^{\beta_{12}X} \right] \\ p_{01}(t) &= \frac{\lambda_{01}e^{\beta X}}{\lambda_{01}e^{\beta X} + \lambda_{02}e^{\beta X} - \lambda_{12}e^{\beta_{12}X}} \left[e^{-t\lambda_{12}e^{\beta_{12}X}} - e^{-t(\lambda_{01}e^{\beta X} + \lambda_{02}e^{\beta X})} \right]. \end{aligned}$$

Consider a simple example of observing $Z(a_0) = 0$, $Z(a_1) = 1$ and then death at t_2 , where $a_1 < t_2 < a_2$, we obtain the score function

$$S_1(\boldsymbol{\theta}) = \frac{\partial \log p_{01}(a_1 - a_0)}{\partial \beta} + \frac{\partial}{\partial \beta} \log \left[p_{11}(t_2 - a_1) \lambda_{12}e^{\beta_{12}X} \right] = \frac{\partial \log p_{01}(a_1 - a_0)}{\partial \beta}.$$

To prove the result

$$E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \mid Y_j^\dagger(a_k) = 1 \right] = 0, \quad j \neq 2,$$

we assume the distribution of the censoring time C is independent of that of the disease process. For $C = c$ we consider the case when $c > a_{k+1}$ and $c \in [a_k, a_{k+1}]$ separately.

When $c > a_{k+1}$:

Note that $Y^\dagger(a_{k+1}) = 0$ implies $a_k < T_2 < a_{k+1} < c$, and the conditional distribution of T_2 is

$$f(t \mid a_k < T_2 < a_{k+1} < c) = \frac{\sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t)}{p_{j2}(a_k, a_{k+1})}.$$

Thus the conditional expectation becomes

$$\begin{aligned} & E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \mid c > a_{k+1}, Y_j^\dagger(a_k) = 1 \right] \\ &= \sum_{\ell=0}^1 \frac{\partial}{\partial \theta_r} p_{j\ell}(a_k, a_{k+1}) + \int_{a_k}^{a_{k+1}} \left[\frac{\partial}{\partial \theta_r} \log \sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t) \right] \sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t) dt \\ &= \sum_{\ell=0}^2 \frac{\partial p_{j\ell}(a_k, a_{k+1})}{\partial \theta_r} \end{aligned}$$

which equals to zero because $\sum_{\ell=0}^2 p_{j\ell}(a_k, a_{k+1}) = 1, j \neq 2$.

When $c \in A_k = [a_k, a_{k+1})$:

$$\frac{\partial \ell_{jk}}{\partial \beta} = \frac{\partial}{\partial \beta} \log \left[\sum_{\ell=0}^1 p_{j\ell}(a_k, t^\dagger) \lambda_{\ell 2}^\delta(t^\dagger) \right].$$

Realize that when $\delta = 0$ we have $T^\dagger = c$ and $P(\delta = 0 \mid Y_j^\dagger(a_k) = 1) = \sum_{\ell=0}^1 p_{j\ell}(a_k, c)$, whereas when $\delta = 1$, we have $a_k < T_2 < c < a_{k+1}$ and $T^\dagger = T_2$ with a conditional distribution

$$f(t \mid a_k < T_2 < c < a_{k+1}) = \frac{\sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t)}{P(T_2 < c \mid Y_j^\dagger(a_k) = 1)}.$$

Thus, we can show that

$$\begin{aligned} & E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \mid c \in A_k, Y_j^\dagger(a_k) = 1 \right] \\ &= \sum_{\ell=0}^1 p_{j\ell}(a_k, c) \left[\frac{\partial}{\partial \theta_r} \log \sum_{\ell=0}^1 p_{j\ell}(a_k, c) \right] + \int_{a_k}^c \left[\frac{\partial}{\partial \theta_r} \log \sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t) \right] \sum_{\ell=0}^1 p_{j\ell}(a_k, t) \lambda_{\ell 2}(t) dt \\ &= \frac{\partial}{\partial \theta_r} \sum_{\ell=0}^2 p_{j\ell}(a_k, C) \end{aligned}$$

which again equals to zero.

APPENDIX B: CALCULATION OF FISHER INFORMATION MATRIX $\mathcal{I}(\boldsymbol{\theta})$

We derive the form of the element (11) in the Fisher information matrix,

$$\begin{aligned} \mathcal{I}_{rq} &= \sum_{m=0}^{K-1} \int_{a_m}^{a_{m+1}} \sum_{k=0}^m \sum_{j=0}^1 E \left[Y_j(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \mid C = c \right] dG(c) + \sum_{k=0}^{K-1} \sum_{j=0}^1 E \left[Y_j(a_k) \frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \right] \mathcal{G}(\tau) \\ &= \sum_{m=0}^{K-1} \int_{a_m}^{a_{m+1}} \sum_{k=0}^m \sum_{j=0}^1 p_{0j}(a_k) E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \mid C = c, Z(a_k) = j \right] dG(c) + \sum_{k=0}^{K-1} \sum_{j=0}^1 p_{0j}(a_k) E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \mid Z(a_k) = j \right] \mathcal{G}(\tau) \end{aligned}$$

where the expectation is taken by conditioning on the random censoring time C and averaging over its distribution.

When the random censoring time $C = c \in A_m$, for the $k = 0, 1, \dots, m-1$ observation intervals we have

$$\begin{aligned} & E \left[\frac{\partial \ell_{jk}}{\partial \theta_r} \frac{\partial \ell_{jk}}{\partial \theta_q} \mid k < m, c \in A_m, Z(a_k) = j \right] \\ &= \sum_{\ell=0}^1 \frac{\frac{\partial}{\partial \theta_r} p_{j\ell}(a_k, a_{k+1}) \frac{\partial}{\partial \theta_q} p_{j\ell}(a_k, a_{k+1})}{p_{j\ell}(a_k, a_{k+1})} + \int_{a_k}^{a_{k+1}} \frac{\left(\frac{\partial}{\partial \theta_r} \sum_{\ell=0}^1 p_{j\ell}(a_k, u) \lambda_{\ell 2}(u) \right) \left(\frac{\partial}{\partial \theta_q} \sum_{\ell=0}^1 p_{j\ell}(a_k, u) \lambda_{\ell 2}(u) \right)}{\sum_{\ell=0}^1 p_{j\ell}(a_k, u) \lambda_{\ell 2}(u)} du, \end{aligned}$$

and for the last observation interval ($k = m$), we have

$$\begin{aligned} & E \left[\frac{\partial \ell_{jm}}{\partial \theta_r} \frac{\partial \ell_{jm}}{\partial \theta_q} \mid c \in A_m, Z(a_m) = j \right] \\ &= \int_{a_m}^c \frac{\frac{\partial}{\partial \theta_r} \sum_{\ell=0}^1 p_{j\ell}(a_m, u) \lambda_{\ell 2}(u) \frac{\partial}{\partial \theta_q} \sum_{\ell=0}^1 p_{j\ell}(a_m, u) \lambda_{\ell 2}(u)}{\sum_{\ell=0}^1 p_{j\ell}(a_m, u) \lambda_{\ell 2}(u)} du + \frac{\frac{\partial}{\partial \theta_r} \sum_{\ell=0}^1 p_{j\ell}(a_m, c) \frac{\partial}{\partial \theta_q} \sum_{\ell=0}^1 p_{j\ell}(a_m, c)}{\sum_{\ell=0}^1 p_{j\ell}(a_m, c)}. \end{aligned}$$