

Bias in progression-free survival analysis due to intermittent assessment of progression

Leilei Zeng,^{*†} Richard J. Cook, Lan Wen and Audrey Boruvka

Cancer clinical trials are routinely designed to assess the effect of treatment on disease progression and death, often in terms of a composite endpoint called progression-free survival. When progression status is known only at periodic assessment times, the progression time is interval censored, and complications arise in the analysis of progression-free survival. Despite the advances in methods for dealing with interval-censored data, naive methods such as right-endpoint imputation are widely adopted in this setting. We examine the asymptotic and empirical properties of estimators of the marginal progression-free survival functions and associated treatment effects under this scheme. Specifically, we explore the determinants of the asymptotic bias and point out that there is typically a loss in power of tests for treatment effects. © 2015 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Keywords: illness-death process; intermittent observation; progression-free survival; model misspecification; multistate model

1. Introduction

A fundamental goal in oncology is the reduction of mortality due to cancer. Therapeutic advances for many cancers and the increasing pressure to evaluate experimental treatments in a timely and cost-effective manner have made it challenging to design adequately powered trials based on the time from randomization to death. This has led to increased use of the composite progression-free survival endpoint [1], where the nominal response is the time from randomization to the first of progression or death. From 2005 to 2009 inclusive, for example, just over one-quarter of the trials published on breast, colorectal and non-small-cell lung cancer in the *Journal of Clinical Oncology* assess treatment effects based on the composite endpoint of progression-free survival [2].

The utility and interpretation of findings based on progression-free survival are currently being discussed in the oncology literature [3]. Miksad *et al.* [4] point out that because interest primarily lies in the effect of treatment on overall survival, there is an implicit assumption that progression-free survival is a surrogate for overall survival. These authors conduct a scientific review to examine the empirical evidence of an association between findings from progression-free survival and overall survival analyses; they conclude that while there is an association between the findings, the predictive accuracy is not high. Sidhu *et al.* [5] study this issue in metastatic colorectal cancer trials involving fluoropyrimidine-based regimens and found a higher correlation, leading them to conclude that progression-free survival is a valid surrogate for overall survival in this context. In another context, Ballman *et al.* [6] examine the association between 6-month progression-free survival and overall survival at 12 months. With a view to gaining insight into the structural nature of the relation between these two endpoints, Broglio and

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

*Correspondence to: Leilei Zeng, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

†E-mail: lzeng@uwaterloo.ca

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

The copyright line for this article was changed on 18 June 2015 after original online publications.

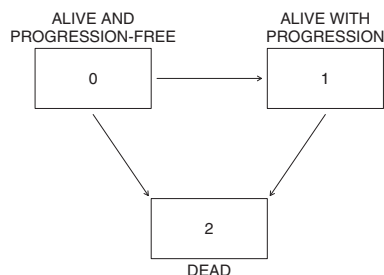


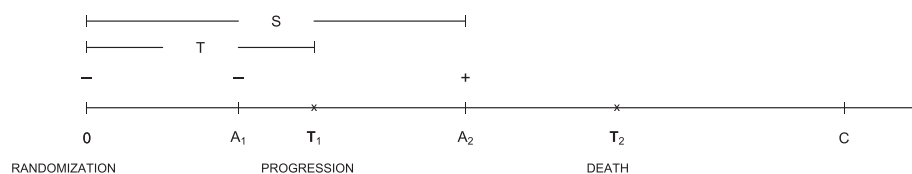
Figure 1. A three-state model for joint consideration of progression and death.

Berry [7] decompose the effect of treatment on overall survival into components, one of which relates to progression-free survival. Saad *et al.* [8] emphasize the importance of a clear definition of progression when analysed on its own or as part of a composite endpoint.

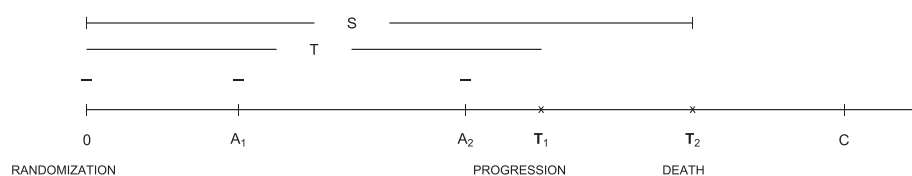
Frydman and Szarek [9] point out that the classical illness-death model (Figure 1) offers a natural framework for joint consideration of a non-terminal event, such as progression, and the terminal event death. In the current setting, we consider individuals to be in state 0 at the time of randomization when they are progression free and alive. Individuals enter state 1 upon progression and subsequently enter state 2 upon death. The possibility of death without progression is accommodated by direct transition from state 0 to state 2. Let T_{jk} denote the potential $j \rightarrow k$ transition time. If $T_{01} < T_{02}$, we let $T_1 = T_{01}$ denote the time of entry to state 1. The time of entry to state 2 is the overall survival time denoted by $T_2 = I(T_{02} < T_{01})T_{02} + I(T_{01} \leq T_{02})T_{12}$. The progression-free survival time is the time of exiting state 0 denoted by $T = \min(T_{01}, T_{02}) = \min(T_1, T_2)$.

Challenges arise when progression-free survival is used as an endpoint, as progression status is only observed at periodic assessment times and the time of progression is at best interval censored. The time to the composite progression-free survival endpoint is then subject to a hybrid censoring scheme involving interval censoring for progression and right censoring for death.

Figure 2(a) shows a timeline diagram where T_1 and T_2 are the times of progression and death, respectively; A_k , $k = 1, 2, \dots$ denote the assessment times; and C is a right censoring time. In this scenario, progression is first detected at A_2 , so one knows that T lies in the interval $(A_1, A_2]$. The most common strategy for dealing with this type of data, however, is to use right-endpoint imputation whereby the surrogate $S = A_2$ is used in lieu of the event time T and standard survival analysis methods are adopted; this approach is recommended by the FDA [10]. A further complication arises when death occurs without prior evidence of progression (Figure 2(b)) as it is unknown whether or not progression occurred between the last negative assessment (A_2) and the time of death. The convention in this case is to assume individuals have not progressed and hence use $S = T_2$ as the progression-free survival time (e.g. see Dejardin *et al.* [11]). The strong implicit assumptions associated with the approaches described earlier warrant



(a) A timeline diagram for events in which progression is detected



(b) A timeline diagram for events in which progression is not detected due to death

Figure 2. Timeline diagrams for progression (T_1), death (T_2), assessments (A_1 and A_2) and censoring (C).

careful consideration when dealing with the intermittently observed progression component in the composite progression-free survival endpoint.

Given the vast number of cancer trials using progression-free survival as the primary endpoint, the challenges in interpreting findings based on composite endpoints and the seemingly self-evident shortcomings of using the documented time of progression rather than the actual progression time, it is surprising that this problem has not received more attention. Frydman and Szarek [9] and Lefonndré *et al.* [12] highlight and examine the biases that arise from some form of imputation in this setting. Panageas *et al.* [13] discuss the role of the assessment schedule on inferences regarding progression-free survival analysis and discuss, in particular, the phenomenon that the Kaplan–Meier (KM) estimate tends to drop around the time of scheduled assessments in a way that reflects both the risk of progression and the assessment process. Binder and Schumacher [14] consider modelling the time to progression and examine the impact of censoring progression times at death for individuals not observed to have progressed. Lefonndré *et al.* [12] advocate suitable methods for interval-censored data when modelling the joint distribution of progression and death times. Much of this work has been based on simulation studies.

In this article, we examine the asymptotic and empirical properties of estimators arising from use of the surrogate ‘observed’ progression-free survival times as is customarily carried out under the intermittent assessment scheme. We consider the settings where the assessment times are pre-scheduled, but there is modest variation between individuals in precisely when they occur, to reflect data arising in clinical trials. We also consider the setting in which assessment times are governed by an independent and possibly progression-dependent stochastic process to correspond to the less regulated assessment schemes of observational cohort studies. We formulate models for each component of the composite endpoint (progression and death) in the framework of an illness-death model with proportional intensities so that the proportional hazards assumption is satisfied for the endpoint of progression-free survival. Asymptotic calculations are carried out by considering an enlarged state space that allows joint consideration of the disease process and the assessment process. From this stochastic model, the distribution of the ‘observed’ progression-free survival time is derived, where it is defined by right-endpoint imputation when progression is detected and the survival time otherwise. We contrast this distribution with that of the latent progression-free survival time in a variety of scenarios. We also use the theory of misspecified Cox models to derive the asymptotic biases of the estimated treatment effects from a Cox proportional hazard model for progression-free survival [15–17].

The remainder of this paper is organized as follows. In Section 2, we introduce further notation for the three-state illness-death model, and we discuss issues in the use of ‘observed’ progression-free survival when progression is intermittently observed in Section 3. We derive the expected partial score functions, which are central to the derivation of the asymptotic bias in estimates of treatment effect from the Cox proportional hazards model. We also derive the asymptotic power associated with Wald-type tests of the null hypothesis of no treatment effect and demonstrate the potential reduction in power from adopting right-endpoint imputation. Empirical investigations are conducted in Section 4 to illustrate the finite sample properties, and concluding remarks are given in Section 5.

2. The Markov illness-death model for progression and death

We let $Z(t)$ denote the state occupied at time t , and $\{Z(u), 0 < u\}$ represent the corresponding stochastic process. At time $t > 0$, the history of the process, denoted as $H(t) = \{Z(u), 0 < u < t\}$, contains information on the timing and nature of any transitions over $(0, t)$. Transitions among the states $\mathcal{S} = \{0, 1, 2\}$ at any time t are governed by the transition intensities

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

where $Y_j(t) = I(Z(t^-) = j)$ indicates that the individual is ‘at risk’ of a transition out of state j at time t , $j = 0, 1$.

If the transition intensities depend on the process history only through state occupied at t^- , we write

$$\lambda_{jk}(t | H(t)) = \lambda_{jk}(t),$$

and the process is Markov. We can then let $\mathbb{A}(t) = [\lambda_{jk}(t)]$ denote the transition intensity matrix, where $\mathbb{A}_{jk}(t) = \lambda_{jk}(t)$ for $j < k \in \mathcal{S}$, $\mathbb{A}_{jj}(t) = -\sum_{k>j} \lambda_{jk}(t)$, and other elements are zero. Let $\mathbb{P}(s, t) = [p_{jk}(s, t)]$

denote the transition probability matrix with elements $p_{jk}(s, t) = P(Z(t) = k | Z(s) = j)$, $s < t$. The matrix $\mathbb{P}(s, t)$ satisfies the Kolmogorov forward differential equation

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

which enables one to express the transition probabilities in terms of intensity functions [18]. When the process is time homogeneous, the calculation becomes much simpler and $\mathbb{P}(t) = \exp(\mathbb{A}t)$. For the three-state illness-death models shown in Figure 1, the explicit expressions for $p_{00}(s, t)$ and $p_{01}(s, t)$ are

$$\begin{aligned} p_{00}(s, t) &= \exp\left(-\int_s^t [\lambda_{01}(u) + \lambda_{02}(u)] du\right), \\ p_{01}(s, t) &= \int_s^t \lambda_{01}(u) \exp\left(-\int_s^u [\lambda_{01}(v) + \lambda_{02}(v)] dv\right) \exp\left(-\int_u^t \lambda_{12}(v) dv\right) du. \end{aligned} \quad (3)$$

Note that the survival function for the progression-free survival endpoint is then

$$\mathcal{F}(t) = P(T > t) = p_{00}(0, t), \quad t > 0, \quad (4)$$

with respective hazard function $h(t) = -d \log(\mathcal{F}(t))/dt = \lambda_{01}(t) + \lambda_{02}(t)$, the sum of the cause-specific hazard functions for $0 \rightarrow 1$ and $0 \rightarrow 2$ transitions.

At times in what follows, it will also be useful to consider the counting process formulation of multi-state processes, and with that in mind, we define $N_{jk}(t)$ as the number of $j \rightarrow k$ transitions over $(0, t]$ and let $\Delta N_{jk}(t) = N_{jk}(t + \Delta t^-) - N_{jk}(t^-)$ count the number of $j \rightarrow k$ transitions over the interval $[t, t + \Delta t)$. Then $dN_{jk}(t) = \lim_{\Delta t \rightarrow 0} \Delta N_{jk}(t)$ indicates the occurrence of a $j \rightarrow k$ transition at time t .

3. Progression-free survival with intermittent assessment schemes

3.1. Distribution of imputed progression-free survival time

When interest lies in evaluating treatment effects on the basis of progression-free survival, standard survival models are often fitted with the time $T = \min(T_1, T_2)$ in mind. As discussed in Section 2, progression status is only determined at periodic assessment times, and hence, among individuals for whom progression is detected, the progression time is interval censored. For such individuals, the convention is to use the assessment time at which progression is detected as the progression-free survival time. For individuals dying without progression having previously been detected, it is unknown whether or not progression occurred, and in this case, it is conventional to assume they were progression-free at the time of death and to take the time of death as the progression-free survival time. These conventions lead to invalid inference because the distributions of the adopted event time and the actual progression-free survival time differ. In what follows, we formally construct a probability model that represents the process generating the available data. We then use this model to evaluate the performance of methods used in this situation.

Let C_A denote an administrative censoring time and R_i a random right censoring time for individual i , so that the survival status is known over the interval $(0, C_i]$, where $C_i = \min(C_A, R_i)$. Suppose progression status is assessed at a sequence of times $0 = a_{i0} < a_{i1} < \dots < a_{iK_i} < C_i$. The surrogate for the progression-free survival time T is given by

$$S = \min\{a_{ik}; I(Z(a_{ik}) = 1), k = 1, \dots, K_i\} \cdot I(Z(a_{iK_i}) = 1) + T_2 \cdot I(Z(a_{iK_i}) = 0), \quad (5)$$

which takes the value of the assessment time if progression is detected, and the time of death otherwise. The progression-free survival time 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 [11]. We derive the distribution function for the surrogate event time S under two different intermittent assessment schemes, deferring discussion of covariates until Section 3.2.

The first scenario is designed to mimic the setting of a randomized trial and the second a less regular assessment scheme such as one might see in a cohort study. The prostate cancer trial by Scher *et al.* [19] is a suitable example in which progression is evaluated every month during the first 6 months of follow-up and every 3 months thereafter. For ease of exposition, we assume there is an assessment at the time of randomization denoted $a_0 = 0$, and K regularly spaced follow-up assessments are scheduled over $(0, C_A]$

at $a_k = kC_A/K, k = 1, \dots, K$. According to (5), the surrogate progression-free survival time S is one of these fixed assessment times if progression is detected, but it otherwise takes on any value over $(0, C_A]$ if death is observed. This S is a mixed-type non-negative random variable with positive mass at a finite number of discrete points (at the assessment times) and a cumulative distribution function

$$F_S(s) = P(S \leq s) = \int_0^s f(u)du + \sum_{k: a_k \leq s} P(S = a_k),$$

where

$$f(u) = p_{00}(1a_0, a_{k-1}) \left[\sum_{j=0}^1 p_{0j}(a_{k-1}, u^-) \lambda_{j2}(u) \right] \text{ for } u \in (a_{k-1}, a_k),$$

$$P(S = a_k) = p_{00}(a_0, a_{k-1})p_{01}(a_{k-1}, a_k),$$

and $p_{00}(u, v)$ and $p_{01}(u, v)$ are given in (3). The cumulative function is a complicated time-varying function of the baseline intensities in the three-state multiplicative illness-death model characterizing the true underlying process of progression and death, as is the hazard function $h_S(s)$.

For cohort studies, point process models offer a useful framework for characterizing the stochastic assessment process. We let A_k denote the random time of the k th assessment, $k = 1, 2, \dots$, and $N_a(t) = \sum_{k=1}^{\infty} I(A_k \leq t)$ count the number of follow-up assessments occurring over $(0, t]$. If $Y(s) = I(s \leq C)$ and the history of the joint process is defined as $H^\dagger(t) = \{(Z(a_k), a_k), k = 0, \dots, N_a(t^-); I(s \leq T_2), Y(s), 0 < s < t\}$ for an arbitrary individual, then their visit intensity is defined as

$$\lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_a(t + \Delta t) = 1 | H^\dagger(t))}{\Delta t} = Y(t) \lambda^\dagger(t | H^\dagger(t)). \tag{6}$$

A joint model for the response, assessment and censoring processes can be constructed by first expanding the state space. Specifically, we define a new multistate process $\{Z^\dagger(s), 0 < s\}$ with a state space $\mathcal{S}^\dagger = \{V_0, V_1, \dots, V_1^p, V_2^p, \dots, P_0, P_1, \dots, D_0, D_1, \dots, \dots\}$ as depicted in Figure 3. The subscript on the states $V_k, k = 1, 2, \dots$, reflects the number of assessment visits for which progression was not detected; for the states $V_k^p, k = 1, \dots$, the superscript p designates an assessment made post-progression. Thus, an individual is in state V_k at time t if they are alive and progression-free and $N_a(t) = k$ (i.e. $A_k \leq t < \min(T_1, T_2, A_{k+1})$). They are in state P_k at time t if they progressed after the k th assessment but have not yet experienced the $(k + 1)$ st assessment or died (i.e. $A_k < T_1 \leq t < \min(T_2, A_{k+1})$). An individual is in state D_k at time t if $I(A_k < T_2 < \min(t, A_{k+1}))$ and in state V_k^p at time t if $I(A_k < T_1 < A_{k+1} \leq t < T_2), k = 0, 1, 2, \dots$. Following the occurrence of the k th assessment, the next event to occur (visit $k + 1$, progression or death) is governed by a competing risk process; transitions to the right (i.e. $V_k \rightarrow V_{k+1}$) occur with the cause-specific intensity $\lambda^\dagger(t | H^\dagger(t)) = \alpha(t)$. Downward transitions from the V_k state (i.e. $V_k \rightarrow P_k$ and $V_k \rightarrow D_k$) correspond to transitions out of state 0 in the original three-state illness-death process and hence have the same transition intensities $\lambda_{01}(t)$ and $\lambda_{02}(t)$; likewise, the $P_k \rightarrow D_k$ transitions have intensity $\lambda_{12}(t)$ under a Markov model. Finally, the $P_k \rightarrow V_{k+1}^p$ transition corresponds to the occurrence of the $(k + 1)$ st assessment following progression and has intensity $\lambda^\dagger(t | T_1 < t, H^\dagger(t)) = \alpha_p(t)$. If we assume the assessment process is independent of the disease process, then $\alpha_p(t) = \alpha(t)$, but otherwise, we may assign a different intensity, such as $\alpha_p(t) = \alpha(t) \exp(\rho)$, to reflect increase intensity of the assessment process post-progression if, for example, $\rho > 0$. The latter may be reasonable if clinic visits are more likely if symptoms associated with progression become evident; note that such a process would violate the sequential missing at random assumption [20] and invalidate even methods based on interval-censored data [21].

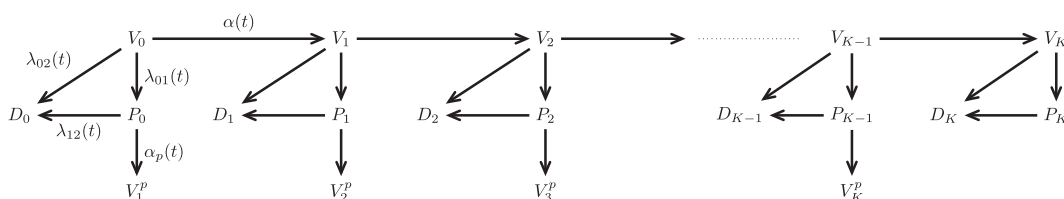


Figure 3. A multistate diagram for joint consideration of progression, death and recurrent assessment times.

We define \mathcal{V} as the set of states V_0, V_1, \dots , and define $\mathcal{V}^p, \mathcal{P}$ and \mathcal{D} similarly. Under such a setup, if K is large enough, the survival function for the surrogate progression-free survival time S is simply $P(Z^\dagger(s) \notin \mathcal{V}^p \cup \mathcal{D})$. In other words, the survival function $\mathcal{F}_S(s)$ can be expressed as a function of the transition probabilities of the multistate process given in Figure 3 such that

$$\mathcal{F}_S(s) = 1 - P(Z^\dagger(s) \in \mathcal{V}^p | Z^\dagger(0) = 0) - P(Z^\dagger(s) \in \mathcal{D} | Z^\dagger(0) = 0). \quad (7)$$

The transition probabilities necessary to compute this can be obtained by the Kolmogorov differential equation of Section 2 but can also be calculated directly from the probability of particular sample paths. For example, for an individual to be in state V_k^p at time s , $k - 1$ assessments have to occur prior to progression, and the individual must survive to the k th assessment at $A_k < s$. Over $[0, s]$, the individual thus follows the path $V_0 \rightarrow \dots \rightarrow V_{k-1} \rightarrow P_{k-1} \rightarrow V_k^p$ at times $0 = a_0 < \dots < a_{k-1} < t_1 < a_k \leq s$, so the probability $P(Z^\dagger(s) = V_k^p | Z^\dagger(0) = 0)$ takes form

$$\int_0^s \int_0^{a_k} \int_0^{t_1} \dots \int_0^{a_2} \left[\prod_{j=1}^{k-1} \alpha(a_j) \right] \lambda_{01}(t_1) \alpha_p(a_k) \exp \left[- \int_0^{t_1} (\lambda_{01}(u) + \lambda_{02}(u) + \alpha(u)) du \right] \times \exp \left[- \int_{t_1}^{a_k} (\lambda_{12}(u) + \alpha_p(u)) du \right] da_1 \dots da_{k-1} dt_1 da_k. \quad (8)$$

An expression for $P(Z^\dagger(s) = D_k | Z^\dagger(0) = 0)$ is derived similarly by noting that over $[0, s]$, an individual must follow either the path $V_0 \rightarrow \dots \rightarrow V_k \rightarrow D_k$ with transition times $0 = a_0 < \dots < a_k < t_2 \leq s$, or path $V_0 \rightarrow \dots \rightarrow V_k \rightarrow P_k \rightarrow D_k$ with transitions at times $0 = a_0 < \dots < a_k < t_1 < t_2 \leq s$. Hence, $P(Z^\dagger(s) = D_k | Z^\dagger(0) = 0)$ takes the form

$$\int_0^s \int_0^{t_2} \dots \int_0^{a_2} \left[\prod_{j=1}^k \alpha(a_j) \right] \lambda_{02}(t_2) \exp \left[- \int_0^{t_2} (\lambda_{01}(u) + \lambda_{02}(u) + \alpha(u)) du \right] da_1 \dots da_k dt_2 + \int_0^s \int_0^{t_2} \int_0^{t_1} \dots \int_0^{a_2} \left[\prod_{j=1}^k \alpha(a_j) \right] \lambda_{01}(t_1) \lambda_{12}(t_2) \exp \left[- \int_0^{t_1} (\lambda_{01}(u) + \lambda_{02}(u) + \alpha(u)) du \right] \times \exp \left[- \int_{t_1}^{t_2} (\lambda_{12}(u) + \alpha_p(u)) du \right] da_1 \dots da_k dt_1 dt_2 \quad (9)$$

When the multistate process is time homogeneous where $\lambda_{jk}(t) = \lambda_{jk}$ and the assessment process is ignorable and time homogeneous ($\alpha(t) = \alpha_p(t) = \alpha$), expressions (8) and (9) can be further simplified. Substituting (8) and (9) in (7) results in an expression for $\mathcal{F}_S(s)$, and the hazard $h_S(s)$ is no longer a simple sum of $\lambda_{01}(s)$ and $\lambda_{02}(s)$.

We compare the survival functions for progression-free survival time (4) and the surrogate progression-free survival time (7) under various parameter configurations for the setting where the assessments arise from a point process. We assume time-homogenous transition intensities $\lambda_{jk}(t) = \lambda_{jk}, j < k \in \mathcal{S}$, with values set such that (i) $P(T_{01} < T_{02}) = \lambda_{01}/(\lambda_{01} + \lambda_{02})$ equals a desired probability of progression, (ii) $\lambda_{12}/\lambda_{02} = 1.5$ corresponds to an increased risk of death following progression and (iii) $\pi_A = P(T > C_A) = p_{00}(0, C_A) = 0.20$ gives a 20% administrative censoring rate for the progression-free survival endpoint; we set $C_A = 1$ without loss of generality. The random right censoring time R_i is assumed to follow an exponential distribution with the hazard set to yield the desired net censoring rates (π_N), where $\pi_N = P(C_i < T)$ with $C_i = \min(C_A, R_i)$. We set $\alpha(t) = \alpha$ and $\alpha_p(t) = \alpha \exp(\rho)$ with $\rho = \{\log 1, \log 2\}$ for an progression independent or dependent assessment process.

Figure 4(a) contains a plot of the survival functions for the surrogate (black dashed line) and actual (black solid line) progression-free survival times for the case when $P(T_{01} < T_{02}) = 0.8$, and the assessment process is event independent with a time homogeneous rate $\alpha = 2$. The asymptotic bias in the estimator of the median progression-free survival time is evident in Figure 4(a) from the distance between the vertical lines; the median progression-free survival time is overestimated by a factor of $(0.31 - 0.12)/0.12 = 2.6$. Similarly, a progression-free survival analysis based on the surrogate time leads to an over-estimation of the probability of being event free. We also find that when all other factors are fixed, the extent of over-estimation is greater when assessments are less frequent (e.g. comparing the

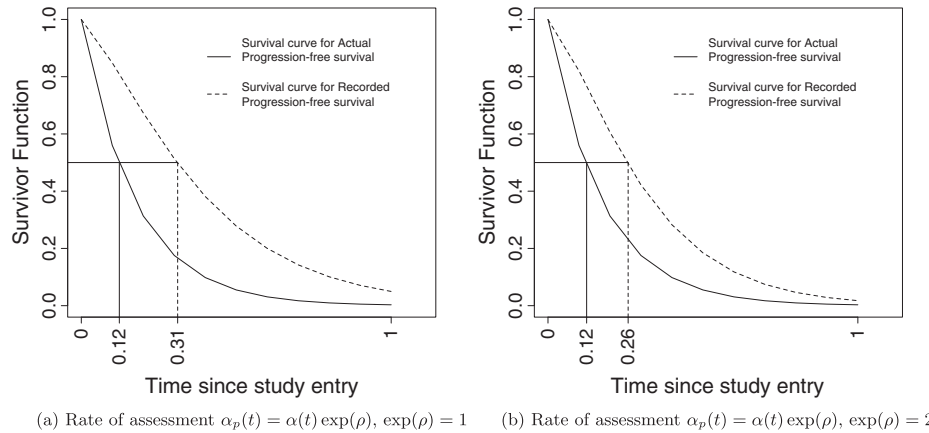


Figure 4. Survival functions for the surrogate progression-free survival time (dotted) and the actual progression-free survival time (solid) when the assessment process is (a) event independent and (b) event dependent; here, $P(T_{01} < T_{02}) = 0.8$, $\alpha(t) = \alpha = 2$, $\lambda_{12}/\lambda_{02} = 1.5$ and $\pi_A = 0.2$.

biases when $\alpha = 2$ versus when $\alpha = 4$. Finally, the asymptotic bias decreases as $P(T_{01} < T_{02})$ decreases below 0.8 because as the probability of failure due to death increases, the probability that the actual failure time will be recorded increases.

Figure 4(b) displays an analogous plot to Figure 4(a) for the case of an event-dependent assessment process with $\rho = \log 2$; here, the intensity for the next assessment doubles after progression occurs. The findings are similar, but the biases are lower because the lag from progression to its detection is stochastically smaller.

3.2. Cox regression based on the surrogate progression-free survival time

Treatment effects are naturally characterized by multiplicative intensity models of the form

$$\lambda_{jk}(t | x) = \lambda_{jk0}(t) \exp(x' \beta_{jk}), \quad (10)$$

where $\lambda_{jk0}(t)$ is the baseline rate of $j \rightarrow k$ transitions and $\beta = (\beta'_{jk}, j < k \in \mathcal{S})'$ is the vector of regression parameters. For simplicity, we focus on a binary treatment indicator x , so that the hazard ratio for the progression-free survival time is

$$\frac{h(t | x = 1)}{h(t | x = 0)} = \frac{\lambda_{010}(t) \exp(\beta_{01}) + \lambda_{020}(t) \exp(\beta_{02})}{\lambda_{010}(t) + \lambda_{020}(t)}. \quad (11)$$

The proportionality assumption holds for $h(t)$ under either (i) $\beta_{01} = \beta_{02}$ or (ii) $\lambda_{010}(t) = \lambda_{020}(t)$ [22, 23]. Under the first scenario, the treatment effect on progression-free survival is the same as the treatment effect on progression and death, $\beta = \beta_{01} = \beta_{02}$. The second scenario leads to a treatment effect on progression-free survival given by $\beta = \log[(\exp(\beta_{01}) + \exp(\beta_{02}))/2]$. When neither of these two conditions are satisfied, the hazard ratio (11) will be a complicated time-varying function of the baseline intensities and the treatment effects for progression and death.

We consider the setting when proportionality holds for the progression-free survival endpoint under condition (i) and a semi-parametric regression model for the hazard takes a form $h(t; x) = h_0(t) \exp(x\beta)$, where $h_0(t)$ is the baseline hazard, x is a binary treatment indicator and β is the true treatment effect. While this is no doubt unrealistic, we do so because our primary objective is to evaluate the impact of using the surrogate event time S in a Cox regression model under the intermittent assessment scheme. If neither condition (i) nor (ii) is satisfied, the Cox model for T is misspecified to start with, and it will be difficult to determine the role of the inspection process in the results.

To distinguish the true model from the one for the surrogate, we write

$$h_S(s; x) = h_{S0}(s) \exp(x\gamma), \quad s > 0, \quad (12)$$

for the working Cox model based on the surrogate time S with baseline hazard $h_{S0}(s)$ and treatment effect γ . Let $Y_i^\dagger(s) = I(s \leq S_i)$ be the at-risk indicator and $dN_i(s) = I(S_i = t)$ be the event indicator, which are

all defined based on the imputed event time S . If $Y_i(s) = I(s \leq C_i)$ indicates individual i is still under observation, then $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)$ indicates they are under observation and at risk of the composite event. The partial likelihood [24, 25] leads to a score function for γ of the form $U(\gamma) = \sum_{i=1}^n U_i(\gamma)$, where

$$U_i(\gamma) = \int_0^\infty \bar{Y}_i(s) \left[X_i - \frac{R^{(1)}(\gamma; s)}{R^{(0)}(\gamma; s)} \right] dN_i(s), \quad (13)$$

$R^{(1)}(\gamma, s) = \sum_{i=1}^n \bar{Y}_i(s)X_i \exp(X_i\gamma)$ and $R^{(0)}(\gamma, s) = \sum_{i=1}^n \bar{Y}_i(s) \exp(X_i\gamma)$. If $\hat{\gamma}$ denotes the estimate of γ by solving $U(\gamma) = 0$, then $\hat{\gamma} \rightarrow \gamma^*$ in probability, where γ^* is the solution to

$$\int_0^\infty \left\{ E[\bar{Y}_i(s)X_i dN_i(s)] - \frac{r^{(1)}(\gamma, s)}{r^{(0)}(\gamma, s)} E[\bar{Y}_i(s) dN_i(s)] \right\} = 0, \quad (14)$$

with $r^{(0)}(\gamma, s) = E\{R^{(0)}(\gamma, s)\}$ and $r^{(1)}(\gamma, s) = E\{R^{(1)}(\gamma, s)\}$, respectively [16]. Lin and Wei [17] prove that $\sqrt{n}(\hat{\gamma} - \gamma^*)$ is asymptotically normal with

$$\text{var} \left(\sqrt{n}(\hat{\gamma} - \gamma^*) \right) = [A^{-1}(\gamma^*)][B(\gamma^*)][A^{-1}(\gamma^*)]', \quad (15)$$

where

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

$$B(\gamma) = \int_0^\infty E \left\{ \bar{Y}_i(s) \left[X_i - \frac{r^{(1)}(\gamma, s)}{r^{(0)}(\gamma, s)} \right]^2 dN_i(s) \right\}.$$

We define $\gamma^* - \beta$ as the asymptotic bias.

The expectations in (14) and (15) are taken with respect to the distribution of the surrogate time S with expressions $E\{\bar{Y}_i(s)X_i dN_i(t)\} = \mathcal{G}(s)f_S(s; x = 1)P(X = 1)$, $E\{\bar{Y}_i(s) dN_i(t)\} = \sum_{x=0}^1 \mathcal{G}(s)f_S(s; x)P(X = x)$, $r^{(1)}(\gamma, s) = n\mathcal{G}(s)\mathcal{F}_S(s; x = 1) \exp(\gamma)P(X = 1)$ and $r^{(0)}(\gamma, s) = n \sum_{x=0}^1 \mathcal{G}(s)\mathcal{F}_S(s; x) \exp(x\gamma)P(X = x)$, where $\mathcal{G}(s) = P(R_i \geq s)$ denotes the survival function for the independent random right censoring time R_i . Note that $f_S(s; x)$ and $\mathcal{F}_S(s; x)$ are the density and survival functions for the surrogate progression-free survival time S derived in Section 3.1, and they are functions of baseline intensities and treatment effects in the true three-state illness-death model as well as the parameters governing the assessment process. These enable one to evaluate how the bias in the estimator from the imputed data analysis changes according to these factors. In the Appendix, we report the results of a small investigation regarding the asymptotic behaviour of Cox regression coefficients of the treatment effect when the surrogate progression-free survival data are used. For the scenarios studied, we find that there is generally a conservative bias, meaning the naive use of the surrogate progression-free survival time underestimates the impact of treatment.

4. Simulations and analyses based on correct and misspecified models

Here, we report on simulation studies designed to empirically assess the bias and frequency properties of tests and confidence intervals in terms of power and coverage probabilities. Individuals were randomized to one of two treatment arms in a balanced fashion with equal probability. Given the treatment assignment, the times to progression or/and death were simulated based on the three-state multiplicative model; T_{01} and T_{02} were simulated based on exponential distributions with rates $\lambda_{01} \exp(x\beta)$ and $\lambda_{02} \exp(x\beta)$, respectively, and if $T_{01} < T_{02}$, we simulate T_{12} according to an exponential distribution with rate $\lambda_{12} \exp(x\beta_{12})$. The values for the baseline intensities are set to satisfy the following constraints: (i) $P(T_{01} < T_{02} | X = 0) = \lambda_{01}/(\lambda_{01} + \lambda_{02}) = \{0.6, 0.8\}$; (ii) $\lambda_{12}/\lambda_{02} = 1.5$; and (iii) $\pi_A = P(T > C_A | X = 0) = 0.20$. We set $\beta_{01} = \beta_{02} = \beta = \log 0.4$ and $\beta_{12} = \log 1$. The random right censoring time R_i is simulated by an exponential distribution with the rate set to achieve $\pi_N = \{0.4, 0.6\}$.

For the case with fixed assessment times, we assume $K = 4$ assessments occur over $[0, C_A]$, where $C_A = 1$, at times $a_k = kC_A/K + \epsilon_k$, where $\epsilon_k \sim N(0, \sigma_e)$ and $k = 1, \dots, K$. The random error ϵ_k is added to mimic the situation where some variation occurs around the evenly spaced assessment times in a protocol. We let $\sigma_e = \{C_A/(6K), C_A/(20K)\}$ for moderate and mild variation so that the chance

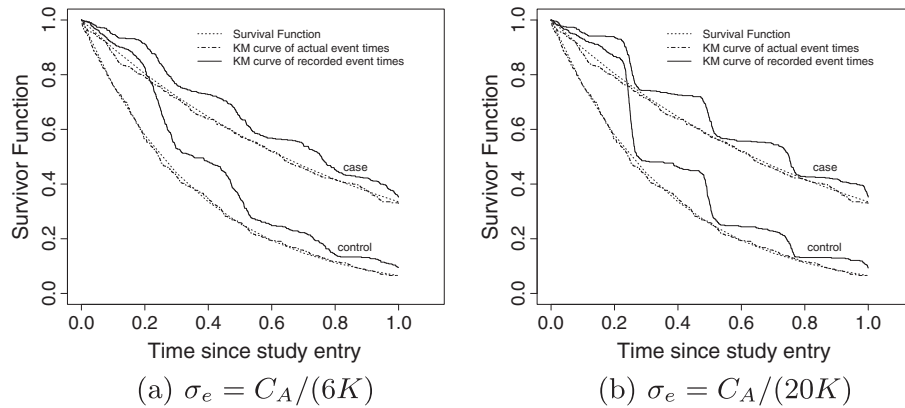


Figure 5. Kaplan–Meier curves estimated using right end point imputed progression-free survival event times and actual progression-free survival event times, contrasted against the theoretical survival function of progression-free survival, for treatment and control groups. Number of fixed assessments $K = 4$ with a random normal noise $\epsilon_k \sim N(0, \sigma_e)$. ($\beta_{01} = \beta_{02} = \beta = \log 0.5$, $\beta_{12} = 0$, $P(T_{01} < T_{02} | X = 0) = 0.80$, $\lambda_{12}/\lambda_{02} = 1.5$, $\pi_A = 0.2$, $\pi_N = 0.4$).

of two consecutive assessment times would be realized in the reverse order is small. We plot (4) along with the KM estimates based on the actual progression-free survival time (dashed line) and the surrogate progression-free survival time (solid line) for a data set of $n = 1000$ individuals. The plots reveal a striking, familiar but misleading stepwise drop in the KM curves based on the surrogate event times as seen in [19], for example. This pattern becomes more distinctive as the variation around pre-scheduled assessment times decreases (as shown in Figure 5(b)) and the risk of progression increases. In addition, the KM curves under right-endpoint imputation sit above KM curves based on actual progression-free survival event times and the theoretical curves.

We also simulated random assessment times based on a Poisson process where the gap time between two consecutive assessments follows an exponential distribution with a rate α . For each parameter configuration, we calculated the sample size required to achieve 80% power for a Wald test at the 5% significance level under the Cox model. Three types of analysis were conducted on each simulated dataset for a given sample size including a Cox model for the true progression-free survival time, a Cox model for the surrogate progression-free survival time and a correct analysis based on the actual panel observation scheme using a time-homogenous illness-death model of Section 2. The first and third analyses yield consistent estimates of the treatment effect, but the first, of course, is not possible in practice. Under the assumption that the inspection process satisfies the sequentially missing at random condition [20] and is non-informative, the partial likelihood has the form

$$\left\{ \prod_{k=1}^{K_i} \Pr [Z(a_{ik}) | Z(a_{i,k-1}), X_i] \right\} \left\{ \prod_{j=0}^1 \Pr [Z(t_{i2}^-) = j | Z(a_{iK_i}), X_i] \lambda_{j2}(t_{i2} | X_i) \right\}^{I(T_{i2} < C_i)}$$

This analysis can be implemented with the MSM package [26].

Table I reports the difference between the average estimated treatment effects and the true β (EBIAS) based on 1000 replications for each parameter configuration, the empirical standard error of the estimates (ESE), the average robust standard error (RSE) based on (15) and the empirical coverage probability of nominally 95% CIs for β (ECP). The proportion of replicates in which a test of $H_0 : \beta = 0$ is rejected under a two-sided Wald test is also reported as the empirical power (EP). The Cox regression analysis of the actual event time and the multistate analysis yielded estimates with negligible empirical biases, but the naive Cox analysis based on the imputed event times gave biased estimates of treatment effect. The empirical coverage probabilities from all three analysis are very close to the nominal level. Although the multistate analysis based on panel data and the Cox model based on the true event time both yield consistent estimates, the former is associated with slightly larger standard errors and lower EP, reflecting the loss of information when event times are interval censored. More appreciable differences in EP are seen when comparing the naive Cox regression analysis with the parametric multistate analysis. The EP from the naive Cox regression analysis are below 80% in general and can be as low as 62.7%.

Table 1. Assessments arise from point process; number of simulations = 1000 for 80% statistical power.

α	P_1	N	Cox model														
			True PFS time [†]					Surrogate PFS time [†]					TH multistate analysis				
			EBIAS	ESE	RSE	ECP%	EP%	EBIAS	ESE	RSE	ECP%	EP%	EBIAS	ESE	RSE	ECP%	EP%
$\beta = \log(0.50), \pi_N = 40\%$																	
2.00	0.60	111	-0.003	0.263	0.253	94.0	78.5	0.059	0.284	0.277	93.9	62.7	-0.006	0.271	0.276	95.4	73.3
2.00	0.80	111	-0.007	0.257	0.252	94.7	80.9	0.078	0.272	0.268	93.0	64.6	-0.006	0.269	0.278	95.9	72.4
4.00	0.60	111	-0.003	0.263	0.253	94.0	78.5	0.028	0.284	0.269	94.4	69.4	-0.008	0.252	0.266	95.9	78.5
4.00	0.80	111	-0.007	0.257	0.252	94.7	80.9	0.037	0.287	0.277	93.6	67.1	-0.008	0.259	0.257	95.6	78.4
8.00	0.60	111	-0.003	0.263	0.253	94.0	78.5	0.008	0.280	0.263	94.2	73.6	-0.005	0.254	0.252	95.3	80.4
8.00	0.80	111	-0.007	0.257	0.252	94.7	80.9	0.006	0.280	0.267	94.2	74.3	-0.003	0.254	0.253	95.3	80.2
$\beta = \log(0.50), \pi_N = 60\%$																	
2.00	0.60	167	-0.004	0.253	0.253	95.1	80.0	0.039	0.272	0.265	93.2	69.8	-0.005	0.267	0.276	95.5	73.6
2.00	0.80	167	0.002	0.246	0.254	96.3	79.1	0.065	0.268	0.272	94.9	64.3	0.004	0.276	0.277	95.4	72.0
4.00	0.60	167	-0.004	0.253	0.253	95.1	80.0	0.014	0.269	0.267	94.9	73.2	-0.004	0.252	0.026	96.1	78.3
4.00	0.80	167	0.002	0.246	0.254	96.3	79.1	0.034	0.271	0.274	94.9	69.2	0.005	0.259	0.258	94.6	77.1
8.00	0.60	167	-0.004	0.253	0.253	95.1	80.0	0.003	0.269	0.266	94.3	74.3	0.009	0.249	0.253	95.5	79.4
8.00	0.80	167	0.002	0.246	0.254	96.3	79.1	0.018	0.265	0.271	95.0	72.2	-0.005	0.251	0.254	95.8	78.7
$\beta = \log(0.75), \pi_N = 40\%$																	
2.00	0.60	634	-0.002	0.104	0.103	94.1	82.0	0.023	0.111	0.114	94.9	65.4	-0.003	0.114	0.113	95.3	73.4
2.00	0.80	634	0.003	0.100	0.103	95.9	79.5	0.036	0.110	0.110	93.4	61.9	-0.006	0.112	0.113	94.1	71.7
4.00	0.60	634	-0.002	0.104	0.103	94.1	82.0	0.012	0.109	0.110	95.1	73.3	-0.002	0.102	0.105	95.8	78.4
4.00	0.80	634	0.003	0.100	0.103	95.9	79.5	0.021	0.111	0.113	95.0	66.6	-0.001	0.107	0.105	94.3	78.1
8.00	0.60	634	-0.002	0.104	0.103	94.1	82.0	0.001	0.107	0.107	94.7	76.4	-0.003	0.104	0.103	93.9	80.6
8.00	0.80	634	0.003	0.100	0.103	95.9	79.5	0.008	0.106	0.109	95.2	72.9	-0.003	0.105	0.104	94.8	78.8
$\beta = \log(0.75), \pi_N = 60\%$																	
2.00	0.60	952	-0.001	0.105	0.103	93.5	80.6	0.017	0.110	0.108	94.6	70.3	-0.006	0.111	0.112	94.7	76.0
2.00	0.80	952	-0.003	0.104	0.103	94.9	81.0	0.023	0.109	0.111	94.7	67.4	-0.004	0.114	0.112	94.9	75.1
4.00	0.60	952	-0.001	0.105	0.103	93.5	80.6	0.011	0.109	0.109	95.1	72.5	-0.003	0.103	0.105	95.5	79.3
4.00	0.80	952	-0.003	0.104	0.103	94.9	81.0	0.011	0.111	0.112	94.9	68.9	-0.004	0.105	0.108	94.3	79.9
8.00	0.60	952	-0.001	0.105	0.103	93.5	80.6	0.004	0.109	0.108	94.3	73.9	-0.000	0.103	0.103	94.8	79.6
8.00	0.80	952	-0.003	0.104	0.103	94.9	81.0	-0.001	0.108	0.110	94.5	74.7	-0.005	0.106	0.103	94.8	78.7

[†]Both the true and surrogate progression-free survival (PFS) times are subject to right censoring. $\beta_{12} = \log 1, P_1 = P(T_{01} < T_{02} | X = 0) = \lambda_{01}/(\lambda_{01} + \lambda_{02}), \lambda_{12}/\lambda_{02} = 1.5, C_A = 1, \pi_A = 0.20$.

The impact of the misspecification due to right-endpoint imputation can therefore be quite substantial. A selection of results in Table I are graphically displayed in Figure A.1 in the Appendix, demonstrating good agreement between the empirical finite sample behaviour and results anticipated based on large sample theory.

5. Discussion

There has been much discussion about the use of composite endpoints in clinical trials [22, 23], and while views are somewhat divided, it is apparent that a clear interpretation of findings based on composite endpoints is difficult given the many factors influencing estimators. More discussion on use of progression-free survival as a primary endpoint in oncology can be found in recent papers by [2, 3, 27, 28].

We have restricted attention to a idealized setting in which the treatment effect is the same for progression and death in order to simplify discussion and focus attention on the effects of intermittent assessments.

Frydman and Szarek [9] discuss the issue of interval-censored progression times in the context of nonparametric estimation of progression-free survival distribution. Boruvka and Cook [29] consider estimation and inference with the Cox model when two competing events are subject to different censoring schemes and interest lies in the transition intensity ratios. They point out that the present problem can be cast in this dual censoring framework, and hence, their methods could be utilized in this context. We have focused here, however, on the study of currently used methods to alert researchers to their limitations. Of particular importance is our finding that the standard practice of using the surrogate progression-free survival time can lead to a substantial drop in the power of the trial when assessments are infrequent. To ensure trials are adequately powered, sample sizes should be increased to address the loss of power associated with the use of the surrogate event time S . One approach to this is to derive the sample size formula based on the three-state illness-death model addressing the interval-censored progression times; piecewise constant baseline hazard function can be adopted to ease the calculation. Alternatively, one can simply adjust the sample size based on the misspecified model by deriving the limiting behaviour of the naive estimator and accommodating the bias and robust large sample variance in the calculations to ensure the power is maintained at the nominal level despite the misspecification.

These issues, while motivated by the setting of cancer clinical trials, arise in many other areas of health research involving joint consideration of non-fatal events and death. Examples include osteoporosis trials in which elderly patients are at risk of both asymptomatic fractures (e.g. vertebral compression) and death where the former events are only detected by periodic bone scans, studies of cognitive impairment where periodic assessments may be scheduled and carried out until death or administrative censoring, among others.

Appendix A: Asymptotic properties of treatment effect estimates from Cox regression on surrogate PFS

We set $P(X = 1) = P(X = 0) = 0.5$ and take the transition intensities in the illness-death model to have a multiplicative form $\lambda_{jk}(t) = \lambda_{jk} \exp(x\beta)$. As before, values for the baseline intensities λ_{jk} are based on specifications for $P(T_{01} < T_{02} | X = 0) = \lambda_{01}/(\lambda_{01} + \lambda_{02})$, $\lambda_{12}/\lambda_{02}$ and π_A ; see Section 3.1. We set $\beta_{01} = \beta_{02} = \beta \in \{\log 0.9, \log 0.75, \log 0.5\}$ for mild, moderate and strong (beneficial) treatment effects. We let $\beta_{12} = \log 1$ for no treatment effect on the intensity for progression to death transition.

Figure A.1 displays percent relative asymptotic bias $(100(\gamma^* - \beta)/\beta)$, asymptotic power associated with a Wald test of no treatment effect and the asymptotic coverage probability of a 95% CI when fitting a Cox model based on the surrogate progression-free survival time under an ignorable random assessment process with intensity $\alpha(t) = \alpha$. The treatment effect is under-estimated in general with the percent relative biases ranging from about 3% to 15%; see Figure 5. As expected, the magnitude of the bias increases as the rate of assessments decreases and as the competing risk of progression over death, $P(T_{01} < T_{02} | X = 0)$, increases. The size of the treatment effect itself seems to have very little impact on the magnitude of the percent relative bias. For each parameter configuration, we derived the sample size required to achieve an 80% power for a two-sided Wald test of treatment effect at the 5% significance level under the Cox model for the progression-free survival analysis. We then calculated the asymptotic power that one can actually obtain when the surrogate time is used for analysis. Figure A.1(c, d) shows that using the surrogate time can lead to an appreciable loss in power, which increases as the assessment rate decreases and the probability of progression increases. For example, when $\alpha = 2$ and $P(T_{01} < T_{02} | X = 0) = 0.8$, the

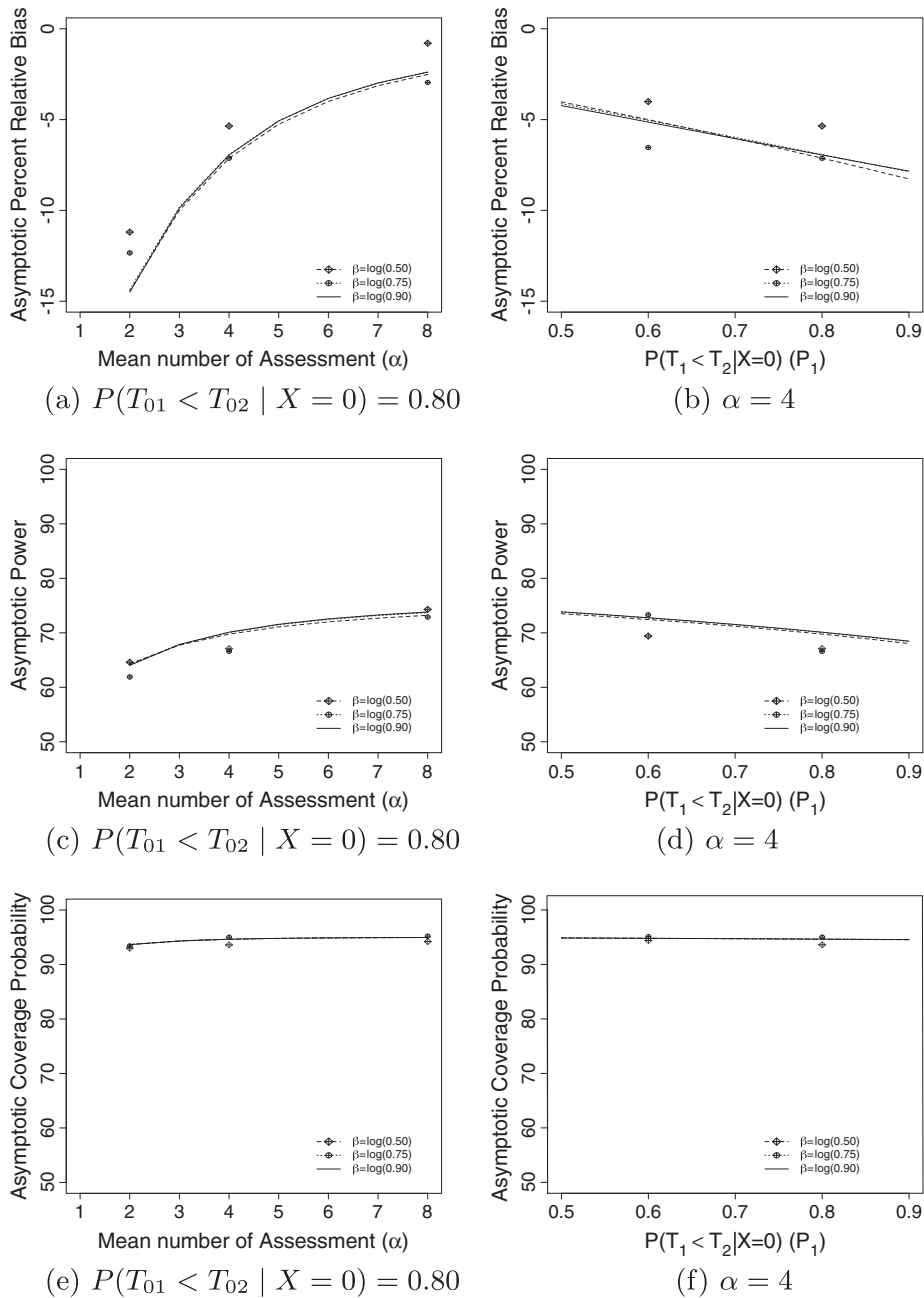


Figure A.1. Asymptotic percent relative bias ($100(\gamma^* - \beta)/\beta$), asymptotic power and asymptotic coverage probability of Cox regression coefficient of treatment effect from progression-free survival analysis using imputed data; here, $\beta_{01} = \beta_{02} = \beta = \{\log 0.5, \log 0.75, \log 0.9\}$, $\beta_{12} = \log 1$, $\lambda_{12}/\lambda_{02} = 1.5$, $\pi_A = 0.2$ and $\pi_N = 0.4$.

power from the naive progression-free survival analysis can drop to as low as 63% when the nominal level is 80%. The asymptotic coverage probability of the 95% CI, on the other hand, is quite robust to the model misspecification. Analogous results are found when the assessment times are fixed and, hence, are not reported here.

Acknowledgements

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to L. Zeng (RGPIN 327107) and R. J. Cook (RGPIN 155849) and the Canadian Institutes for Health Research to R. J. Cook (FRN 13887). L. Zeng is a Graham Trust Chair in Health Statistics, and R. J. Cook is a Canada Research Chair in Statistical Methods for Health Research.

References

1. Saaed ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. *Journal of Clinical Oncology* 2010; **28**(11):1958–1962.
2. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *Journal of Clinical Oncology* 2012; **30**(10):1030–1033.
3. Fleming TR, Rothmann MD, Lu H L. Issues in using progression-free survival when evaluating oncology products. *Journal of Clinical Oncology* 2009; **27**(17):2874–2880.
4. Miksad RA, Zietemann V, Gothe R, Schwarzer R, Conrads-Frank A, Schnell-Inderst P, Stollenwerk B, Siebert U. Progression-free survival as a surrogate endpoint in advanced breast cancer. *International Journal of Technology Assessment in Health Care* 2009; **24**(4):371–383.
5. Sidhu R, Rong A, Dahlberg S. Evaluation of progression-free survival as a surrogate endpoint for survival in chemotherapy and targeted agent metastatic colorectal cancer trials. *Clinical Cancer Research* 2013; **19**(5):969–976.
6. Ballman KV, Buckner JC, Brown PD, Giannini C, Flynn PJ, LaPlant BR, Jaeckle KA. 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* 2007; **9**(1):29–38.
7. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute* 2009; **101**(23):1642–1649.
8. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Annals of Oncology* 2009; **20**(3):460–464.
9. Frydman H, Szarek M. Nonparametric estimation in a Markov “illness-death” process from interval censored observations with missing intermediate transition status. *Biometrics* 2009; **65**(1):143–151.
10. FDA. *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, 2007. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM259421.pdf> [Accessed on 5 May 2015].
11. Dejardin D, Lesaffre E, Verbeke G. Joint modeling of progression-free survival and death in advanced cancer clinical trials. *Statistics in Medicine* 2010; **29**:1724–1734.
12. Leffondré K, Touraine C, Helmer C, Joly P. Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? *International Journal of Epidemiology* 2013; **42**(4):1177–1186.
13. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute* 2007; **99**(6):428–432.
14. Binder N, Schumacher M. Missing information caused by death leads to bias in relative risk estimates. *Journal of Clinical Epidemiology* 2014; **67**(10):1111–1120.
15. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; **50**(1):1–26.
16. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika* 1986; **73**(1):363–369.
17. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 1989; **84**(408):1074–1078.
18. Cox DR, Miller HD. *The Theory of Stochastic Processes* First edn. Wiley: New York, 1965.
19. Scher HI, Fizazi K, Saad F, Taplin M, Sternberg CN, Miller K, Wit R, Mulders P, Chi KN, Shore ND, Armstrong AJ, Flaig TW, Flechon A, Mainwaring P, Fleming M, Hainsworth JD, Hirmand M, Selby B, Seely L, Bono JS for the AFFIRM Investigators. Increased survival with enzalutamide in prostate cancer after chemotherapy. *The New England Journal of Medicine* 2012; **367**(13):1187–1197.
20. Hogan JW, Roy J, Korkontzelou C. Handling dropouts in longitudinal studies. *Statistics in Medicine* 2004; **23**(9):1455–1497.
21. Grüger J, Kay R, Schumacher M. The validity of inferences based on incomplete observations in disease state models. *Biometrics* 1991; **47**(2):595–605.
22. Wu L, Cook RJ. Misspecification of Cox regression models with composite endpoints. *Statistics in Medicine* 2012; **31**(28):3545–3562.
23. Cook R J, Lee K-A. Statistical considerations in the use of composite endpoints in clinical trials. In *Developments in statistical evaluation in clinical trials*, van Montfort K, Oud J, Ghidry W (eds). Springer Science + Business Media: New York, NY, 2014.
24. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data* Second, Wiley Series in Probability and Statistics. Wiley- Interscience: A John Wiley & Sons, Inc: Hoboken, New Jersey, 2002.
25. Lawless JF. *Statistical models and methods for lifetime data* Second, Wiley Series in Probability and Statistics. Wiley-Interscience: A John Wiley & Sons, Inc: Hoboken, New Jersey, 2003.
26. Jackson CH. Multi-state models for panel data: the MSM package for R. *Journal of Statistical Software* 2011; **38**(8):1–28.
27. Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL, Elfring GL, Pignon JP, Peidbois P. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *Journal of Clinical Oncology* 2007; **25**(33):5218–5224.
28. Soria JC, Massard C, Chevalier TL. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? *Annals of Oncology* 2010; **21**(12):2324–2332.
29. Boruvka A, Cook RJ. Sieve estimation in a Markov illness-death process under dual censoring. Resubmitted, *Biostatistics* 2015.