

# The Design of Intervention Trials Involving Recurrent and Terminal Events

LONGYANG WU

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

RICHARD J. COOK

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

## Summary

Clinical trials are often designed to assess the effect of therapeutic interventions on the incidence of recurrent events in the presence of a dependent terminal event such as death. Statistical methods based on multistate analysis have considerable appeal in this setting since they can incorporate changes in risk with each event occurrence, a dependence between the recurrent event and the terminal event, and event-dependent censoring. To date, however, there has been limited development of statistical methods for the design of trials involving recurrent and terminal events. Based on the asymptotic distribution of regression coefficients from a multiplicative intensity Markov regression model, we derive sample size formulas to address power requirements for both the recurrent and terminal event processes. We consider the design of trials for which separate marginal hypothesis tests are of interest for the recurrent and terminal event processes and deal with both superiority and non-inferiority tests. Simulation studies confirm that the designs satisfy the nominal power requirements in both settings, and an application to a trial evaluating the effect of a bisphosphonate on skeletal complications is given for illustration.

*Keywords:* Multistate model, non-inferiority, recurrent events, sample size, superiority

**The final publication is available at Springer via <http://dx.doi.org/10.1007/s12561-013-9083-z>.**

## 1 INTRODUCTION

### 1.1 BACKGROUND

Clinical trials must be designed with appropriate power to address scientific needs, ethical demands, and financial restrictions. In parallel group randomized trials involving failure time outcomes, power objectives are typically met for a given model (e.g. Cox model) by specifying the event rate in the reference arm, the clinically important effect, the censoring rate and the size of the test, and then by deriving a suitable sample size based on large sample theory [1]. Under this general framework, a number of authors have developed methods for planning trials based on analyses of the time to the first event [2–5].

Sample size formulas have been developed [6] for recurrent event outcomes based on mixed Poisson models with multiplicative rate functions [7, 8]. Power and sample size considerations were subsequently developed for more general multiplicative models [9] using counting process theory [10]. Another approach to the analysis of recurrent event data in clinical trials is to use the robust methods for the analysis of multivariate survival data [11] under a working independence assumption and sample size formula for this approach are available [12]. More recently there has been interest in trial design based on covariate-adjusted log-rank statistics for recurrent event analyses and associated sample size formula have been developed [13].

To date no methods have been developed for the design of clinical trials in which the aim was to test treatment effects on recurrent and terminal event processes. We address this problem under the framework of a Markov model with transient states corresponding to the recurrent events and a single absorbing state for death. The treatment effect on the recurrent events is formulated by specifying multiplicative intensity models with time-dependent strata based on the cumulative event history and a common treatment effect; this formulation is in the spirit of the Prentice et al. [14] approach to the analysis of recurrent events. Multiplicative intensity-based models are also incorporated for mortality with the same stratification criteria. Under this formulation we derive the limiting value of partial score statistics for the treatment effect on the recurrent and terminal event processes, along with the asymptotic variances under the null and alternative hypotheses. Sample size criteria are then obtained to satisfy power objectives when separate marginal hypothesis tests are of interest for the two types of events.

Non-inferiority designs are being used increasingly often in cancer and cardiovascular research [15, 16] since many treatments with proven efficacy are available and placebo-controlled trials are therefore unethical. In such settings new interventions are required to have some advantages over standard care, such as a lower cost, a lower rate of adverse events, or a less invasive mode of administration [17]. Rothmann et al. [15] provides an excellent discussion about the various approaches to hypothesis testing in the context of non-inferiority oncology trials and extensions have recently been made for recurrent event analyses based on mixed Poisson models or robust marginal methods [18]. We consider design issues when there are superiority or non-inferiority hypotheses for the recurrent event and survival processes in the context of the multistate model.

The remainder of this paper is as follows. In the next subsection we give further background information on the setting of the palliative trials for patients with cancer metastatic to bone. In Section 2 we define notation, describe the multistate model, and derive the relevant partial score statistics. The limiting distribution of the partial score statistics are derived in Section 3 which facilitate sample size calculation in Section 4. Simulation studies in Section 5 confirm that the empirical frequency properties are compatible with the nominal levels under the null and that power requirements are met. An application is given in Section 6 and general remarks and topics for further research are discussed in Section 7.

## 1.2 TRIAL DESIGN FOR PATIENTS WITH SKELETAL METASTASES

Cancer patients with skeletal metastases are at increased risk of a variety of clinical events including pathological and nonpathological fractures, bouts of acute bone pain, and episodes of hypercalcemia. These events are typically grouped together to form a composite recurrent “skeletal related event” which is used as a basis for the evaluation of treatments designed to reduce the occurrence of skeletal complications in cancer patients to help maintain functional ability and quality of life and minimize health service utilization [19]. Because the patient population has metastatic cancer, they are also at considerable risk of death. In breast cancer, twelve month survival in recent studies has been approximately 78.9% in treated patients; in lung, prostate and other solid tumors the 12 month survival rates were 28.0%, 66.0% and 33.6% respectively.

While bisphosphonate therapy is palliative and not expected to impact survival, an assessment of the effect on survival times is warranted for a complete evaluation of the consequences of treatment. Simultaneous consideration of treatment effects on the recurrent skeletal related events and survival is therefore essential and analyses must accommodate a possible association between the recurrent event and terminal death process.

## 2 LIKELIHOOD FOR RECURRENT AND TERMINAL EVENTS

We adopt the framework of a continuous time multistate Markov process to jointly model the recurrent events and terminal event. Let  $\{Z_i(s), 0 < s\}$  denote this process for individual  $i$  with a countable number of states in the state space  $\mathcal{S} = \{0, 1, \dots, D\}$  and a right continuous sample path. The integers  $0, 1, 2, \dots$  represent the number of recurrent events experienced over time and  $D$  represents an absorbing death state. Figure 1 displays a multi-state diagram for the recurrent events and terminal event process. If individual  $i$  is alive at time  $t$  and has experienced precisely  $j$  events over  $(0, t]$ , then  $Z_i(t) = j$  and if individual  $i$  dies at time  $s$ ,  $Z_i(t) = D$  for  $t \geq s$ . We assume that all subjects are at state 0 at time  $t = 0$ , the time of randomization. Let  $v_i$  be a binary treatment indicator for individual  $i$  such that  $v_i = 1$  if individual  $i$  was randomized to the experimental treatment and  $v_i = 0$  otherwise.

Let  $T_{ij}$  be the time individual  $i$  enters state  $j$ ,  $j = 1, \dots$ , and  $T_i^d$  their time of death,  $i = 1, \dots, m$ . Let  $dN_{ij}(t) = I(Z_i(t^-) = j - 1, Z_i(t) = j)$ , indicate that a  $(j - 1) \rightarrow j$  transition was made at time  $t$  for individual  $i$ , so  $dN_{ij}(t) = 1$  at  $t_{ij}$  but is zero otherwise,  $j = 1, \dots$ . Let  $dN_{ij}^d(t) = I(Z_i(t^-) = j - 1, Z_i(t) = D)$  indicate that a  $(j - 1) \rightarrow D$  transition is made at time  $t$  (i.e. that the  $j$ th event was death). Let  $N_i(t) = (N_{ij}(t), j = 1, \dots)$  and  $N_i^d(t) = (N_{ij}^d(t), j = 1, \dots)$  jointly be the multivariate counting process for individual  $i$ . The history of the process is the information observed up to  $t^-$  and we let  $H_i(t) = \{N_i(s), N_i^d(s), 0 \leq s < t, v_i\}$  denote the history for individual  $i$ ,  $i = 1, \dots, m$ . A stochastic model for this multistate process must be assumed to derive sample size calculations. We formulate this model by specifying the respective intensity functions [21]. The intensities for event occurrence or death are defined as

$$\lambda_j(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}(t) = 1 | Z_i(t^-) = j - 1, H_i(t))}{\Delta t}$$

and

$$\gamma_j(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}^d(t) = 1 | Z_i(t^-) = j - 1, H_i(t))}{\Delta t},$$

respectively, where  $\Delta N_{ij}(t) = N_{ij}((t + \Delta t)^-) - N_{ij}(t^-)$  and  $\Delta N_{ij}^d(t) = N_{ij}^d((t + \Delta t)^-) - N_{ij}^d(t^-)$  count the number of the  $(j - 1) \rightarrow j$  and  $(j - 1) \rightarrow D$  transitions over  $(t, t + \Delta t)$  respectively.

Consider a study with planned follow-up over the interval  $(0, \tau]$ , where  $\tau$  is called the administrative censoring time. Individuals may withdraw prematurely from a study and so we let  $\tau_i^\dagger$  be the random right censoring time and let  $\tau_i = \min(\tau_i^\dagger, \tau)$  be the net censoring time for individual  $i$ ; we let  $X_i = \min(T_i^d, \tau_i)$  denote the total time on study and  $\delta_i = I(X_i = T_i^d)$  indicate whether the terminal event was observed. Let  $Y_i(t) = I(t \leq \tau_i)$  indicate whether individual  $i$  is under observation at  $t$  and  $Y_{ij}(t) = I(Z_i(t^-) = j - 1)$ ,  $j = 1, \dots$  indicate that individual  $i$  is at risk of a transition out of state  $j - 1$  at time  $t$  (i.e. they are at risk for the  $j$ th event of either type), so  $\bar{Y}_{ij}(t) = Y_i(t)Y_{ij}(t)$  indicates they are both at risk *and* under observation. Then  $d\bar{N}_{ij}(t) = \bar{Y}_{ij}(t)dN_{ij}(t)$  and  $d\bar{N}_{ij}^d(t) = \bar{Y}_{ij}(t)dN_{ij}^d(t)$  are so-called the observable counting processes for the recurrent event and terminal events respectively. The observed data can then be written  $\{d\bar{N}_i(s), d\bar{N}_i^d(s), Y_i(s), 0 < s, v_i\}$ ,  $i = 1, \dots, m$ . The history of the observable process is the information observed up to  $t^-$  and denoted  $\bar{H}_i(t) = \{\bar{N}_i(s), \bar{N}_i^d(s), \bar{Y}_i(s), 0 \leq s < t, v_i\}$ ,  $i = 1, \dots, m$ .

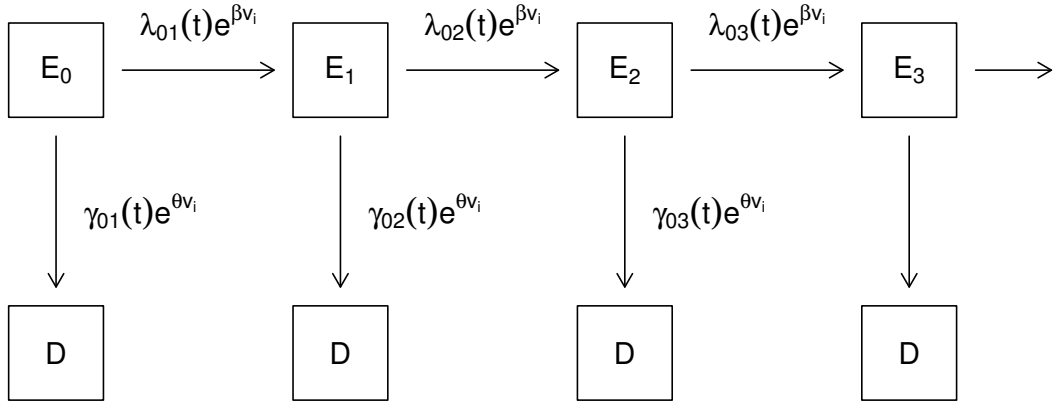


Figure 1: Recurrent events with terminal event diagram representing the model formation based on counting processes.  $\lambda_{0j}(t) \exp(\beta v)$ ,  $j = 1, 2, \dots$ , are the transition intensities for the recurrent events from state  $(j - 1)$  to state  $j$ . State D represents the terminal event of death and  $\gamma_{0j}(t) \exp(\theta v)$ ,  $j = 1, 2, \dots$ , are the corresponding event-dependent transition intensities; for convenience state  $E_j$  is simply referred to as state  $j$

Under conditionally independent censoring [22], the intensities for event occurrence and death of the observable processes are given by  $\bar{\lambda}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\lambda_j(t|H_i(t))$  and  $\bar{\gamma}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\gamma_j(t|H_i(t))$ , respectively. Thus if individual  $i$  experienced  $J_i > 0$  recurrent events at times  $t_{i1}, \dots, t_{iJ_i}$  over  $[0, X_i]$ , their likelihood contribution is proportional to

$$\prod_{j=1}^{J_i} \bar{\lambda}_j(t_{ij}|\bar{H}_i(t_{ij}))[\bar{\gamma}_{J_i}(X_i|\bar{H}_i(X_i))]^{\delta_i} \exp\left(-\sum_{j=1}^{J_i+1} \int_{t_{i,j-1}}^{t_{ij}} (\bar{\lambda}_j(u|\bar{H}_i(u)) + \bar{\gamma}_j(u|\bar{H}_i(u)))du\right),$$

where  $t_{i0} = 0$  and for notational convenience we let  $t_{i,J_i+1} = X_i$ .

A specification is required for the intensity functions and here we adopt a multiplicative intensity Markov model [1] and we set the two intensities to

$$\bar{\lambda}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\lambda_{ij}(t) = \bar{Y}_{ij}(t)\lambda_{0j}(t) \exp(\beta v_i), \quad (1)$$

and

$$\bar{\gamma}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\gamma_{ij}(t) = \bar{Y}_{ij}(t)\gamma_{0j}(t) \exp(\theta v_i), \quad (2)$$

where  $\lambda_{0j}(t)$  and  $\gamma_{0j}(t)$  are non-negative baseline intensity functions for the recurrent event and terminal event for state  $j$ , respectively. Through the time-dependent stratification on the cumulative number of events, this model accommodates an association between the recurrent and terminal events. The multiplicative effect of  $v_i$  is assumed to be constant (i.e. not event dependent) for the two processes to give a parsimonious parameterization of the treatment effect. This model was discussed by Prentice et al. [14] is sometimes referred to as the stratified Anderson-Gill model [23].

The likelihood can be factored into two parts, one part involving  $\beta$  and the other part involving  $\theta$ . The likelihood contribution for the recurrent event process involves  $\beta$  as is given by

$$\prod_{j=1}^{J_i} \lambda_{ij}(t_{ij}) \exp\left(-\sum_{j=1}^{J_i+1} \int_{t_{i,j-1}}^{t_{ij}} \bar{Y}_{ij}(u)d\Lambda_{ij}(u)\right), \quad (3)$$

where  $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u) du$  is the cumulative intensity function for individual  $i$  in stratum  $j$ . The partial likelihood for a sample of size  $m$  is then the product of  $m$  such terms.

The partial score estimating function for  $\beta$  is then

$$\sum_{i=1}^m \sum_{j=1}^{J_i} \int_0^\tau \bar{Y}_{ij}(u) (dN_{ij}(u) - d\Lambda_{0j}(u) e^{\beta v_i}) v_i . \quad (4)$$

The Breslow profile estimate of  $d\Lambda_{0j}(u)$  is

$$d\hat{\Lambda}_{0j}(u) = \frac{\sum_{i=1}^m \bar{Y}_{ij}(u) dN_{ij}(u)}{\sum_{i=1}^m \bar{Y}_{ij}(u) \exp(\beta v_i)} , \quad (5)$$

and substituting (5) into (4) gives the ‘‘profile’’ partial score function

$$U(\beta) = \sum_{i=1}^m \sum_{j=1}^{J_i} \int_0^\tau \bar{Y}_{ij}(u) \left( v_i - \frac{R_j^{(1)}(\beta, u)}{R_j^{(0)}(\beta, u)} \right) dN_{ij}(u) , \quad (6)$$

where  $R_j^{(a)}(\beta, u) = m^{-1} \sum_{i=1}^m \bar{Y}_{ij}(u) v_i^a \exp(\beta v_i)$  and  $a = 0, 1$ . Similarly, we obtain the corresponding score functions for the terminal event intensities as

$$U^d(\theta) = \sum_{i=1}^m \sum_{j=1}^{J_i+1} \int_0^\tau \bar{Y}_{ij}(u) \left( v_i - \frac{S_j^{(1)}(\theta, u)}{S_j^{(0)}(\theta, u)} \right) dN_{ij}^d(u) , \quad (7)$$

where  $S_j^{(a)}(\theta, u) = m^{-1} \sum_{i=1}^m \bar{Y}_{ij}(u) v_i^a \exp(\theta v_i)$  and  $a = 0, 1$ . The score functions (6) and (7) are those of a stratified Cox regression model with one binary covariate. These two score functions form the basis of partial score statistics we used to calculate sample size.

### 3 ASYMPTOTIC PROPERTIES OF PARTIAL SCORE STATISTICS

In this section, we investigate the asymptotic properties of the partial score statistic (6) and (7) under the null and the alternative hypotheses. We focus on trials for which separate marginal hypothesis tests are of interest for the recurrent and terminal event processes; common type I and II error rates are assumed for the two tests but accommodation of different type I and II error rates is trivial. We suppose here that analyses are to be based on at most  $J$  events, but note that  $J$  can be chosen to be large enough to capture all events in any given setting with probability approaching one. Suppose the treatment effect is  $\beta_0$  in (1) under the null hypothesis and  $\beta_A$  under the alternative hypothesis. Under regularity conditions A to D of Andersen and Gill [23] and the assumption that  $mP(Z_i(t) = j | Z_i(0) = 0) \rightarrow \infty$ , for every  $j$  and  $t$ , as  $m \rightarrow \infty$ ,  $U(\beta_0)$  is asymptotically equivalent to

$$\sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_0(R_j^{(1)}(\beta_0, u))}{E_0(R_j^{(0)}(\beta_0, u))} \right) dM_{ij}^r(u) , \quad (8)$$

where  $E_0(\cdot)$  is the expectation taken under the null hypothesis and

$$dM_{ij}^r(u) = d\bar{N}_{ij}(u) - \bar{Y}_{ij}(u) \exp(\beta_0 v_i) d\Lambda_{0j}(u) \quad (9)$$

is the associated martingale under the null. Note that (8) is a sum of  $m$  independent and identically distributed random variables with expectation zero, so it follows from the central limit theorem that

$m^{-\frac{1}{2}}$  times (8) converges in distribution to a zero-mean normal random variable with asymptotic variance

$$\sum_{j=1}^J \int_0^\tau \left[ \frac{r_j^{(2)}(\beta_0, u)}{r_j^{(0)}(\beta_0, u)} - \left( \frac{r_j^{(1)}(\beta_0, u)}{r_j^{(0)}(\beta_0, u)} \right)^2 \right] E_0[\bar{Y}_{ij}(u)e^{\beta_0 v_i} d\Lambda_{0j}(u)] , \quad (10)$$

where  $r_j^{(a)}(\beta_0, u) = E_0[R_j^{(a)}(\beta_0, u)]$ ,  $a = 0, 1, 2$ . This asymptotic variance is similar to the expected information from a stratified Cox regression where the strata are defined by the state of the Markov process.

Under the same set of regularity conditions as under the null hypothesis, the partial score statistic (6) evaluated at  $\beta_0$  is asymptotically equivalent to

$$\sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} \right) d\bar{N}_{ij}(u) , \quad (11)$$

under the alternative hypothesis, where the expectation is taken under the alternative hypothesis. Note that (11) is also a sum of  $m$  independent and identically distributed random variables and it follows from the central limit theorem that  $m^{-\frac{1}{2}}$  times (11) converges in distribution to a normal random variable with mean

$$\sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left\{ E_A(\bar{Y}_{ij}(u)v_i e^{\beta_A v_i} d\Lambda_{0j}(u)) - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} E_A(\bar{Y}_{ij}(u)e^{\beta_A v_i} d\Lambda_{0j}(u)) \right\} . \quad (12)$$

If we let

$$H_{ij}(u) = v_i - E_A(R_j^{(1)}(\beta_0, u))/E_A(R_j^{(0)}(\beta_0, u)) ,$$

the asymptotic variance of  $m^{-\frac{1}{2}}$  times (11) is

$$\sum_{j=1}^J \int_0^\tau E_A(\bar{Y}_{ij}(u)[H_{ij}(u)]^2 e^{\beta_A v_i} d\Lambda_{0j}(u)) , \quad (13)$$

under the alternative.

Thus we have expressions for  $E_A(m^{-\frac{1}{2}}U(\beta_0))$  by (12), the asymptotic variance  $V_0 = \text{Var}_0(m^{-\frac{1}{2}}U(\beta_0))$  of the score statistic under the null by (10), and the asymptotic variance of  $V_A = \text{Var}_A(m^{-\frac{1}{2}}U(\beta_0))$  under the alternative by (13). These results will be used for the sample size calculations in the next section. Details on how the requisite expectations can be carried out are given in Appendix.

For the terminal event under the null hypothesis,  $m^{-\frac{1}{2}}$  times the partial score statistics can be shown to asymptotically equivalent to

$$m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_0(S_j^{(1)}(\theta_0, u))}{E_0(S_j^{(0)}(\theta_0, u))} \right) dM_{ij}^d(u) , \quad (14)$$

where

$$dM_{ij}^d(u) = d\bar{N}_{ij}^d(u) - \bar{Y}_{ij}(u)e^{\theta_0 v_i} d\Gamma_{0j}(u)$$

is the associated martingale process for the terminal event of subject  $i$  at the state  $j$  and  $\Gamma_{0j}(t) = \int_0^t \gamma_{0j}(u)du$  is the baseline cumulative intensity function for the terminal event in stratum  $j$ . The asymptotic variance of (14) is

$$\sum_{j=1}^J \int_0^\tau \left[ \frac{s_j^{(2)}(\theta_0, u)}{s_j^{(0)}(\theta_0, u)} - \left( \frac{s_j^{(1)}(\theta_0, u)}{s_j^{(0)}(\theta_0, u)} \right)^2 \right] E_0[\bar{Y}_{ij}(u)e^{\theta_0 v_i} d\Gamma_{0j}(u)] , \quad (15)$$

under the null, where  $s_j^{(a)}(\theta_0, u) = E_0[S_j^{(a)}(\theta_0, u)]$ ,  $a = 0, 1, 2$ , and under the alternative hypothesis,  $m^{-\frac{1}{2}}$  times the partial score statistic (7) is asymptotically equivalent to

$$m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_A(S_j^{(1)}(\theta_0, u))}{E_A(S_j^{(0)}(\theta_0, u))} \right) d\bar{N}_{ij}^d(u). \quad (16)$$

The asymptotic variance of (16) is

$$\sum_{j=1}^J \int_0^\tau E_A(\bar{Y}_{ij}(u)[H_{ij}^d(u)]^2 e^{\theta_A v_i} d\Gamma_{0j}(u)) \quad (17)$$

under the alternative, where

$$H_{ij}^d(u) = v_i - E_A(S_j^{(1)}(\theta_0, u))/E_A(S_j^{(0)}(\theta_0, u)).$$

## 4 SAMPLE SIZE DERIVATION BASED ON PARTIAL SCORE STATISTICS

### 4.1 SAMPLE SIZE FOR THE DESIGN OF SUPERIORITY TRIALS

In this section, using the partial score statistics of Section 3 we use a score test to calculate sample size requirements for a clinical trial involving recurrent events and terminal event. We illustrate this procedure by testing a treatment effect on the recurrent events. In superiority trials interest is in demonstrating the effectiveness of a new therapy for both the recurrent event process and the terminal event. In particular, we consider the case where  $H_0 : \beta = \beta_0$  and  $H_A : \beta \neq \beta_0$ , where  $\beta_0$  is the null value, and  $\beta_A < \beta_0$  is the value under the alternative that represents the minimal clinically important treatment effect we wish to detect for the recurrent event process. If we assume a follow-up period  $(0, \tau]$ , then under the null hypothesis, the partial score statistics based on (6) is

$$Z = \frac{m^{-1/2}U(\beta_0)}{\sqrt{V_0(\beta_0)}} \quad (18)$$

which converges in distribution to a standard normal random variable.

The approximate one-sided  $100\alpha_1\%$  level partial score test involves rejecting the null if  $Z < z_{\alpha_1}$ , where  $z_\alpha$  is the  $100\alpha\%$  percentile of the standard normal distribution. Under the alternative hypothesis, if we set the power to  $100(1 - \alpha_2)\%$ , we require  $P(Z < z_{\alpha_1} | H_A) = 1 - \alpha_2$ . Straightforward calculations show that the required sample size  $m$  to detect the effect of a reduction in the intensity of events under the new treatment at the significance level of  $100\alpha_1\%$  with power  $100(1 - \alpha_2)\%$  is

$$m = \frac{\left( z_{1-\alpha_1} \sqrt{V_0(\beta_0)} + z_{1-\alpha_2} \sqrt{V_A(\beta_0)} \right)^2}{E_A(U_i(\beta_0))^2}, \quad (19)$$

where  $U_i(\cdot)$  is the contribution of a single individual  $i$  to the partial score statistic (6).

Similarly, the required sample size for detecting superiority of the treatment on the terminal event with power  $100(1 - \alpha_2)\%$  at size  $100\alpha_1\%$  is

$$m^d = \frac{\left( z_{1-\alpha_1} \sqrt{V_0^d(\theta_0)} + z_{1-\alpha_2} \sqrt{V_A^d(\theta_0)} \right)^2}{E_A(U_i^d(\theta_0))^2}. \quad (20)$$

Then the minimum required sample size to detect the superiority of the new treatment on both the recurrent events and terminal event is  $\max(m, m^d)$ .

The asymptotic variances of the test statistics require specification of the baseline intensities, effects of interest (e.g.  $\beta_A$  and  $\theta_A$ ) and information regarding the censoring including the time of administrative censoring and the rate of withdrawal. Parametric assumptions are typically made at the design stage of clinical trials and with time-homogeneous baseline transition intensities  $\lambda_{0j}(t) = \lambda_0 \exp(\psi_\beta(j-1))$  and  $\gamma_{0j}(t) = \gamma_0 \exp(\psi_\gamma(j-1))$ ,  $j = 2, 3, \dots$ , a transition probability matrix can be easily obtained for the control group which is governed by four parameters,  $\lambda_0$ ,  $\psi_\beta$ ,  $\gamma_0$ , and  $\psi_\gamma$  (see Appendix). Historical data on the increase in risk of death with event occurrence and the increase in risk of future events with event occurrence and inform the choice of  $\psi_\beta$  and  $\psi_\theta$ , leaving  $\lambda_0$  and  $\gamma_0$  to specify. Clinical researchers will typically be able to specify a mortality rate over a given period of time  $(0, \tau]$  which represents a constraint on  $P_{0D}(0, \tau|v = 0)$ ; see equation (29) in Appendix. Specification of the expected number of events over  $(0, \tau]$  yields the second constraint which enables specification of the process for the control arm. Specification of the treatment effects on the event and death intensities enables computation of the analogous transition probability matrix for the experimental arm. An assumption of a constant rate of withdrawal is typical in sample size calculations for survival endpoints which can be modeled using an exponential distribution. All calculations required and outlined in Appendix are then possible with this specification; code is available from the authors upon request to facilitate these calculations. The same procedure can be carried out for non-inferiority designs described in the next section.

#### 4.2 SAMPLE SIZE FOR THE DESIGN OF NON-INFERIORITY TRIALS

In this section we address design issues when testing for non-inferiority of a new treatment for both recurrent events and terminal event when compared to a existing active-control. We adopt common notation to formulate the non-inferiority hypotheses [18]. Let  $\text{LRR}(C_1/P_1)$  denote the log-relative risk reflecting the effect of the active-control ( $C$ ) to a placebo ( $P$ ) treatment on the risk of events. The subscript ‘1’ on  $C_1$  and  $P_1$  to denote that this estimate must be known or estimated from historical studies. Similarly, we let  $\text{LRR}(C_2/P_2)$  denote the effect of the active-control to a placebo in the context of planned study. We also let  $\text{LRR}(E_2/P_2)$  denote the log-relative risk for the planned new treatment *versus* a placebo. Though no placebo will be used in the planned study, it is helpful to make indirect comparisons with the effect of the active-control to placebo. In particular, the non-inferiority trial is intended to show that the experimental trial retains a pre-stated percentage of the active-control effect against placebo with a specified power and type I error rate. We formulate the non-inferiority hypotheses for the recurrent events as follows. Let  $\delta_0$  be the percentage of the active-control effect to placebo necessary to retain for non-inferiority claims for the new treatment. The null hypothesis can be formulated as

$$H_0 : \text{LRR}(E_2/C_2) \geq (1 - \delta_0)\text{LRR}(P_1/C_1) \quad (21)$$

which is to be tested against the alternative hypothesis

$$H_A : \text{LRR}(E_2/C_2) < (1 - \delta_0)\text{LRR}(P_1/C_1). \quad (22)$$

For the purpose of sample size calculation, it is sometime desirable to consider a particular value of  $\text{LRR}(E_2/C_2)$  in the alternative hypothesis, which may be expressed as a percentage of the effect of active-control to the placebo. We let  $1 - \delta_A$  denote the percentage of the active-control effect that the experiment treatment retains once the null hypothesis is rejected so that  $\text{LRR}(E_2/C_2) = (1 - \delta_A)\text{LRR}(P_1/C_1) < (1 - \delta_0)\text{LRR}(P_1/C_1)$ . In this study, we examine different values of  $\delta_A$  in sample size calculations.

For testing non-inferiority of the treatment based on the recurrent event, we let  $\beta_0 = \text{LRR}(P_1/C_1)$  and evaluate the partial score statistic (6) at the boundary of the null hypothesis of (21). If we further



suppose that the follow-up duration is  $(0, \tau]$ , the partial score statistic

$$\frac{m^{-1/2}U((1 - \delta_0)\beta_0)}{\sqrt{V_0((1 - \delta_0)\beta_0)}} \quad (23)$$

then converges in distribution to a standard normal random variable  $Z$ , where  $V_0(\cdot)$  is the asymptotic variance of the partial score statistic under the null hypothesis according to (10). Based on a one-sided  $\alpha_1$  level partial score test, to reject the null hypothesis with the power  $1 - \alpha_2$ , one can obtain the required sample size  $m$  for a non-inferiority of the new treatment on the recurrent events as

$$\frac{(z_{1-\alpha_1}\sqrt{V_0((1 - \delta_0)\beta_0)} + z_{1-\alpha_2}\sqrt{V_A((1 - \delta_0)\beta_0)})^2}{E_A(U_i((1 - \delta_0)\beta_0))^2}, \quad (24)$$

where  $V_A(\cdot)$  is the asymptotic variance of the partial score statistic under the alternative hypothesis (13) and  $U_i(\cdot)$  is the contribution of individual  $i$  to the partial score statistic (6). The expectation  $E_A(\cdot)$  is taken with respect to the true model under the alternative as in (8) with  $\beta_A = (1 - \delta_A)\beta_0$ . The required sample size  $m^d$  for testing non-inferiority of new treatment on the terminal event may be obtained by replacing the corresponding quantities in (24) by the ones from the partial score statistic for the terminal event (7) as follows:

$$\frac{(z_{1-\alpha_1}\sqrt{V_0^d((1 - \delta_0^d)\theta_0)} + z_{1-\alpha_2}\sqrt{V_A^d((1 - \delta_0^d)\theta_0)})^2}{E_A(U_i^d((1 - \delta_0^d)\theta_0))^2}, \quad (25)$$

where  $V_0^d(\cdot)$  and  $V_A^d(\cdot)$  are the asymptotic variances for the partial score statistics for the terminal event under the null and the alternative hypotheses, respectively; the expectation  $E_A$  is taken with respect to the model for the terminal event under the alternative with  $\theta_A = (1 - \delta_A^d)\theta_0$ .

The minimum requirement for testing the non-inferiority of the new treatment on both recurrent events and terminal event is  $\max(m, m^d)$  for one-sided test with the level of  $\alpha_1$  and the power of  $1 - \alpha_2$ .

## 5 AN EMPIRICAL STUDY OF FREQUENCY PROPERTIES

We simulate the Markov process with the multiplicative model of (1) for recurrent events and (2) for the terminal event. For planning purposes we set an upper limit to the number of states and set the maximum number of events to  $J = 10$ ; only approximately 2% patients had eight or more skeletal complications in Hortobagyi et al. [20]. For computational convenience, we further specify the intensity function for recurrent event (1) and for the terminal event (2) as  $\lambda_{0j}(t) = \lambda_0 \exp(\psi_\beta \cdot (j - 1))$  and  $\gamma_{0j}(t) = \gamma_0 \exp(\psi_\theta \cdot (j - 1))$   $j = 1, \dots, 10$ , respectively. The constants  $\psi_\beta$  and  $\psi_\theta$  represent the relative increase in the event and death intensity with the occurrence of each additional event. In the simulation study, we consider  $\psi_\beta = \log 1.0 = 0$  for a constant baseline intensity (rate) which is independent of the number of previous events and  $\psi_\theta = \log 1.0 = 0$  to correspond to the setting where mortality is independent of event occurrence. We set  $\psi_\beta = \log 1.1$  to reflect the setting where the event intensity increases with each event and  $\psi_\theta = \log 1.1$  to correspond to the case where the mortality rate increases with event occurrence. The coefficients  $\beta$  and  $\theta$  are the effects of the experiment treatment on recurrent events and death, respectively, and are chosen to represent modest improvements.

The Markov model has eleven states  $(0, 1, \dots, 10)$  corresponding to the cumulative number of recurrent events and one absorbing state for death; we number these states 1 to 12 and consider a  $12 \times 12$  transition intensity matrix denoted  $\mathcal{Q}^v$  for an individual with  $v_i = v$  having  $(k, \ell)$  entry  $q_{kl}^v$  given by  $\lambda_{0k} \exp(\beta v)$  for  $k = 1, \dots, 10$  and  $\ell = k + 1$ ,  $\gamma_{0k} \exp(\theta v)$  for  $k = 1, \dots, 10$  and  $\ell = 12$ ,

$-(\lambda_{0k} \exp(\beta v) + \gamma_{0k} \exp(\theta v))$  for  $k = \ell = 1, \dots, 10$ , and zero otherwise. The transition probability matrix has elements  $P_{\ell k}(t|v) = P(Z(t) = \ell | Z(t^-) = k, v)$  and can be obtained as described in Appendix. We further specify the baseline intensities  $\lambda_0$  and  $\gamma_0$  by setting the probabilities that for a control subject the first event is a recurrent event to  $q = \lambda_0 / (\gamma_0 + \lambda_0)$  and setting the probability that a control subject has died by  $t = 1$  to  $q$  for some pre-specified values of  $p$  and  $q$ .

## 5.1 EMPIRICAL STUDY OF SUPERIORITY DESIGNS

For simulation studies involving superiority designs, under the null hypothesis of no treatment effect we set  $\beta_0 = \theta_0 = 0$ . Under the alternative we set  $\beta_A = \log 0.8$  and  $\theta_A = \log 0.9$ . The duration of the study is set to  $\tau = 1$ . A random censoring time is simulated for each individual using an exponential random variable with a probability of  $P(\tau_i < 1) = 0.2$ . We investigate the performance of the proposed methods for sample size calculations for different scenarios. For each setting the sample sizes are determined according to formulae in (19) and (20). All simulations were implemented in R, and the `coxph` function in the `survival` package was used to obtain the partial score statistics. By setting the `iter.max` and `init` options to zero, the partial score statistics are obtained using the function `coxph.detail`. Under the null hypothesis, the variance of partial score statistic was obtained by summing up the observed information at each event time. Under the alternative, this variance was calculated using the sample variance of the partial score statistics at each event time. For each setting, we conducted 2000 replicates and reported the percentage of those replicates leading to rejection of the null hypothesis as the empirical type I error rate under the null hypothesis, and as the power under the alternative. Table 1 displays the empirical type I error rate and the power for different settings of superiority trials. The empirical type I error rates are consistent with the nominal level of 0.025. For testing for superiority of a new treatment with respect to both the recurrent event and the terminal event,  $\max(m, m^d)$  was the selected sample size which ensured the empirical powers satisfied the nominal requirement of 80%.

Table 1: Sample sizes and empirical rejection rates for tests of superiority for recurrent and terminal events;  $\beta_A = \log(0.80)$  and  $\theta_A = \log(0.9)$ ;  $\%REJ_0$  and  $\%REJ_A$  are the empirical type I error rate and empirical power respectively; the nominal type I error rate is 2.5% and the nominal power is 80%

$\psi_\beta$	Endpoint <sup>†</sup>	Setting <sup>‡</sup>	$\psi_\theta = 1.0$			$\psi_\theta = 1.1$		
			$m$	$\%REJ_0$	$\%REJ_A$	$m$	$\%REJ_0$	$\%REJ_A$
1.0	Recurrent	$\theta = \theta_0$	728	2.45	84.45	771	2.00	83.10
	Recurrent	$\theta = \theta_A$	710	2.65	84.20	753	2.40	82.90
	Death	$\beta = \beta_0$	6636	2.40	80.35	6673	2.30	80.85
	Death	$\beta = \beta_A$	6740	2.50	80.50	6816	2.75	80.50
1.1	Recurrent	$\theta = \theta_0$	691	2.85	84.70	737	2.40	84.25
	Recurrent	$\theta = \theta_A$	674	2.70	84.15	719	2.10	84.25
	Death	$\beta = \beta_0$	6674	2.60	79.45	6691	2.25	81.15
	Death	$\beta = \beta_A$	6759	2.45	80.50	6836	2.40	83.00

<sup>†</sup> Endpoint is the outcome used for the sample size calculation

<sup>‡</sup> Setting is the value of the parameter for the complementary outcome when testing the corresponding endpoint

## 5.2 EMPIRICAL STUDY OF NON-INFERIORITY DESIGNS

In this section we present simulation studies conducted to validate the proposed methods for sample size calculations for testing non-inferiority of the experiment treatment on both recurrent events and terminal event. We demonstrate that the empirical rejection rates are consistent with the nominal levels. In particular, we set  $\text{LRR}(C_1/P_1)=\log 0.6 (\beta_A)$  for the effect of active-control against a placebo for the recurrent events and  $\text{LRR}(C_1/P_1)=\log 0.8 (\theta_A)$  for the terminal event. We also assume the constancy assumption so that  $\text{LRR}(P_2/C_2)=\text{LRR}(P_1/C_1)$ .

We consider these designs where the aim is to demonstrate that the experimental treatment retains at least 50 per cent of the effect of the active-control, so that  $\delta_0 = 0.5$ . In this simulation study, we consider one-sided test with nominal level of type I error rate  $\alpha_1=0.025$  and the power is set to 80 per cent ( $1 - \alpha_2=0.8$ ). The effect of the experiment treatment under the alternative hypothesis is represented by  $\text{LRR}_A(E_2/C_2)=(1 - \delta_A)\text{LRR}(P_1/C_1)$  and we let  $\delta_A = 0.90$  and  $1.00$  to correspond to a retention of 90 and 100 per cent of the active-control effect, respectively. The duration of the follow-up  $\tau$  is set to be 1. A random censoring process is simulated for each subject using an exponential distribution with parameter  $\rho$ , which is specified so that each subject may withdraw from the study with a probability of 0.20 ( $\rho = \log 5/4$ ).

For each simulation setting, the sample size is determined according to the formula (24) and (25). The simulation was implemented in R and the partial score statistics are obtained using `coxph` function in the `survival` package by setting the `iter.max` option equal to zero. The partial score statistics was obtained by setting the `init` option as  $(1 - \delta_0)\beta_A$ . Under the null hypothesis, the corresponding variance was obtained by summing up the observed information of each event time. Under the alternative hypothesis, this variance was calculated by the sample variance of the partial score statistics at all event times.

We conducted 2000 replicates and the percentage of those replicates leading to rejection of the null hypothesis is the empirical type I error rate under the null and the power under the alternative. Table 2 presents the empirical type I error rate and the power for different non-inferiority configurations. The empirical type I error rates are all consistent with the nominal level of 0.025. The empirical powers are all close to the nominal levels for modest and large sample sizes. For simultaneous detecting the superiority of a new treatment on both recurrent events and the terminal event,  $\max(m, m^d)$  equal to the sample size calculated for the terminal event. The empirical powers for simultaneous testing for the superiority are consistent with the nominal level of 80%. Additional simulation studies were conducted with larger effect sizes yielding smaller sample sizes. These demonstrated excellent control of the type I error rate for both the recurrent event and terminal event analyses, and slightly higher empirical power than the nominal level.

We also examine the sensitivity of the sample size calculations to misspecification of the censoring process at the request of a referee. We do this by using the sample size formula of Section 4, simulating the response processes correspondingly, but simulating withdrawal times from a Weibull distribution with shape and scale parameters given by  $a = 2$  and  $b = 2.1199$  respectively to ensure that  $P(\tau_i < 1) = 0.2$  remains satisfied. The same parameter configurations adopted earlier in this section were used here. We again evaluate the empirical type I error rate and empirical power and display these results in Table 3. In general, the results suggest the proposed method is moderately robust to misspecification of the censoring distribution as the empirical type I error rate and empirical power remain quite close to the previous results. There are few cases that the proposed method yielded either lower power (as low as 75%) and higher power (over 90%) but given the degree of misspecification this seems in line with expectations.

Additional sensitivity studies involved exploration of the effect of more general history dependence through the incorporation of subject-specific frailties common to all transitions. Of course if the variability of this frailty is small the state-dependence accommodated in the sample size calculations will be close to adequate and the resultant sample sizes will be reasonable. If this variance is

Table 2: Sample sizes and empirical rejection rates for tests of non-inferiority for recurrent and terminal events;  $\beta_A = \log(0.60)$ ,  $\theta_A = \log(0.8)$  and  $\delta_0 = 0.50$ ; %REJ<sub>0</sub> and %REJ<sub>A</sub> are the empirical type I error rate and empirical power respectively; the nominal type I error rate is 2.5% and the nominal power is 80%

Endpoint <sup>†</sup>	Setting <sup>‡</sup>	$1-\delta_A^d=0.9$			$1-\delta_A^d=1.0$		
		$m$	%REJ <sub>0</sub>	%REJ <sub>A</sub>	$m$	%REJ <sub>0</sub>	%REJ <sub>A</sub>
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	986	2.20	83.05			
Recurrent	$\theta = \theta_A$	967	2.20	83.50	962	2.40	83.60
Death	$\beta = \beta_0$	9665	2.65	81.20	6296	2.65	80.65
Death	$\beta = \beta_A$	9850	2.65	81.40	6405	2.55	81.40
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	664	2.55	83.70			
Recurrent	$\theta = \theta_A$	657	2.65	83.35	655	2.75	82.60
Death	$\beta = \beta_A$	9904	2.75	82.10	6429	2.85	81.85
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	945	2.65	84.30			
Recurrent	$\theta = \theta_A$	934	2.85	84.85	931	2.10	84.90
Death	$\beta = \beta_0$	9669	2.30	80.15	6276	2.30	79.10
Death	$\beta = \beta_A$	9860	2.25	82.35	6401	2.40	81.60
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	639	2.60	83.75			
Recurrent	$\theta = \theta_A$	631	2.90	84.40	629	2.30	84.70
Death	$\beta = \beta_A$	9918	2.20	82.20	6438	2.75	81.55
			$\psi_\theta = 1.1$	$\psi_\beta = 1.0$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	1042	2.15	82.50			
Recurrent	$\theta = \theta_A$	1030	2.25	83.05	1027	2.05	82.10
Death	$\beta = \beta_0$	9761	2.05	81.75	6322	2.65	80.05
Death	$\beta = \beta_A$	9964	2.05	80.50	6475	2.75	79.35
			$\psi_\theta = 1.1$	$\psi_\beta = 1.0$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	701	2.05	83.75			
Recurrent	$\theta = \theta_A$	693	2.65	83.65	691	2.65	82.70
Death	$\beta = \beta_A$	10029	2.75	79.55	6507	2.35	81.05
			$\psi_\theta = 1.1$	$\psi_\beta = 1.1$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	1004	2.85	84.25			
Recurrent	$\theta = \theta_A$	992	2.75	83.20	990	2.95	83.9
Death	$\beta = \beta_0$	9752	2.05	80.52	6329	2.50	80.70
Death	$\beta = \beta_A$	9986	2.25	80.30	6482	2.50	80.15
			$\psi_\theta = 1.1$	$\psi_\beta = 1.1$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	678	2.60	83.70			
Recurrent	$\theta = \theta_A$	670	2.35	84.70	665	2.75	83.65
Death	$\beta = \beta_A$	10053	2.35	79.75	6526	2.25	80.80

<sup>†</sup> Endpoint is the outcome used for the sample size calculation

<sup>‡</sup> Setting is the value of the parameter for the complementary outcome when testing the corresponding endpoint

Table 3: Sensitivity of empirical rejection rates for tests of non-inferiority for recurrent and terminal events to non-uniform withdrawal;  $\beta_0 = \theta_0 = 0$ ,  $\beta_A = \log(0.50)$ ,  $\theta_A = \log(0.5)$  and  $\delta_0 = 0.50$ ;  $\%REJ_0$  and  $\%REJ_A$  are the empirical type I error rate (nominal level 2.5%) and empirical power (nominal level 80%) respectively; random early withdrawal time generated by a Weibull distribution with shape  $a = 2$  and scale  $b = 2.1199$ .

Endpoint <sup>†</sup>	Setting <sup>‡</sup>	1- $\delta_A=0.9$			1- $\delta_A=1.0$		
		$m$	$\%REJ_0$	$\%REJ_A$	$m$	$\%REJ_0$	$\%REJ_A$
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	610	2.50	86.70			
Recurrent	$\theta = \theta_A$	594	2.40	84.45	591	2.60	85.80
Death	$\beta = \beta_0$	1304	3.00	84.20	881	2.05	85.70
Death	$\beta = \beta_A$	1346	2.05	85.30	910	2.75	69.00
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	424	2.90	86.50			
Recurrent	$\theta = \theta_A$	413	2.50	73.75	411	2.55	89.35
Death	$\beta = \beta_A$	1358	2.30	84.80	918	2.35	86.25
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	603	2.90	86.80			
Recurrent	$\theta = \theta_A$	587	2.75	87.45	584	2.30	86.65
Death	$\beta = \beta_0$	1313	2.40	83.00	887	2.55	86.10
Death	$\beta = \beta_A$	1352	2.25	85.15	913	3.00	85.30
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	421	2.20	87.75			
Recurrent	$\theta = \theta_A$	410	2.00	72.90	408	2.30	89.40
Death	$\beta = \beta_A$	1363	2.65	85.35	921	2.20	86.35
			$\psi_\theta = 1.1$	$\psi_\beta = 1.0$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	662	2.27	87.85			
Recurrent	$\theta = \theta_A$	645	2.15	89.00	642	2.20	88.60
Death	$\beta = \beta_0$	1317	2.40	85.30	890	2.05	86.95
Death	$\beta = \beta_A$	1366	1.90	85.25	923	2.05	85.85
			$\psi_\theta = 1.1$	$\psi_\beta = 1.0$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	459	2.75	75.50			
Recurrent	$\theta = \theta_A$	450	2.20	75.75	445	2.20	75.40
Death	$\beta = \beta_A$	1379	2.15	86.95	932	2.35	87.75
			$\psi_\theta = 1.1$	$\psi_\beta = 1.1$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	642	2.45	88.30			
Recurrent	$\theta = \theta_A$	625	2.35	88.80	622	2.15	97.20
Death	$\beta = \beta_0$	1320	2.90	87.00	926	2.85	88.25
Death	$\beta = \beta_A$	1370	2.50	86.45	891	2.35	85.35
			$\psi_\theta = 1.1$	$\psi_\beta = 1.1$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	447	2.40	76.60			
Recurrent	$\theta = \theta_A$	444	2.25	75.60	442	2.40	90.50
Death	$\beta = \beta_A$	1384	3.00	88.35	935	2.10	87.80

<sup>†</sup> Endpoint is the outcome used for the sample size calculation

<sup>‡</sup> Parameter setting for the complementary outcome when testing the corresponding endpoint

large, a stronger state dependence exists and the model is more seriously misspecified; poor frequency properties of the design and analysis will then result.

### 5.3 AN EXTENSION TO TIME NON-HOMOGENEOUS TRANSITION INTENSITIES

It is apparent from Appendix that a key step in the sample size derivation is the calculation of the transition probability matrix for a continuous time Markov process. This Markov process has a finite state space with  $J + 1$  states and one absorbing state  $D$  corresponding to the terminal event. Let  $P(s, t|v)$  denote the  $(J + 2) \times (J + 2)$  transition probability matrix for  $0 \leq s \leq t$ , with  $[k, l]$  entry

$$P(Z(t) = l | Z(s) = k, v) = p_{kl}(s, t|v) ,$$

for  $l = k + 1$  or  $D$ ,  $k = 0, 1, \dots, J$ . This Markov process can be fully specified through the transition intensity matrix as discussed in Cox and Miller [36]. Let  $P(s, t|v)$  and  $Q(v)$  denote the  $(J + 2) \times (J + 2)$  matrix of transition probabilities  $p_{kl}(s, t|v)$  and respective transition intensity matrix.

For a time homogenous Markov process considered in our derivation,  $q_{kl}(t|v) = q_{kl}(v)$  and  $p_{kl}(s, t|v) = p_{kl}(t - s|v)$  for  $l = k + 1$  or  $D$ ,  $k = 0, 1, \dots, J$ . The transition probability matrix is then  $P(t|v) = \exp\{Q(v)t\} = \sum_{x=0}^{\infty} [Q(v)t]^x / x!$  which can be evaluated through Jordan decomposition.

The proposed method can be easily extended to accommodate time nonhomogeneous Markov processes through use of a time transform of the original time scale. Suppose there exists a transformation of the time scale such that  $t = g(u; \varsigma)$  defines a time scale on which the process is homogeneous with transition intensity matrix  $Q_0(v)$  given  $v$ . Then

$$P(u_1, u_2|v) = \exp\{Q_0(g(u_2; \varsigma) - g(u_1; \varsigma)|v)\} . \quad (26)$$

We consider the exponential time transformation [26]

$$g(u; \varsigma) = \varsigma u^\varsigma ,$$

under which the rate of the process may be increasing ( $\varsigma > 1$ ) or decreasing ( $\varsigma < 1$ ), and when  $\varsigma = 1$  the process is time homogenous. In the following simulation study, we set  $\varsigma = 1.2$  and the transition probability matrix was calculated using (26). All sample size formula are still applicable in this case with the specified  $\varsigma$  following the time-transform. The simulation results in Table 4 indicate that the sample sizes derived from the proposed method achieve the nominal type I error rate (nominal level 2.5%) and power (nominal level 80%).

## 6 TRIAL DESIGN IN CANCER METASTATIC TO BONE

Hortobagyi et al. [20] report on the effectiveness of the bisphosphonate pamidronate for the prevention of skeletal related events in breast cancer patients with skeletal metastases. Here we report on analyses of this data to furnish information helpful for the design of a future study planned to have one year duration.

Figure 2 displays the estimates of the cumulative transition intensities for the placebo group for both event occurrence and death. Separate transition intensities were specified for the first to third events (i.e.  $\bar{\lambda}(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\lambda_{0j}(t)$  where  $N_i(t^-) = j$ ,  $j = 0, 1, 2$ ), but the baseline intensity was assumed to be the same for fourth and subsequent events due to sparse data (i.e.  $\bar{\lambda}(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\lambda_{03}^*(t)$  if  $N_i(t^-) = j \geq 3$ ). The risk of the first event appears roughly constant over two years and could be represented with a time homogeneous rate of  $\lambda_0 = 1$  with time measured in years. The slope of the Nelson-Aalen estimates for the event intensities (left panel) are increasing with event occurrence indicating increased risk of future events with each event occurrence. For design purposes a parsimonious representation is required, and the results of fitting a

Table 4: Empirical properties of design based on time nonhomogeneous Markov process with  $\varsigma = 1.2$ : sample sizes and empirical rejection rates for tests of non-inferiority for recurrent and terminal events;  $\beta_0 = \theta_0 = 0$ ,  $\beta_A = \log(0.50)$ ,  $\theta_A = \log(0.5)$  and  $\delta_0 = 0.50$ ; %REJ<sub>0</sub> and %REJ<sub>A</sub> are the empirical type I error rate (nominal level 2.5%) and empirical power (nominal level 80%) respectively.

Endpoint <sup>†</sup>	Setting <sup>‡</sup>	1- $\delta_A=0.9$			1- $\delta_A=1.0$		
		<i>m</i>	%REJ <sub>0</sub>	%REJ <sub>A</sub>	<i>m</i>	%REJ <sub>0</sub>	%REJ <sub>A</sub>
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	657	2.70	81.35			
Recurrent	$\theta = \theta_A$	648	2.60	80.85	647	3.10	81.10
Death	$\beta = \beta_0$	1469	2.40	81.75	985	2.40	83.75
Death	$\beta = \beta_A$	1513	1.95	82.30	1016	2.80	82.60
			$\psi_\theta = 1.0$	$\psi_\beta = 1.0$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	452	3.05	82.85			
Recurrent	$\theta = \theta_A$	446	2.75	83.45	445	2.60	83.55
Death	$\beta = \beta_A$	1526	2.60	82.65	1024	2.35	82.85
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 0.9$		
Recurrent	$\theta = \theta_0$	647	2.45	81.15			
Recurrent	$\theta = \theta_A$	639	2.35	81.70	637	2.05	80.35
Death	$\beta = \beta_0$	1494	2.50	82.10	1002	2.70	82.60
Death	$\beta = \beta_A$	1538	2.95	82.85	1032	3.20	83.80
			$\psi_\theta = 1.0$	$\psi_\beta = 1.1$	$1 - \delta_A = 1.0$		
Recurrent	$\theta = \theta_0$	446	2.50	82.85			
Recurrent	$\theta = \theta_A$	441	2.15	83.50	440	2.35	83.60
Death	$\beta = \beta_A$	1551	2.80	82.05	1041	2.45	81.85

<sup>†</sup> Endpoint is the outcome used for the sample size calculation

<sup>‡</sup> Parameter setting for the complementary outcome when testing the corresponding endpoint

regression model  $\bar{\lambda}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\lambda_0(t) \exp(\psi_\beta N_i(t^-))$  gives  $\hat{\psi}_\beta = 1.41$ . A similar model was specified for the death intensities and the Nelson-Aalen estimates plotted in the right panel of Figure 2 reveal increasing risk of death with the occurrence of each event. When the regression model  $\bar{\gamma}_j(t|\bar{H}_i(t)) = \bar{Y}_{ij}(t)\gamma_0(t) \exp(\psi_\theta N_i(t^-))$  was fit the estimate obtained is  $\hat{\psi}_\theta = 1.36$ ; based on the mortality rate over one year we set  $\gamma_0 = 0.1$ . The censoring rate over the course of a planned study is assumed to be 10% over the 24 months suggesting  $\rho = 0.5^{-1} \log(10/9)$ .

SCENARIO I: Consider the planning of future study aiming to demonstrate that a new treatment is superior with respect to the occurrence of skeletal complication and superior with respect to mortality. We suppose that the overall type I error rate is 5% and a Bonferroni adjustment yields a 2.5% type I error rate for each hypothesis. Suppose two two-sided tests are to be conducted, with each at the 2.5% level to control the overall type I error rate at 5%. Suppose 90% power is required to detect a 20% reduction ( $\beta_A = \log 0.80$ ) in the risk of recurrent events and a 10% reduction in mortality ( $\theta = \log 0.90$ ). We find minimum sample sizes of 700 and 707 individuals, respectively.

SCENARIO II: Suppose a non-inferiority design is of interest and we have margins of 50% for both the recurrent events and death. Suppose the type I error rate for each test is controlled at 2.5% and 80% power is desired for each test. Suppose the true effect of treatment corresponds to a 20% loss of the effect of the active control on survival and a 10% loss of effect on the recurrent event outcome. To ensure 80% power to claim non-inferiority for the survival endpoint, 9052 individuals will be required, and 8506 individuals will be required for the recurrent event outcome.

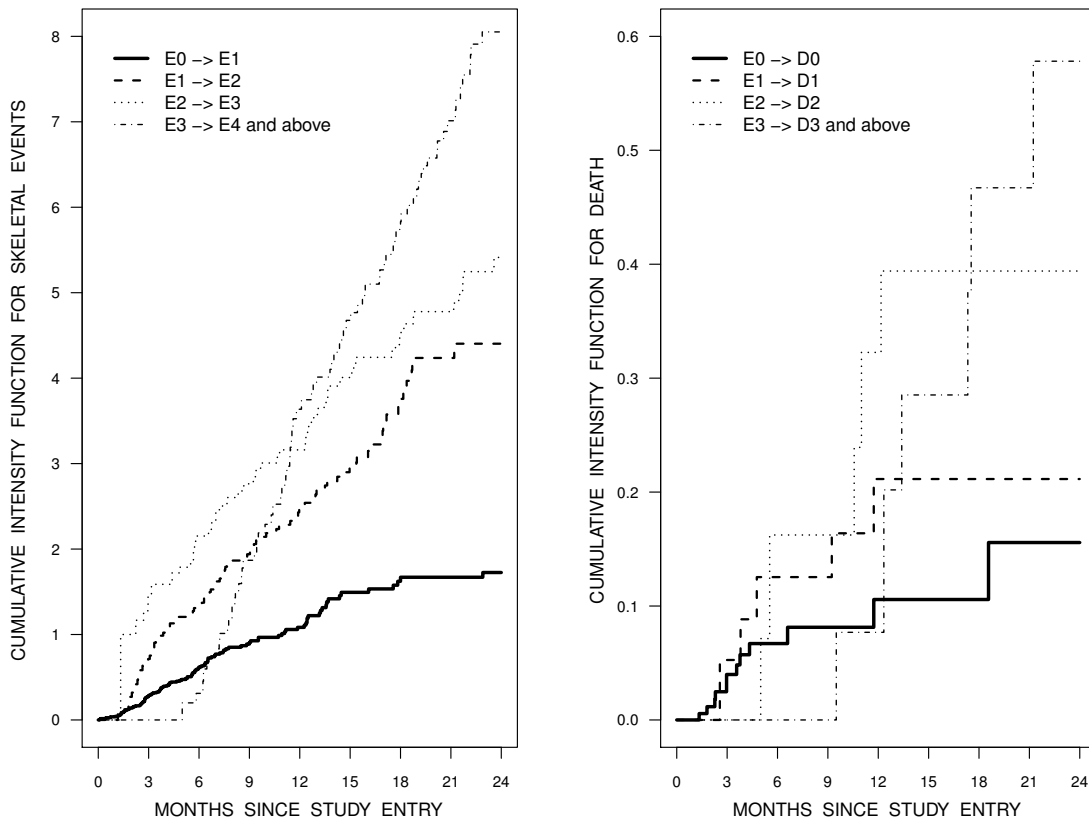


Figure 2: Nelson-Aalen estimates of the cumulative transition intensities for the placebo group in Hortobagyi et al. [20].



## 7 DISCUSSION

This article has provided design criteria for randomized trials with the objective of comparing two treatment groups with respect to the incidence of recurrent events and a terminal event. The motivating setting involves the palliative treatment of cancer patients with skeletal metastases who are at risk of both skeletal related events and death. Recurrent and terminal events arise in many other settings in medical research including transplant studies in which recipients may experience transient graft rejection episodes and total graft rejection [24]. In trials designed to investigate the effect of treatment for advanced chronic obstructive pulmonary disease patients are at risk of recurrent exacerbations and death [25].

The multistate framework adopted is appealing for modeling such processes because it structurally incorporates the terminal events as an absorbing state [22]. This is in contrast to many joint models which incorporate an association between recurrent and terminal events through shared or correlated random effects arising from parametric models. The proposed analysis represents a compromise between use of intensity-based models reliant on full model specification and marginal models. The proposed recurrent event model is in line with the Prentice et al. [14] approach in which the baseline intensity is stratified on the cumulative number of events but has the added implicit condition that subjects must be alive to contribute to the risk set; they are sometimes called “partially” conditional models. The terminal event state therefore enters in the asymptotic calculations by reducing the expected size of the risk sets.

The Nelson-Aalen estimates of the cumulative transition intensities and Aalen-Johansen estimates of the transition probability functions which are estimated under a Markov assumption, are robust in the sense that they remain consistent estimates for non-Markov processes under independent censoring [27, 28]. This is not true for the estimates of treatment effect in multiplicative intensity-based models where there is greater reliance on the model assumptions for valid interpretation of covariate effects. It would be of interest to study the performance of the separate and joint tests of treatment effect in this setting, which involve no conditioning on the event history [29].

Between subject variation in risk of events routinely arises in recurrent event datasets and mixed Poisson models are often adopted since they account for this heterogeneity. The marginal intensity of mixed Poisson processes features a sudden change in risk following event occurrence [21]. This feature is present in the proposed multistate framework but the change in risk is not transient. Boher and Cook [30] showed empirically that the multistate analysis based on the Prentice et al. [14] formulation retains good control of the type I error rate even with naive (i.e. non-robust) variance estimation, so the multistate partially conditional analysis offers some protection against heterogeneity.

Mixed models have also been proposed by several authors for modeling the association between the recurrent and terminal events through correlated or shared random effects [31–33]. Likelihood and semiparametric methods based on estimating functions can be used for analysis of a dataset, but parametric assumptions could be made to derive required sample sizes. We prefer the multistate framework however, since the terminal nature of death is reflected in its designation as an absorbing state. Moreover, with the multistate analysis in which we adopt time-dependent stratification on the cumulative number of events, our sample size formula is directly relevant for analyses based on the so-called Prentice-Williams-Peterson approach [14] to analyze recurrent events in the absence of mortality. While the multistate framework requires that more parameters be specified, the multiplicative increase in risk with event occurrence is seen in a diverse range of datasets and offers some degree of parsimony.

We have restricted attention to settings where the event times are at most right censored. Frequently recurrent events are not observed directly but are only detectable under careful examination in a clinic. Studies aiming to prevent the occurrence of skeletal metastases involve quarterly examinations of patients at which bone scans are conducted to assess whether new metastases have developed.

The same multistate model can be used to characterize the incidence of skeletal metastases and death, but the onset times of the metastases become interval-censored. If the Markov framework remains appropriate, the methods of Kalbfleisch and Lawless [34] may be employed with the multistate model package `msm` in `R/Spplus`. Sample size calculations must be suitably modified and this is a topic of ongoing research.

## ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (LW and RJC) and the Canadian Institutes for Health Research (RJC). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research. The authors thank Jerry Lawless for helpful discussions, the editorial reviewers for helpful comments, and Novartis Pharmaceuticals for permission to use the data from the clinical trial.

## APPENDIX

### ASYMPTOTIC EQUIVALENCE OF THE PARTIAL SCORE STATISTICS

Under the null hypothesis,  $m^{-\frac{1}{2}}$  times the partial score statistic (6) can be written as

$$m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_0(R_j^{(1)}(\beta_0, u))}{E_0(R_j^{(0)}(\beta_0, u))} \right) dM_{ij}^r(u) - m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( \frac{R_j^{(1)}(\beta_0, u)}{R_j^{(0)}(\beta_0, u)} - \frac{E_0(R_j^{(1)}(\beta_0, u))}{E_0(R_j^{(0)}(\beta_0, u))} \right) dM_{ij}^r(u).$$

Using similar arguments in the proofs of Theorem 4.2.1 and 4.3.1 of Gill [35], one can show that the second term of the above expression converges in probability to zero as  $m \rightarrow \infty$  for every  $\beta_0$ .

Similarly, let

$$dM_{ij}(u) = d\bar{N}_{ij}(u) - \bar{Y}_{ij}(u) e^{\beta_A v_i} d\Lambda_{0j}(u)$$

be the associated martingale process for the recurrent event under the alternative hypothesis. One can write the partial score statistic (6) under the alternative hypothesis as follows,

$$m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( v_i - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} \right) d\bar{N}_{ij}(u) - m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \left( \frac{R_j^{(1)}(\beta_0, u)}{R_j^{(0)}(\beta_0, u)} - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} \right) dM_{ij}(u) - m^{-\frac{1}{2}} \sum_{i=1}^m \sum_{j=1}^J \int_0^\tau \bar{Y}_{ij}(u) \left( \frac{R_j^{(1)}(\beta_0, u)}{R_j^{(0)}(\beta_0, u)} - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} \right) e^{\beta_A v_i} d\Lambda_{0j}(u). \quad (27)$$

Using similar arguments as for the null hypothesis, one can show the second term converges in probability to zero as  $m \rightarrow \infty$  for every  $\beta_0$ . We now show the last term of the above expression converges in probability to zero as  $m \rightarrow \infty$ . From the regularity conditions of Andersen and Gill [23], the integrand is locally bounded for every  $u \in (0, \tau]$ . Note that the last term can be written as

$$\sum_{j=1}^J \int_0^\tau \left( \frac{R_j^{(1)}(\beta_0, u)}{R_j^{(0)}(\beta_0, u)} - \frac{E_A(R_j^{(1)}(\beta_0, u))}{E_A(R_j^{(0)}(\beta_0, u))} \right) m^{-\frac{1}{2}} \sum_{i=1}^m \bar{Y}_{ij}(u) e^{\beta_A v_i} d\Lambda_{0j}(u), \quad (28)$$

where  $R_j^{(a)}(\beta_0, u)$  converges almost surely to  $E_A(R_j^{(a)}(\beta_0, u))$  at each time point  $u$ ,  $a = 0, 1$ . It follows from the Slutsky's theorem that the first term of the integrand in (28) converges almost surely to zero as  $m \rightarrow \infty$  at every  $u$ . By the central limit theorem,

$$m^{-\frac{1}{2}} \sum_{i=1}^m \bar{Y}_{ij}(u) e^{\beta_A v_i} d\Lambda_{0j}(u) \quad (29)$$

converges in distribution to a normal random variable at every  $u$  with mean  $\mu$  as  $d\Lambda_{0j}(u)P(\tau_i > u)$  times

$$\sum_{v=0}^1 P(Z_i(u) = j | Z_i(0) = 0, v_i = v) P(v_i = v) e^{\beta_A v}$$

and the variance

$$d\Lambda_{0j}^2(u) P(\tau_i > u) \sum_{v=0}^1 P(Z_i(u) = j | Z_i(0) = 0, v_i = v) P(v_i = v) e^{2\beta_A v} - \mu^2$$

Then, for every  $u$  the integrand in (28) converges in probability to zero. Therefore, it follows from the Lebesgue's dominated convergence theorem that (28) converges in probability to zero as  $m \rightarrow \infty$ .

A similar approach can be used to prove the asymptotic equivalence of the partial score statistics (7) is (14) under the null hypothesis and (16) under the alternative hypothesis.

#### EVALUATION OF EXPECTATIONS UNDER THE TRUE MODEL

The necessary expectations require the evaluation of the probability being in state  $j$  at time  $t$ ,  $P(Z_i(t) = j | Z_i(0) = 0)$ , for the proposed Markov process in Figure 1. As an example, the calculation of  $E_0(\bar{Y}_{ij}(u) e^{\beta_0 v_i} d\Lambda_{0j}(u))$  in (10), is carried out as follows:

$$\begin{aligned} E_0[E_0(\bar{Y}_{ij}(u) e^{\beta_0 v_i} d\Lambda_{0j}(u) | v_i)] &= E_0[e^{\beta_0 v_i} P(\tau_i > u) P(Z_i(u) = j | Z_i(0) = 0, v_i)] \\ &= P(\tau_i > u) \sum_{v=0}^1 e^{\beta_0 v} P(Z_i(t) = j | Z_i(0) = 0, v_i = v) P(v_i = v) . \end{aligned}$$

The transition probabilities are computed as described in the following section.

#### EVALUATION OF THE TRANSITION PROBABILITY MATRIX

The evaluation of expectations under particular models requires the calculation of the Markov transition probability matrix; for notational convenience we suppress the dependence on  $i$ . We consider a finite state space with  $J + 1$  states corresponding to the cumulative number of recurrent events from 0 to  $J$  and one absorbing state  $D$  for the terminal event. For  $0 \leq s \leq t$ , let  $P(s, t|v)$  be the  $(J + 2) \times (J + 2)$  transition probability matrix with  $(k, \ell)$  entry

$$P_{k,\ell}(s, t|v) = P(Z(t) = \ell | Z(s) = k, v) , \quad (30)$$

for  $\ell = k + 1$  or  $D$ ,  $k = 0, 1, \dots, J$ . Let  $Q^v(t)$  denote the transition intensity matrix for individuals in treatment group  $v$ , the elements of which are based on the intensities  $\lambda_k(t|H(t))$  and  $\gamma_k(t|H(t))$  defined in Section 2.

For a time-homogeneous process adopted at the design stage, let  $\lambda_k(t|H(t)) = \lambda_k$  and  $\gamma_k(t|H(t)) = \gamma_k$  be the intensities for  $k - 1 \rightarrow k$  and  $k - 1 \rightarrow D$  transitions, respectively. The transition intensity matrix can then be written simply as  $Q^v$  and has  $(k, \ell)$  entry given by  $\lambda_k$  for  $k = 1, \dots, J$  and

$\ell = k + 1, \gamma_k$  for  $k = 1, \dots, J$  and  $\ell = J + 2, -(\lambda_k + \gamma_k)$  for  $k = \ell = 1, \dots, J$ , and zero otherwise. Under such a time-homogeneous Markov model,  $P(s, s+t) = P(0, t) = P(t)$  and  $P(t) = \exp(Q^v t)$ .

There are several approaches available to compute  $P(t)$  for a given transition intensity matrix  $Q^v$ . If  $Q^v$  has  $J + 2$  linearly independent eigenvectors, let  $A$  be a matrix of eigenvectors, and note that  $AQ^v A^{-1}$  is a diagonal matrix with the eigenvalues  $d_1, d_2, \dots, d_{J+2}$  of  $Q^v$  along its diagonal. Then by the spectral value decomposition [34],

$$\exp(Q^v t) = A \operatorname{diag}(e^{d_1 t}, \dots, e^{d_{J+2} t}) A^{-1}.$$

If  $Q^v$  does not have  $J + 2$  linearly independent eigenvectors, the Jordan canonical form can be used instead [36]. For some nonsingular matrix  $B$ , the Jordan canonical form of  $Q^v$  is  $BQ^v B^{-1} = \mathcal{J} = \operatorname{diag}(\mathcal{J}_1(d_1), \mathcal{J}_2(d_2), \dots, \mathcal{J}_p(d_p))$  and

$$\mathcal{J}_k(d_k) = \begin{pmatrix} d_k & 1 & & \\ & d_k & \ddots & \\ & & \ddots & 1 \\ & & & d_k \end{pmatrix} \quad (31)$$

is a  $n_k \times n_k$  matrix and  $n_1 + n_2 + \dots + n_p = J + 2$ . The matrix exponential  $\exp(Q^v t)$  can be computed [37] as

$$\exp(Q^v t) = B f(\mathcal{J}) B^{-1} = B f(\mathcal{J}_k(d_k)) B^{-1},$$

and in this case  $f(\mathcal{J}_k(d_k))$  takes the form

$$\begin{pmatrix} e^{d_k t} & d_k e^{d_k t} & \dots & \frac{d_k^{n_k-1} e^{d_k t}}{(n_k-1)!} \\ & e^{d_k t} & \ddots & \vdots \\ & & \ddots & d_k e^{d_k t} \\ & & & e^{d_k t} \end{pmatrix}.$$

Numerically, the Jordan decomposition can be obtained through the MATLAB function `jordan` for a given  $Q^v$  and the construction of (31) and hence the transition probability matrix  $P(t)$  can be easily computed in MATLAB. Other methods for computing matrix exponentials are reviewed in Moler and Van Loan [38]. Another numerically stable approach is the method of scaling and squaring [39], which has been employed by MATLAB function `expm` based on an optimal approach [40]. We used this function in sample size calculations for the trial design in cancer metastatic to bone in Section 6.

## SUPPLEMENTARY MATERIALS

The software for sample size calculations using the proposed method is available from the authors upon request.

## REFERENCES

- [1] Andersen PK, Borgan Ø, Gill RD, Keiding N (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- [2] George SL, Desu MM (1974). Planning the size and duration of a clinical trial studying the time to some critical event. *Journal of Chronic Diseases* 27: 15–24.

- [3] Freedman LS (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.
- [4] Schoenfeld DA (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics* 39: 499–503.
- [5] Lachin JM, Foulkes MA (1986). Evaluation of sample size and power for analyses of survival with allowance for non-uniform patient entry, losses to follow-up, non-compliance and stratification. *Biometrics* 42: 507–519.
- [6] Cook RJ (1995). The design and analysis of randomized trials with recurrent events. *Statistics in Medicine* 14: 2081–2098.
- [7] Lawless JF, Nadeau JC (1995). Nonparametric estimation of cumulative mean functions for recurrent events. *Technometrics* 37: 158–168.
- [8] Cook RJ, Lawless JF, Nadeau C (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* 52: 557–571.
- [9] Bernardo MVP, Harrington DP (2001). Sample size calculations for the two-sample problem using the multiplicative intensity model. *Statistics in Medicine* 20: 557–579.
- [10] Fleming TR, Harrington DP (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- [11] Wei LJ, Lin DY, Weissfeld L (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 84: 1065–1073.
- [12] Hughes MD (1997). Power considerations for clinical trials using multivariate time-to-event data. *Statistics in Medicine* 16: 865–882.
- [13] Song R, Kosork MR, Cai J (2008). Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data. *Biometrics* 64: 741–750.
- [14] Prentice RL, Williams BJ, Peterson AV (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68: 373–379.
- [15] Rothmann M, Li N, Chen G, Chi GYH, Temple R, Tsou H-H (2003). Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine* 22: 239–264.
- [16] Kaul S, Diamond GA (2006). Good Enough: A Primer on the Analysis and Interpretation of Noninferiority Trials. *Annals of Internal Medicine* 145: 62–69.
- [17] D’Agostino Sr. RB, Massaro JM, Sullivan LM (2003). Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. *Statistics in Medicine* 22: 169–186.
- [18] Cook RJ, Lee K-A, Li H (2007). Non-inferiority trial design for recurrent events. *Statistics in Medicine* 26: 4563–4577.
- [19] Hortobagyi GN, Theriault RL, Lipton A, Porter L, Blayney D, Sinoff C, Wheeler H, Simeone JF, Seaman JJ, Knight RD, Heffernan M, Mellars K, Reitsma DJ (1998). Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. Protocol 19 Aredia Breast Cancer Study Group. *Journal of Clinical Oncology* 16: 2038–2044.

- [20] Hortobagyi GN, Theriault RL, Porter L, Blayney D, Lipton A, Sinoff C, Wheeler H, Simeone JF, Seaman J, Knight RD (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. Protocol 19 Aredia Breast Cancer Study Group. *The New England Journal of Medicine* 335: 1785–1791.
- [21] Cook RJ, Lawless JF (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.
- [22] Cook RJ, Lawless JF, Lakhali-Chaieb L, Lee K-A (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association* 104: 60–75.
- [23] Andersen PK, Gill RD (1982). Cox regression model for counting processes: a large sample study. *The Annals of Statistics* 10: 1100–1120.
- [24] Cole EH, Cattran DC, Farewell VT, Aprile M, Bear RA, Pei YP, Fenton SS, Tober JA, Cardella CJ (1994). A comparison of rabbit antithymocyte serum and OKT3 as prophylaxis against renal allograft rejection. *Transplantation* 57: 60–67.
- [25] Calverley PMA, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, Yates JC, Vestbo J, for the TORCH investigators (2007). Salmeterol and Fluticasone Propionate and survival in Chronic Obstructive Pulmonary Disease. *The New England Journal of Medicine* 356: 775–789.
- [26] Hubbard RA, Inoue LYT, and Fann JR (2008). Modeling nonhomogeneous Markov processes via time transformation. *Biometrics* 64: 843–850.
- [27] Aalen OO, Borgan Ø., Fekjær H (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* 57: 993–1001.
- [28] Datta S, Satten GA (2001). Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics & Probability Letters* 55: 403–411.
- [29] Ghosh D, Lin DY (2000). Nonparametric analysis of recurrent events and death. *Biometrics* 56: 554–562.
- [30] Boher J, Cook RJ (2006). Implications of model misspecification in robust tests for recurrent events. *Lifetime Data Analysis* 12: 69–95.
- [31] Huang CY, Wang MC (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 99: 1153–1165.
- [32] Liu L, Wolfe RA, Huang X (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60: 747–756.
- [33] Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 8: 708–721.
- [34] Kalbfleisch JD, Lawless JF (1985). The Analysis of Panel Data Under a Markov Assumption. *Journal of the American Statistical Association* 80: 863–871.
- [35] Gill RD (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124. Amsterdam: Mathematical Centrum.

- [36] Cox DR, Miller HD (1965). *The Theory of Stochastic Processes*. Wiley, New York.
- [37] Horn RA, Johnson CR (1994). *Topics in Matrix Analysis*. Cambridge University Press, Cambridge.
- [38] Moler C, Van Loan C (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45: 3–49.
- [39] Moler C, Van Loan C (1978). Nineteen dubious ways to compute the exponential of a matrix. *SIAM Review* 20: 801–836.
- [40] Higham NJ (2005). The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications* 26: 1179–1193.