Published in final edited form as:

Stat Biopharm Res. 2016 ; 8(1): 12–21. doi:10.1080/19466315.2015.1093539.

A Multi-state Model for Designing Clinical Trials for Testing Overall Survival Allowing for Crossover after Progression

Fang Xia, M.S., Stephen L. George, Ph.D, and Xiaofei Wang, Ph.D.*

Department of Biostatistics and Bioinformatics, Duke University School of Medicine

Abstract

In designing a clinical trial for comparing two or more treatments with respect to overall survival (OS), a proportional hazards assumption is commonly made. However, in many cancer clinical trials, patients pass through various disease states prior to death and because of this may receive treatments other than originally assigned. For example, patients may crossover from the control treatment to the experimental treatment at progression. Even without crossover, the survival pattern after progression may be very different than the pattern prior to progression. The proportional hazards assumption will not hold in these situations and the design power calculated on this assumption will not be correct. In this paper we describe a simple and intuitive multi-state model allowing for progression, death before progression, post-progression survival and crossover after progression and apply this model to the design of clinical trials for comparing the OS of two treatments. For given values of the parameters of the multi-state model, we simulate the required number of deaths to achieve a specified power and the distribution of time required to achieve the requisite number of deaths. The results may be quite different from those derived using the usual PH assumption.

Keywords

Crossover; Multi-state survival model; Overall survival; Progression-free survival; Sample size

1 Introduction

In any clinical trial involving patients with a life-threatening disease or condition, overall survival (OS), the time to death from any cause, is an obviously important endpoint, and improvement in OS is widely regarded as the ultimate measure of treatment effectiveness. It is a direct and easily interpreted measure of clinical benefit, and does not suffer from problems inherent in endpoints such as progression-free survival (PFS) that require assessments at various times. However, other treatments given after disease progression or after the protocol-specified therapy has been completed may complicate interpretation of the OS analysis. For example, crossover from the control treatment to the experimental treatment at disease progression leads to bias and various statistical analysis techniques have been proposed to deal with this problem (Ishak et al. (2014)).

*Research reported in this paper was supported in part by the National Cancer Institute of the National Institutes of Health under award number CA142538.

In this paper we address the design of cancer clinical trials for comparing the OS of a standard (control) treatment with a new experimental treatment based on a multi-state model allowing for deaths prior to progression, differential post-progression survival and crossovers from the control treatment to the experimental treatment after progression. Multi-state modeling approaches provide a useful framework for design and analysis when there are potentially intermediate disease states (e.g., a progressed state) experienced prior to death (Andersen et al. (2002); Putter et al. (2007); Farewell and Tom (2014)). These models have been applied, implicitly or explicitly, to the situation we are considering in this paper, but not generally as a formal design tool (Broglia and Berry (2009); Fleischer et al. (2009); Dejardin et al. (2010); Redman et al. (2013); Zhang et al. (2013); Belkacemi et al. (2014); Jönsson et al. (2014)).

Under the multi-state model the proportional hazards assumption does not hold and, since the hazard ratio is not a constant, no single hazard ratio independent of time has any meaning except as an “average” HR for OS over the time period of study, making it generally unsuitable for design or analysis purposes (Hernán (2010)). We compare results using the multi-state model to the usual design approach that incorrectly assumes proportional hazards. The results demonstrate that the required number of events and the distribution of time to achieving those events using our approach may be quite different from those using the PH approach. The proposed method could find broad application in designing clinical trials for many disease areas. Multi-state survival models have been used to describe diseases that involve more than one type of event (Putter et al. (2007)). Some examples include bone marrow transplantation (Klein et al. (2000), Keiding et al. (2001)), HIV infection and AIDS (Longini et al. (1989), Collaboration et al. (2010)), coronary heart disease after the first heart attack (Lloyd-Jones et al. (1999), Wolbers et al. (2014)), and cancer screening (Kirby et al. (1992), Duffy et al. (1995)). These diseases often have death as the event of interest and an intermediate non-fatal event that may change the risk of death. Thus, the impact of non-fatal event should be taken into account in the study design.

2 Design Approaches

Consider the design of a two-treatment randomized clinical trial to compare a standard or control treatment (C) to a new experimental treatment (E) with respect to OS. The null hypothesis is $H_0 : S_C(t) = S_E(t)$, for all t , where $S_i(t)$ is the survival function for treatment $i = C, E$. The test statistic to be used is the ordinary two-sample logrank test, the commonly used test statistic in this setting. The type I error rate is α and the power to be achieved is $1 - \beta$ for given specifications of the design parameters.

2.1 The Multi-state Model Approach

Multi-state models provide a natural and useful conceptual framework for design of trials for OS when patients may enter various disease states prior to death (Hougaard (1999); Broglia and Berry (2009); Fleischer et al. (2009); Beyersmann et al. (2011); Redman et al. (2013); Zhang et al. (2013)). Consider the simple three-state model given in Figure 1, where each patient starts in an initial state (alive, no progression) and thereafter moves to one of the other two states (progressed; dead) with transition intensity (hazard) functions $\lambda_{ji}(t)$ for

progression and $\lambda_{22}(t)$ for death before progression ($i = C$ or E). These hazard rates are the “cause-specific” hazard rates in a competing risks model. The “all-cause” hazard rate, i.e., the progression-free survival (PFS) hazard rate, is simply $\lambda_{PFS}(t) = \lambda_{11}(t) + \lambda_{22}(t)$. Note that no independence assumptions are necessary for the two components of PFS. For those who progress prior to death, the intensity function for moving from the “progressed” state to the “dead” state is $\lambda_{23}(t)$. This formulation allows for differences between control and experimental treatments in any or all of the three intensity functions.

Based on the notation of Fleischer et al. (2009), the random variables associated with each of the three processes for the control and experimental arms are denoted as TTP_i for time to progression, X_i for time to death without progression and OS'_i for time to death after progression. Thus, $PFS_i = \min(TTP_i, X_i)$ and $OS_i = TTP_i + OS'_i$ if $TTP_i < X_i$; otherwise $OS_i = PFS_i = X_i$. In principle, the transition probabilities between states could be modeled in any reasonable way. To avoid complexities in the design, we assume constant transition probabilities (i.e., the time in each state has an exponential distribution with hazard functions $\lambda_{ij}(t) = \lambda_{ij}$ for all t). We are thus assuming a Markov model since the transition hazard depends only on the current state. In general these hazards might depend on the time since the entry to the state or on the time of arrival in the state (Putter et al. (2007)), but these add complexities and assumptions that are difficult to assess at the design stage and these are not considered in the current paper. However, in principle such complexities can be handled via simulation in the same manner outlined below for the exponential distributions. Note that even in the exponential case, although the distribution of PFS is exponential, the distribution of OS is not exponential so the design problem for OS is not a standard one.

We write $Z \sim \text{Exp}(\lambda)$ to indicate that the random variable Z has an exponential distribution with hazard λ . Since $TTP_i \sim \text{Exp}(\lambda_{11})$ and $X_i \sim \text{Exp}(\lambda_{22})$, it follows that $PFS_i = \min(TTP_i, X_i) \sim \text{Exp}(\lambda_{11} + \lambda_{22})$ and the probability ω_i that death occurs before progression is $\omega_i = \lambda_{22}/(\lambda_{11} + \lambda_{22})$. Alternatively, for a given PFS median M_i and a given ω_i , we have $\lambda_{11} = (1 - \omega_i) \ln 2/M_i$ and $\lambda_{22} = \omega_i \ln 2/M_i$.

Although the distribution of overall survival is not exponential, it can be derived explicitly. If there are no crossovers, the cumulative distribution function for OS_i is (Fleischer et al. (2009)):

$$P(OS_i \leq t) = 1 - \frac{\lambda_{i1}}{\lambda_{i1} + \lambda_{i2} - \lambda_{i3}} \exp^{-\lambda_{i3}t} + \frac{\lambda_{i3} - \lambda_{i2}}{\lambda_{i1} + \lambda_{i2} - \lambda_{i3}} \exp^{-(\lambda_{i1} + \lambda_{i2})t} \quad (1)$$

When crossover to the experimental treatment after progression on the control arm is allowed, we assume that crossover occurs immediately after progression. Figure 2 illustrates this situation for the control arm, where it is important to remember that crossover is not based on randomization as part of the design. Thus, patients who crossover are likely to have quite different characteristics and prognoses than patients who do not crossover and the non-crossover subgroup may also receive treatment other than the experimental treatment. The hazard of death for control patients who crossover to the experimental treatment ($\lambda_{C4}(t)$) will

usually differ from those who do not crossover ($\lambda_{C3}(t)$), either because of the effect of the experimental treatment or the differential prognosis of crossover patients.

When crossovers are allowed, the distribution function for OS for patients assigned to the control arm is shown in Appendix 1 to be:

$$P(OS_C \leq t) = 1 - \frac{\lambda_{C1}}{\lambda_{C1} + \lambda_{C2} - \lambda_{C4}} p \exp^{-\lambda_{C4}t} - \frac{\lambda_{C1}}{\lambda_{C1} + \lambda_{C2} - \lambda_{C3}} (1-p) \exp^{-\lambda_{C3}t} + \frac{\exp^{-(\lambda_{C1} + \lambda_{C2})t}}{\lambda_{C1} + \lambda_{C2}} \left[\frac{p\lambda_{C1}\lambda_{C4}}{\lambda_{C1} + \lambda_{C2} - \lambda_{C4}} + \frac{(1-p)\lambda_{C1}\lambda_{C3}}{\lambda_{C1} + \lambda_{C2} - \lambda_{C3}} - \lambda_{C2} \right] \quad (2)$$

The hazard function and the hazard ratio (experimental to control) can be derived from equations (1) and (2) in the usual way.

2.2 The Usual Approach

In the usual design approach, the hazard ratio $= h_E(t)/h_C(t)$ is assumed to be independent of t . For example, in the simple exponential case (i.e., constant hazards: $h_i(t) = h_i$) a close approximation to the required number of events D^* under $\rho: 1 - \rho$ randomization, where ρ the probability of being randomized to the control arm, is given by George and Desu (1974):

$$D^* = \left\lceil \frac{(Z_{\alpha/2} + Z_{\beta})^2}{(\ln \Delta)^2 \rho(1-\rho)} \right\rceil \quad (3)$$

where Z_x is the 100(1 - x) percentile of the standard normal distribution and $\lceil k \rceil$ is the smallest integer $\geq k$.

Straight-forward generalizations to equation (3) that allow for stratification and loss to follow-up have been derived (Bernstein and Lagakos (1978); Rubinstein et al. (1981)). In the general proportional hazards setting, Schoenfeld (1981) derived the required number of events as:

$$N = \left\lceil \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\psi (\ln \Delta)^2 \rho(1-\rho)} \right\rceil \quad (4)$$

where N is the sample size and ψ is the probability of not being censored by the end of the trial, the same result as for exponential distributions but expressed in terms of N and ψ . The quantity ψ depends on the accrual period, the follow-up time, the accrual rate, the individual hazard rates, and the loss to follow-up rates. These approximations work reasonably well as long as the hazard ratio is close to constant. Under the multi-state model described earlier, this assumption does not hold and the usual approach does not yield the correct number of required events.

2.3 Required Number of Events, Sample Size and Duration of Study

Although the first part of the multi-state model described above can also be used in the design of studies comparing PFS and its two components (e.g., see Rauch and Beyersmann (2013)), our concern is with testing OS. In this case, the resulting hazard ratio is not a constant and there are no simple expressions for the required number of events. Lakatos

(1988) presents a method for estimating the required sample sizes in complex clinical trials under non-proportional hazards based on a simulation approach. Lakatos and Lan (1992) compare this method to others and argue for the use of simulation as a basis for sample size calculation in complex settings. Barthel et al. (2006) derive some general results for comparing two or more survival distributions in the non-proportional hazards setting and use simulation to validate their results. Royston and Parmar (2014) propose using a composite null hypothesis and a joint test for the hazard ratio and for the time-dependence of the hazard ratio. In the present paper, we use simulation of the proposed clinical trial under the multi-state model to determine a design to achieve the correct power.

The primary quantity needed at the design stage is the number of events (deaths) required. The sample size, the number of patients to be entered in the trial, must be greater than or equal to the required number of events but is otherwise not uniquely determined by this number. Also, an important characteristic of studies involving time-to-event endpoints with a planned analysis at a targeted number of events is that the time of the final analysis is a random variable. Smaller sample sizes yield longer study durations; larger sample sizes yield shorter study durations.

The required design parameters (Table 1) are the error rates α and β , the accrual rate (a), assumed to be constant; median PFS_i and median OS'_i for both arms; the probability of death before progression (ω_i) for both arms and the probability of crossover after progression for the control patients (p). From these design parameters, we determine, via an iterative simulation procedure, the required number of events D^* , the minimum number of events that yields the required power $1 - \beta$ specified in the design. In contrast to the proportional hazards case, the required number of events D^* in the non-proportional hazards case depends on the accrual time T and a , where $N = aT$. The accrual rate is one of the specified design parameters but N can be any integer D^* , so we also need to specify a desired ratio D^*/N to set the value of N . Details of the simulation approach are given in Appendix 2.

In Table 1, the time-to-event median values are specified in terms of the distributions of PFS and post-progression survival OS' since these are the most natural specifications for the multi-state model. Alternatively, one may wish to specify the median OS for control and experimental treatments rather than the various post-progression survival median OS' values. Unfortunately, as can be seen from Appendix 1, when there are crossovers the distribution of OS depends on the post-progression survival distributions of those who crossover and those who do not. Thus, the knowledge of the distribution of the median OS alone does not provide enough information to define completely the design configuration.

3 Example

3.1 CALGB 30607

In this section we present an example based on a trial (CALGB 30607) conducted by the Cancer and Leukemia Group B (CALGB) in advanced lung cancer (Socinski et al. (2014)). This randomized, phase III, double-blind trial was designed to compare the effect (PFS and OS) of sunitinib (experimental arm) to placebo (control arm) in advanced non-small cell lung cancer (NSCLC) patients who have completed four cycles of first-line platinum-based

therapy. The sample size for the trial was based on comparing the PFS distributions using a two-sample logrank test with a two-sided type I error rate of 0.05. Assuming exponential distributions for PFS, a total of 222 events and a total accrual of 232 patients were stated as required to achieve 90% power for PFS assuming a median PFS of 11 weeks (2.5 months) for the control arm and a median of 17 weeks (3.9 months) for the experimental arm (true HR = 0.647). Patients were to be accrued uniformly over 4 years. The analysis was based on the intent-to-treat principle. In this particular example, crossovers from the placebo arm to the experimental arm were not allowed but we consider the design implications if such an exclusion had not been mandated.

3.2 The Multi-state Model Approach

Since the actual trial design was based only on PFS and no details were given for OS, we make reasonable assumptions for the design parameters for the NSCLC setting, to illustrate our multistate model approach and compare it to the usual approach. The parameters are summarized in Table 2. The median OS' for the experimental arm is assumed to be 9 months; the median OS' for the control arm is assumed to be 6.5 months without crossover and 9 months with crossover. The probability of death before progression is assumed to be 0.1 for both arms. Under these assumptions, the OS hazard functions for the experimental arm and for the control arm (for the four crossover probabilities) are given in Figure 3a and the OS hazard ratios (experimental to control) are given in Figure 3b. The non-constant hazard rates and ratios and the dramatic effect of increasing the proportions of crossovers are clearly apparent in these figures. In particular, for $p > 0$ the hazard ratio, although < 1 , becomes quite large as t increases implying that the detection of OS differences will become increasingly difficult for large p .

In order to give some idea of the OS power in the original design, we consider two hypothetical approaches that might have been used in the original protocol (as noted, no details on OS were given in the protocol): one in which OS is analyzed at the same time that the PFS results are available (i.e., when the 222 PFS events required by the original design are obtained) and the other in which OS is analyzed only after patients are followed for an additional 3 years after this time. The expected number of deaths and the estimated power for both cases are determined by simulation and given in Table 3. As anticipated, the power decreases dramatically as the probability of crossover increases. If there are no crossovers, the power for testing OS is approximately 0.86 at the time of the PFS analysis and approximately 0.90 after three years of follow-up. For $p > 0$ the power after an additional 3 years of follow-up does not increase much, if at all, with additional follow-up and more events. This is due to the non-proportional hazards illustrated in Figure 3b.

The results for a design with power and Type I error rate set at 0.90 and 0.05, based on assuming proportional hazards equal to the ratio of the OS medians, a common but incorrect assumption, may be derived. Although the OS hazard functions are not constant, the OS medians are easily calculated from the distribution function derived in Appendix 1. Using these median values, the ratios of medians are 0.688 ($p = 0$); 0.725 ($p = 0.25$); 0.761 ($p = 0.50$); 0.797 ($p = 0.75$); and 0.848 ($p = 1$). The required number of deaths D^* under this approach can then be calculated using equation (3) and the true power and expected time

$E(T)$ to achieving D^* can be determined via simulation as described in Appendix 2. The sample sizes chosen for each case are set at $N = 1.10D^*$ although any $N > D^*$ could be chosen. The results are given in the first part of Table 4. The required numbers of events calculated under the PH assumption are 302 for $p = 0$; 406 for $p = 0.25$; 563 for $p = 0.50$; 818 for $p = 0.75$ and 1543 for $p = 1$. These results can be compared to the correct D^* , given in the second part of Table 4, assessed via simulation of the multi-state model as described in Appendix 2. For the multi-state model, the required numbers of events are 225 for $p = 0$; 323 for $p = 0.25$; 518 for $p = 0.50$; 947 for $p = 0.75$ and 2114 for $p = 1$. Under either approach, the requisite number of deaths and the expected time to achieve those deaths rise sharply as a function of the proportion of control patients who crossover. However, the required number of deaths under the incorrect PH approach is substantially larger than for the multi-state approach for small $p \leq 0.25$, resulting in unnecessarily large power, and substantially smaller for large $p \geq 0.75$, resulting in an underpowered trial.

4 Discussion

In this paper we describe an approach for designing a clinical trial for testing overall survival in the presence of death before progression, differential post-progression survival and crossover for the control patients to the experimental treatment after progression. The method is based on a multi-state model and uses simulation to determine the required number of OS events and the expected duration of the trial. We require an assumption about post-progression survival after each of the initial treatments in the common situation in which post-progression treatment is not determined by randomization. Alternatively, one could design a study that randomizes patients to alternative treatments at progression if the available alternative treatments are able to be specified beforehand and it is likely that most patients who progress will be able to be randomized. In that case, such a design likely would be preferable to an uncontrolled design. However, in our experience with actual trials, these conditions are rarely met. For that reason, we consider the common uncontrolled situation.

Even without crossovers and even if the assumed medians are correct, the usual approach to designing trials based on a proportional hazards assumption can lead to very different results from those of the multi-state approach. Thus, consideration of the potential effect on overall survival of post-progression survival, crossovers and survival after crossover is important. In particular, a trial that would be feasible with no crossovers may not be feasible with a high percentage of crossovers or with a strong effect on survival after crossover. The sensitivity of the design to the various assumptions concerning death before progression, post-progression survival and crossover after progression should be checked via simulation before the trial is started.

This article addresses the design of clinical trials when there is treatment switching or crossover and an ITT analysis is used. However, if one is interested in testing or estimating the “true” treatment effect that takes the bias introduced by treatment switching into account, other methods may be used. Here we briefly review the two most prominent approaches, the inverse probability of censoring weighting (IPCW) method and the rank preserving structural failure time method (RPSFTM). Comprehensive reviews of these methods can be found elsewhere (Cain and Cole (2009); Zeng et al. (2012); Watkins et al. (2013); Latimer et

al. (2014)). The IPCW method attempts to correct bias by treating patients as artificially censored at the time of switching. The remaining observations are weighted by a model of the censoring probability given covariates. The IPCW method has been studied by Robins and his colleagues (e.g. Robins and Finkelstein (2000); Hernán et al. (2001)) and others. The key assumption of this method is the “no unmeasured confounders” assumption; that is, data must be available on all baseline and time-dependent prognostic variables for mortality that predict informative censoring (switching) and the model of censoring risk must be correctly specified. In practice, the performance of the IPCW method is dependent on large sample sizes and limited levels of covariates entailing treatment switching (Cain and Cole (2009); Watkins et al. (2013); Latimer et al. (2014)). Another method, RPSFTM, was proposed to correct bias introduced by non-compliance (Robins and Tsiatis (1991); Branson and Whitehead (2002)). The RPSFTM is an instrumental variables (IV) method and thus it does not require the “no unmeasured confounders” assumption. However, this method requires the “common treatment effect” assumption that the treatment effect received by switchers is the same as the treatment effect received by patients initially randomized to the experimental therapy. This assumption is not likely to be valid for clinical trials in which patients switch to experimental therapy when their disease has progressed (Watkins et al. (2013); Latimer et al. (2014)). Because of these assumptions, with some exceptions (Finkelstein and Schoenfeld (2011), Jin et al. (2012)), these methods have not been frequently used to analyze clinical trial data when treatment switching occurs (e.g. Latimer et al. (2014)). However, if one of these methods is chosen as the primary analysis to estimate the “true” treatment effect, simulation similar to that described in this paper can be adapted to determine the sample size for a given power or the power for a given sample size.

The assumptions considered here are deliberately simple to illustrate the approach without introducing unnecessary complications and provide an approach that can be easily used in practice. It would be possible to investigate approaches including time-varying intensity functions for the multi-state model, prognostic factors, interim analyses, loss to follow up, crossovers after progression from the experimental treatment to the control treatment, and so on, but at the cost of complexity not usually justified in the design of clinical trials as well as a potential for obscuring the basic principles. However, one of the advantages of using simulation in the design of clinical trials is that such complexity can be easily accommodated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. *Statistical Methods in Medical Research*. 2002; 11:203–215. [PubMed: 12040697]
- Barthel FMS, Babiker A, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine*. 2006; 25:2521–2542. [PubMed: 16479563]

- Belkacemi MC, Castelli C, Remita M, Fournel P, Daurés J. Modelling of overall survival by an association between progression-free and post-progression survival using a conditional distribution. *Statistical Modelling*. 2014; 14:77–98.
- Bernstein D, Lagakos SW. Sample size and power determination for stratified clinical trials. *J Statist Comput Simul*. 1978; 8:65–73.
- Beyersmann, J.; Allignol, A.; Schumacher, M. *Competing Risks and Multistate Models with R*. Springer; 2011.
- Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine*. 2002; 21:2449–2463. [PubMed: 12205692]
- Broglio KR, Berry DA. Overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute*. 2009; 101:1642–1649. [PubMed: 19903805]
- Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident aids or death. *Statistics in Medicine*. 2009; 28:1725–1738. [PubMed: 19347843]
- Collaboration ATC, et al. Causes of death in hiv-1-infected patients treated with antiretroviral therapy, 1996–2006: collaborative analysis of 13 hiv cohort studies. *Clinical Infectious Diseases*. 2010; 50(10):1387–1396. [PubMed: 20380565]
- 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. [PubMed: 20572123]
- Duffy SW, Chen HH, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in medicine*. 1995; 14(14):1531–1543. [PubMed: 7481190]
- Farewell V, Tom BM. The versatility of multi-state models for the analysis of longitudinal data with unobservable features. *Lifetime Data Analysis*. 2014; 20:51–75. [PubMed: 23225140]
- Finkelstein DM, Schoenfeld DA. Correcting for discretionary treatment crossover in an analysis of survival in the breast international group big 1–98 trial by using the inverse probability of censoring weighted method. *Journal of Clinical Oncology*. 2011; 29(9):1093–1095. [PubMed: 21321297]
- Fleischer F, Gaschler-Markefski B, Bluhmki E. A statistical model for the dependence between progression-free survival and overall survival. *Statistics in Medicine*. 2009; 28:2669–2686. [PubMed: 19579225]
- George SL, Desu MM. Planning the size and duration of a clinical trial studying the time to some critical event. *Journal of Chronic Diseases*. 1974; 27:15–24. [PubMed: 4592596]
- Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010; 21:13–15. [PubMed: 20010207]
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*. 2001; 96:440–448.
- Hougaard P. Multi-state models: A review. *Lifetime Data Analysis*. 1999; 5:239–264. [PubMed: 10518372]
- Ishak KJ, Proskorovsky I, Korytowsky B, Sandin R, Faivre S, Valle J. Methods for adjusting for bias due to crossover in oncology trials. *Pharmacoeconomics*. 2014; 32:533–546. [PubMed: 24595585]
- Jin H, Tu D, Zhao N, Shepherd LE, Goss PE. Longer-term outcomes of letrozole versus placebo after 5 years of tamoxifen in the ncic ctg ma. 17 trial: analyses adjusting for treatment crossover. *Journal of Clinical Oncology*. 2012; 30(7):718–721. [PubMed: 22042967]
- Jönsson L, Sandin R, Ekman M, Ramsberg J, Charbonneau C, Huang X, Jönsson B, Weinstein M, Drummond M. Analyzing overall survival in randomized controlled trials with crossover and implications for economic evaluation. *Value in Health*. 2014; 17:707–713. [PubMed: 25236994]
- Keiding N, Klein JP, Horowitz MM. Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in medicine*. 2001; 20(12):1871–1885. [PubMed: 11406847]
- Kirby A, Spiegelhalter D, Day N, Fenton L, Swanson K, Mann E, Macgregor J. Conservative treatment of mild/moderate cervical dyskaryosis: long-term outcome. *The Lancet*. 1992; 339(8797):828–831.
- Klein JP, Szydlo RM, Craddock C, Goldman JM. Estimation of current leukaemia-free survival following donor lymphocyte infusion therapy for patients with leukaemia who relapse after

- allografting: application of a multistate model. *Statistics in medicine*. 2000; 19(21):3005–3016. [PubMed: 11042629]
- Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*. 1988; 44:229–241. [PubMed: 3358991]
- Lakatos E, Lan KK. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine*. 1992; 11:179–191. [PubMed: 1579757]
- Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. Adjusting survival time estimates to account for treatment switching in randomized controlled trials-an economic evaluation context: Methods, limitations, and recommendations. *Medical Decision Making*. 2014; 34:387–402. [PubMed: 24449433]
- Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *The Lancet*. 1999; 353(9147):89–92.
- Longini IM, Clark WS, Byers RH, Ward JW, Darrow WW, Lemp GF, Hethcote HW. Statistical analysis of the stages of hiv infection using a markov model. *Statistics in medicine*. 1989; 8(7): 831–843. [PubMed: 2772443]
- Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*. 2007; 26:2389–2430. [PubMed: 17031868]
- Rauch G, Beyersmann J. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. *Statistics in Medicine*. 2013; 32:3595–3608. [PubMed: 23553898]
- Redman MW, Goldman BH, LeBlanc M, Schott A, Baker LH. Modeling the relationship between progression-free survival and overall survival: The phase ii/iii trial. *Clinical Cancer Research*. 2013; 19:2646–2656. [PubMed: 23669424]
- Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics*. 2000; 56:779–788. [PubMed: 10985216]
- Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics -Theory and Methods*. 1991; 20:2609–2631.
- Royston P, Parmar MK. An approach to trial design and analysis in the era of non-proportional hazards of the treatment effect. *Trials*. 2014; 15:314. [PubMed: 25098243]
- Rubinstein LV, Gail MH, Santner TJ. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *Journal of Chronic Diseases*. 1981; 34:469–479. [PubMed: 7276137]
- Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981; 68:316–319.
- Socinski MA, Wang X, Baggstrom M, Gu L, Stinchcombe T, Edelman M, Baker S, Mannuel H, Crawford J, Vokes E. Switch maintenance with sunitinib (s) in advanced non-small cell lung cancer (nslc): an alliance (calgb 30607), randomized, placebo-controlled phase 3 trial. *International Journal of Radiation Oncology, Biology, and Physics*. 2014; 90:S1.
- Watkins CL, Huang X, Latimer N, Tang Y, Wright EJ. Adjusting over-all survival for treatment switches: Commonly used methods and practical application. *Pharmaceutical Statistics*. 2013; 12:348–357. [PubMed: 24136868]
- Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondré K, Heinze G. Competing risks analyses: objectives and approaches. *European heart journal*. 2014; 353:ehu131.
- Zeng D, Chen Q, Chen MH, Ibrahim JG. Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. *Biometrika*. 2012; 99:167–184. [PubMed: 23049136]
- Zhang L, Ko CW, Tang SH, Sridhara R. Relationship between progression-free survival and overall survival benefit: A simulation study. *Therapeutic Innovation & Regulatory Science*. 2013; 47:95–100.

Figure 1.
Three-state Model.

Figure 2.
Multi-state Model for Control Patients Including Crossovers.

Figure 3.

Table 1

Design Parameters Required for a Multi-state Approach

α : Type I error rate
$1 - \beta$: Power
λ : accrual rate
ω_i : Probability of death before progression (control and experimental)
p : Probability of crossover (control)
Median PFS (control and experimental)
Median OS' (control-no crossover; control-crossover; experimental)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Design Parameters: CALGB 30607

$\alpha = 0.05$
$1 - \beta = 0.90$
$a = 4.8$ patients/month
$\omega_C = \omega_E = 0.10$
$p = 0; 0.25; 0.50; 0.75; 1$
Median PFS = 2.5 months (control) 3.9 months (experimental)
Median OS = 6.5 months (control-no crossover) 9.0 months (control-crossover) 9.0 months (experimental)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Results for Original Design: CALGB 30607

<i>p</i>	<u>At time of PFS analysis</u>		<u>Three years after PFS analysis</u>	
	Exp Deaths	OS Power	Exp Deaths	OS Power
0	185	0.859	230	0.901
0.25	183	0.709	229	0.756
0.50	181	0.514	229	0.531
0.75	179	0.343	229	0.341
1	178	0.205	228	0.198

Table 4

Simulation Results for CALGB 30607 Examples

p^a	PH		Multi-state		
	D^*b	Power	$E(T)^c$	D^*	Power
0	302	0.971	80	225	0.900
0.25	406	0.952	101	323	0.899
0.50	563	0.916	133	518	0.901
0.75	818	0.858	186	947	0.904
1	1543	0.797	338	2114	0.899

^aProbability of crossover
^bRequired number of events (deaths)
^cExpected time (months) to achieving D^* events