Journal of Modern Applied Statistical Methods

Volume 16 | Issue 1 Article 6

5-1-2017

Guidelines for Generating Right-Censored Outcomes from a Cox Model Extended to Accommodate Time-Varying Covariates

Maria E. Montez-Rath
Stanford University, mmrath@stanford.edu

Kristopher Kapphahn
Stanford University, kikapp@stanford.edu

Maya B. Mathur
Stanford University, mmathur@stanford.edu

Aya A. Mitani
Boston University, amitani@bu.edu

David J. Hendry
London School of Economics, david.hendry@ps.au.dk

See next page for additional authors

Recommended Citation
Montez-Rath, M. E., Kapphahn, K., Mathur, M. B., Mitani, A. A., Hendry, D. J., & Desai, M. (2017). Guidelines for generating right-censored outcomes from a Cox model extended to accommodate time-varying covariates. Journal of Modern Applied Statistical Methods, 16(1), 86-106. doi: 10.22237/jmasm/1493597100

Guidelines for Generating Right-Censored Outcomes from a Cox Model Extended to Accommodate Time-Varying Covariates

Cover Page Footnote
Contract sponsor: Patient-Centered Outcomes Research Institute Grant ME-1303-5989

Erratum
The initial publication indicated that "studies" was the subject of the first sentence of the abstract. The subject is more properly "simulating," and the verb "is" is conjugated accordingly. 2017.06.05

Authors
Maria E. Montez-Rath, Kristopher Kapphahn, Maya B. Mathur, Aya A. Mitani, David J. Hendry, and Manisha Desai

Journal of Modern Applied Statistical Methods
May 2017, Vol. 16, No. 1, 86-106.
doi: 10.22237/jmasm/1493597100

ISSN 1538 − 9472

Guidelines for Generating Right-Censored Outcomes from a Cox Model Extended to Accommodate Time-Varying Covariates

Maria E. Montez-Rath
Stanford University
Stanford, CA

Kristopher Kapphahn
Stanford University
Stanford, CA

Maya B. Mathur
Stanford University
Stanford, CA

Aya A. Mitani
Boston University
Boston, MA

David J. Hendry
London School of Economics
London, UK

Manisha Desai
Stanford University
Stanford, CA

Simulating studies with right-censored outcomes as functions of time-varying covariates is discussed. Guidelines on the use of an algorithm developed by Zhou and implemented by Hendry are provided. Through simulation studies, the sensitivity of the method to user inputs is considered.

Keywords: Right-censored outcomes, extended Cox model, time-varying covariates, simulation studies, censoring distribution

Introduction

The development and evaluation of methods for data analysis are often facilitated through simulation studies, particularly when closed-form solutions are unknown (Burton, Altman, Royston, & Holder, 2006). Simulation studies can be especially useful for assessing the behavior of analytic techniques under various conditions that present complexities in practice. For example, Collins, Schafer, and Kam (2001) described the bias that resulted from multiple imputation methods that utilized varying degrees of auxiliary data by simulating data under conditions that varied the percentage missing, the reasons for missingness, and the strength and availability of auxiliary information. Desai, Bryson, and Robinson (2013) performed a simulation study to evaluate properties of robustly-estimated standard errors in the presence of clustering when clustering membership is mis-specified. In research to evaluate and develop methods for handling missing data,

Dr. Montez-Rath is a Biostatistician in the Division of Nephrology. Email her at: mmrath@stanford.edu.

simulating studies with right-censored outcomes as functions of time-varying covariates is critical. There is particular interest in simulating studies with characteristics, including correlation structures over time and across features, that closely resemble a complicated motivating data set.

A large body of research has been devoted to generating right-censored survival times from time-invariant covariates. For example, Leemis, Shih, and Reynertson (1990) demonstrated that survival times that followed a Cox proportional hazards model could be generated by inverting the cumulative hazard function. Independently, Bender, Augustin, and Blettner (2005) offered details on simulating survival times from such a model where the hazard function was assumed to follow exponential, Weibull or Gompertz distributions.

However, generating right-censored outcomes as functions of time-varying covariates is more complicated; a subject's outcome corresponds to multiple values of a covariate over time where the number of values for the covariate may vary across subjects. Using the approach described by Bender et al. (2005) for this purpose is challenging as it would require inversion of the expression $-H_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}(t))$ which includes the cumulative hazard function. Sylvestre and Abrahamowicz (2008) argue that such inversion cannot be easily done since it is only possible if the baseline hazard can be represented by a parametric function. A possible solution is to express changes over time in the covariate, $\mathbf{x}(t)$, as a parametric function that is well-defined over the range of time studied. To that end, Austin (2012) extended the work of Bender et al. (2005) although the extension is limited in that it can only accommodate one time-varying covariate.

Alternatively, Sylvestre and Abrahamowicz (2008) evaluated extending an algorithm first introduced by Abrahamowicz, MacKenzie, and Esdaile (1996) for time-invariant covariates; this algorithm did not require inverting the cumulative hazard function. Instead, the algorithm matches, one-to-one, survival times and covariates that have been generated independently, based on a probability law derived from the partial likelihood of the Cox proportional hazards model. This method allows for any number of time-invariant as well as time-varying covariates without a need to specify a functional form for how they vary over time, but the proposed process of generating the survival times has no closed-form solution. Time-dependent effects, i.e., effects that would vary depending on the time interval, can be introduced directly in the vector of survival times provided to the algorithm but generating those survival times is challenging. Similarly, Crowther and Lambert (2013) proposed a method that relies on numerical integration and allows explicit modeling of the baseline and estimation of the

absolute hazard but it can be computationally expensive if the number of covariates is large.

Independently, Zhou (2001) showed that right-censored outcomes can be generated by transforming a random variable that follows a piecewise exponential distribution, where the hazard is assumed to be constant within a time interval but can vary across time intervals that are defined by changes in the covariate. A closed-form solution for generating the data was also provided. Hendry (2014) developed a general algorithm (with code in R) that implements Zhou's method to generate right-censored survival times under the Cox model with any number of both time-invariant and time-varying covariates that vary at integer-valued steps of the time scale.

This study focuses on Zhou's method for three important reasons. The first is that it is supported by readily accessible software developed by Hendry (2014), providing easy access to a wide audience of potential users. The second is that it can accommodate any number of time-invariant and/or time-varying covariates. Finally, although not highlighted in our study here, Zhou's method provides the additional flexibility of enabling relaxation of the proportionality assumption by allowing the effects to vary between time-intervals (time-dependent effects). Note that the latter is not a feature shared by other methods.

There are multiple user-supplied parameters involved in applying Hendry's implementation of Zhou's method, but properties of the distribution of the outcome may be sensitive to their specification. The primary purpose of this paper is to evaluate these sensitivities and provide guidelines on the use of the Hendry algorithm. To that end, based on an extensive simulation study, we suggest a flexible form for the baseline hazard and characterize the sensitivity of the method to other user inputs under a variety of conditions. Specifically, sensitivities of the algorithm to the censoring distribution are addressed, the shape of the hazard, the degree of correlation between covariates, and the type of covariates. The performance of the algorithm is evaluated through standardized bias and mean squared error of the fitted coefficients and use these statistics to inform guidelines on use of the algorithm.

Cox Regression Models with Time-Varying Covariates via the Piecewise-Exponential Distribution

Zhou (2001) showed that if Y_j, $j = 1,\ldots, J$ are random variables that follow a piecewise exponential distribution, where J indicates the number of intervals, and g(.) is a monotone increasing function such that $g(0) = 0$ and $g^{-1}(t)$ is

differentiable, then $g(Y_j)$ follows a Cox model with a time-varying covariate and a baseline hazard $h_0(t) = d/dt[g^{-1}(t)]$. To incorporate covariates, one can specify the piecewise exponential variables with varying rates γ_j such that they depend on any number of time-invariant and/or time-varying covariates $\mathbf{Z}_j = Z_{j1},\ldots, Z_{jP}$ and regression parameters $\boldsymbol{\beta} = \beta_1,\ldots, \beta_P$ where $\boldsymbol{\gamma}_j = \exp(\boldsymbol{\beta}\mathbf{Z}'_j)$. In this form, effect sizes can easily be introduced as the components of the rates for the piecewise exponential variates where the hazard of $g(Y_j)$ is defined by $h_0(t)\exp(\boldsymbol{\beta}\mathbf{Z}'_j)$. Time-dependent effects can be introduced by allowing the effects to vary between time intervals ($\boldsymbol{\beta}_j = \beta_{j1},\ldots, \beta_{jP}$) and so $\gamma_j = \exp(\boldsymbol{\beta}_j\mathbf{Z}'_j)$.

Hendry (2014) demonstrated that piecewise exponential random variables with support $[a, b]$ such that $0 < a < b$ (truncated piecewise exponential random variables), can be generated through an accept-reject algorithm where realizations outside of the support are discarded and those within are included. A full proof for how one can generate survival times that follow a Cox model with time-dependent covariates using a truncated piecewise exponential distribution can be found in Hendry (2014). Key parameters of the algorithm that need to be defined are: the bounds of truncation (a, b), the parameters corresponding to the piecewise exponential random variables or rates γ_j, the transformation function g, and the censoring mechanism.

The bounds of truncation relate to the limits of observed survival times, which can be informed by an empirical data set. For example, a lower bound $a > 0$ can correspond to a lower bound on subject eligibility (e.g., it may be that only subjects who are considered "active" users of a health system – i.e., who exceed a minimum duration of observation – are eligible for study). Note this form of truncation is not to be confused with left truncated time-to-event data, where the latter would constrain observational times to begin at the lower bound. In contrast, here observational times begin at zero but are only included if they exceed the lower bound. The upper bound corresponds to the maximum allowable time observed for an individual. The larger the upper bound, the larger the number of records per individual. This has implications not only for the time needed to generate the data, but also for the run time of any application of the simulated data.

The g function has an important role. It is defined such that $g^{-1}(t) = H_0(t)$, the cumulative baseline hazard of some known function. It should be specified to best represent the disease or process of interest. Options described by Hendry are mostly power functions and tend to lead to large hazards such that events occur soon after the start of observation. Hendry suggests exploring a variety of functional forms to appropriately capture the process studied but does not offer much guidance on parameter choice.

It is recommended the g function be defined through the use of a Weibull distribution so that the variates generated in the process have a baseline hazard of a Weibull random variable. This distribution is a flexible choice defined by two parameters: 1) shape, v, which determines whether the hazard is increasing over time ($v > 1$), constant ($v = 1$), or decreasing ($v < 1$); and 2) scale, λ, which shifts the hazard distribution right or left, depending on the overall survival time. The Weibull distribution has a hazard function defined by $h_0(t) = \lambda v t^{v-1}$ and the cumulative hazard equal to $H_0(t) = \lambda t^v$. By fixing v, one can generate outcomes with a pre-determined median survival time informed by the empirical data (which we will call the target median).

Assume $g^{-1}(t)$ to be the cumulative baseline hazard from the Weibull distribution with shape parameter v and scale parameter λ. The estimated median survival time, $\hat{t}(50)$, for an individual whose vector of explanatory variables is $\mathbf{Z} = (Z_1, \ldots, Z_p)$ with estimated effects $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$, is defined by

$$\hat{t}(50) = \left\{ \frac{\log 2}{\hat{\lambda} e^{\hat{\boldsymbol{\beta}} \mathbf{z}'}} \right\}^{\frac{1}{\hat{v}}}$$

(Collett, 2003, p. 177). This formula can be used to compute a value for λ, given v and a target median. This is done for a hypothetical individual whose covariate values are at the mean $\left(\hat{\boldsymbol{\beta}} \bar{\mathbf{Z}}' = \hat{\beta}_1 \bar{Z}_1 + \ldots + \hat{\beta}_p \bar{Z}_p\right)$.

There are a variety of options that the user can consider for incorporating censoring into the data generation process. Often studies impose administrative censoring where subjects are no longer observed beyond the study end date. This is fairly straightforward to define once times to the event have been generated. However, censoring may arise for other reasons, like when subjects drop out of or withdraw from a study and are lost to follow-up.

There are two main ways to implement this type of censoring. One is referred to as traditional censoring, in which both a survival time and a censoring time are generated and then the minimum value of the two is chosen as the time the subject was observed. If the minimum value was the survival time, an indicator for whether the subject was observed to have the event will equal 1. Otherwise, if the subject's time was censored, the indicator will equal 0. Hendry's algorithm can also be used to impose traditional censoring. To obtain the intended percent of observations being censored, though, additional parameters need to be specified and refined by iteration. The second alternative, referred to as random

censoring, is easier to implement and computationally more efficient. In this approach, each patient's observation is simply censored at random with a probability determined by the percentage of censored observations desired. An indicator for whether the time was censored follows a Bernoulli distribution with a pre-specified probability. For more on incorporating censoring into simulations, see Crowther and Lambert (2013), Burton et al. (2006), and Bakoyannis and Touloumi (2012).

Consider the impact of these parameters on generating data that closely mimic a motivating data set. Specifically, the investigation in this study is on the impact of the parameters for the Weibull distribution, censoring mechanism, correlation among variables, and variable type on properties of estimates obtained from fitting an extended Cox model to data generated using this approach, as well as a generated survival time distributions and variation in computation time.

Methodology

Design of Simulation Study

The parameters of the simulation study follow a full factorial design of the following parameters:

Bounds of truncation (a-b): (20-300), (20-150), (20-50)
Covariate combinations: 2 Normal, $Z_1 \sim N(50, 10^2)$ and $Z_2 \sim N(30, 52)$; 1 Normal + 1 Binary, $Z_1 \sim N(50, 10^2)$ and $Z_2 \sim Bern(0.5)$
Weibull shape parameter (v): 2, 1, 0.5
Target median: 35; 75; 150
Censoring distribution: None; Random; Traditional; Administrative
Percent censored patients (if censoring applied): 20%; 50%; 80%

Data were generated using all possible combinations of the parameters listed with the exception of the percent of patients censored, which is relevant only when an actual censoring distribution is being applied. Details on the choice of parameters are described here.

Data Generated

Using Hendry's algorithm, survival times were generated to fall within 2 bounds of truncation defining the range of possible survival times. The lower bound, *a*, was fixed at 20 and the upper bound, *b*, was allowed to vary (50, 150, and 300).

Survival times depended on two independent time-varying covariates (Z_1, Z_2) in two possible combinations: one in which both covariates are normally distributed random variables and a second in which one covariate is normally distributed and the other is a Bernoulli random variable. Specifically, Z_1 is always assumed to be N(50, 10^2) and Z_2 could either be N(30, 5^2) or Bernoulli(0.5).

The transformation function g was specified to be derived from a Weibull distribution with shape parameter corresponding to an increasing (v = 2), constant (v = 1), or decreasing (v = 0.5) hazard. The scale parameter is computed after providing the shape parameter and the target median survival time, which vary (35, 75, and 150). Note that some target medians fell outside the bounds, demonstrating the impact of parameter choice. For a given shape parameter (v), target median (M), and vector of regression parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)$, the scale parameter λ and the g function are defined as follows:

$$\lambda = \frac{\log 2}{\boldsymbol{\beta}\bar{\mathbf{Z}}'} M^{-v},\ \mathrm{g}(t) = \left(\lambda^{-1} t\right)^{\frac{1}{v}},\ \text{and}\ \mathrm{g}^{-1}(t) = \lambda t^{v}$$

where $\bar{\mathbf{Z}} = \left(\bar{Z}_1, \bar{Z}_2\right)$ is the vector of means of the covariates. Under the scenario where $Z_1 \sim \mathrm{N}(50, 10^2)$ and $Z_2 \sim \mathrm{N}(30, 5^2)$ then $\boldsymbol{\beta}\bar{\mathbf{Z}}' = 50\beta_1 + 30\beta_2$. However, if $Z_2 \sim$ Bernoulli(0.5), then $\boldsymbol{\beta}\bar{\mathbf{Z}}' = 50\beta_1 + (0.5)\beta_2$.

The algorithm computes survival times within the defined limits, which might be considered the "true" event times and may or may not be observed depending on the censoring method applied. For administrative and traditional censoring, these are the uncensored times. We then imposed 3 types of censoring (administrative, traditional, and random) with various percentages of patients being censored (20%, 50%, or 80%). In administrative censoring, patients are observed until a fixed time (end of study). In traditional censoring, censoring times were generated in parallel with the uncensored survival times using an independent implementation of the Hendry algorithm. The parameters of the censoring distribution are chosen by iteration to yield the correct amount of censored observations and are different than the parameters used in the creation of the uncensored times, thus reflecting non-informative censoring. The final observed time is defined as the minimum of the two survival times. The event indicator is set to 0 if the censoring time is smaller than the uncensored time. In random censoring, each subject has a probability p_c, set to 0.2, 0.5, or 0.8 (depending on the percentage of censoring desired), of being censored at the end

of the subject's generated survival time. Event indicators were thus distributed as Bernoulli random variables with $p = 1 - p_c$.

The influence of the correlation between the covariates on properties of estimates obtained from fitting the Cox model is examined. The covariates, as defined above, were allowed to be correlated, with correlations ranging from -0.8 to 0.8, using the mvrnorm function in R when generating two Normal random variables and the binnor package in R when generating one Normal and one Bernoulli variable (Demirtas & Doganay, 2012). In this scenario, survival times were set to be bounded between 20 and 300, the shape parameter, v, was fixed at 2, and the target median was fixed at 150.

Number of Replications

For each scenario or combination of the simulation parameters, we drew 1000 simulated data sets (replicates) each with 1000 individuals with varying number of observations per individual depending on the scenario being simulated and the data generated.

Parameters to Be Estimated

We fit the true model (an extended Cox model) to the data generated and obtained estimates for the regression coefficients corresponding to the two covariates. Parameters were set to $\beta_1 = 0.02$ and $\beta_2 = 0.04$ when covariates were two Normal variables and $\beta_1 = 0.02$ and $\beta_2 = -0.5$ when covariates were one Normal and one Bernoulli.

Evaluation Criteria

The performance of the algorithm was assessed by three statistics, which were computed for each parameter estimate $\left(\hat{\beta}_1, \hat{\beta}_2\right)$: the standardized bias (difference between the average estimate and the true value as a percentage of the estimate's empirical standard error), the mean squared error (MSE, squared difference between the true and estimated parameter averaged over the number of simulations), and the coverage percentage (percentage of time the 95% confidence interval contains the true parameter). As suggested by Collins et al. (2001), standardized bias larger than 40% (in absolute value) is considered to indicate poor performance. Although nominal coverage percentage is 95%, Collins and others defined acceptable coverage as 90% or higher. In order to

assess how close the distribution of the generated survival times is to the distribution of times in the empirical data

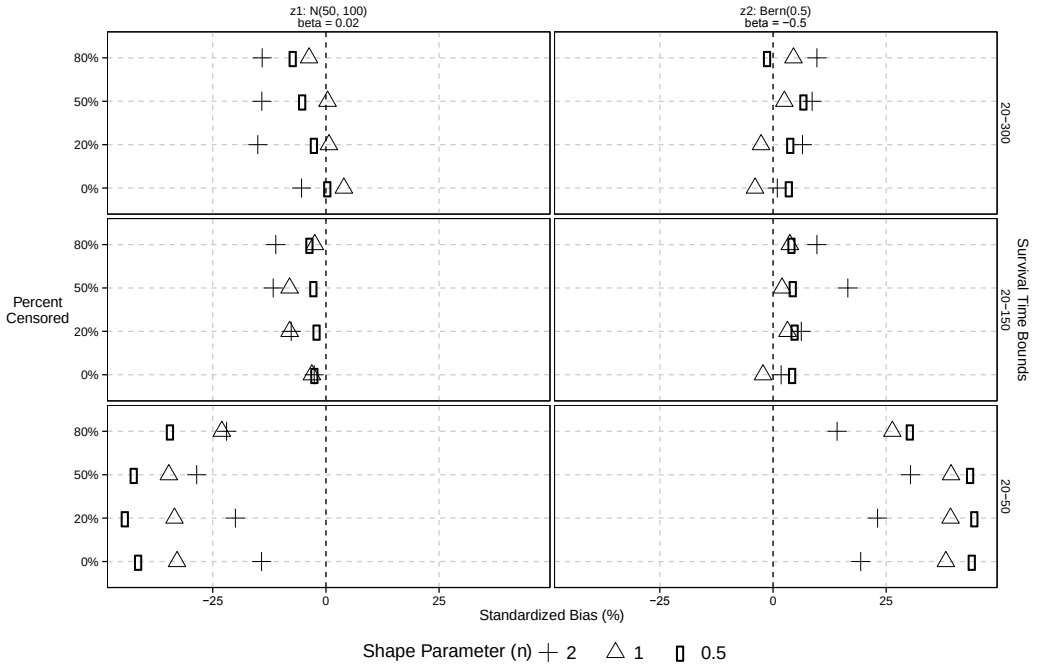

Figure 1. Standardized bias in fitted coefficients when survival times are generated with target median of 35, random censoring, and using mixed covariates

set, we graphically assess the median survival times generated. Finally, we compare algorithm run times across different combinations of simulation parameters.

Results

Impact of Limits on the Survival Times Generated

Data generated under the most restrictive bound (20-50) with a median goal equal to 35, independent of the types of covariates, yielded large standardized bias relative to the other two bounds (e.g., the range of standardized bias for Normal covariates was -48.4 to -14.2, -11.6 to -0.7, and -15.0 to 4.0 for the 20-50, 20-150, and 20-300 bounds, respectively) (Figure 1, left column). For both traditional and administrative censoring, under the most restrictive bound, the standardized bias

decreased as the percent of censored observations increased (e.g., the range of standardized bias for two Normal covariates assuming traditional censoring was -40.9 to -14.2 and -6.3 to -3.2 for 0% and 80% censored, respectively). This was not the case for random censoring, however. For example, the range of standardized bias when the two covariates are normally distributed was -40.9 to -14.2 and -33.9 to -22.0 for 0% and 80% censored observations, respectively (Figure 2, left column). Results are not shown for administrative censoring.

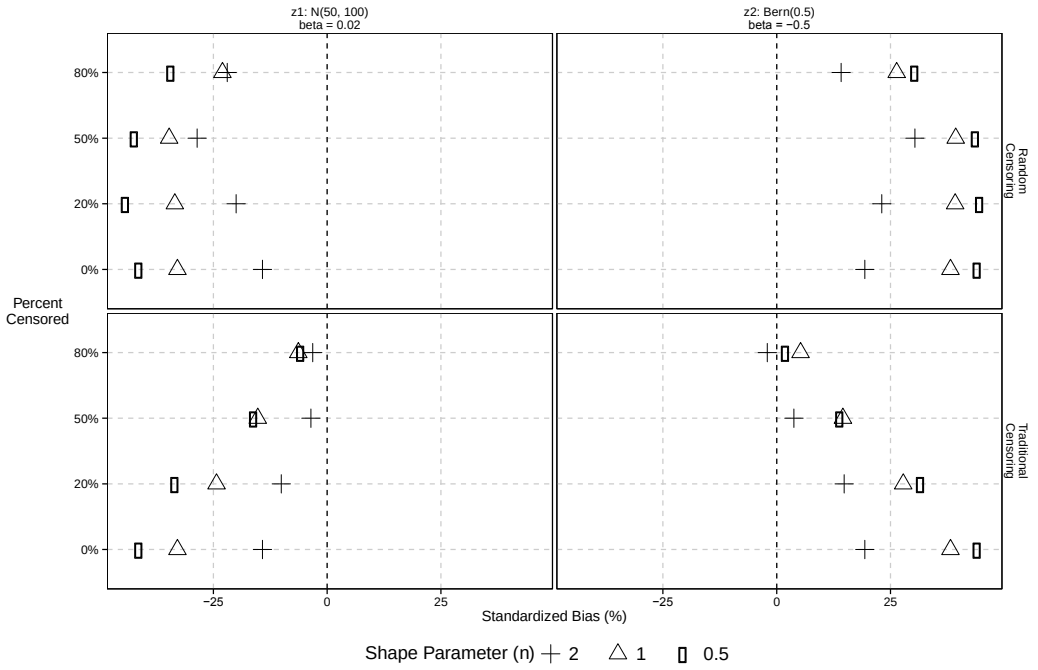

Figure 2. Standardized bias in fitted coefficients when survival times are generated bounded between 20 and 50, target median of 35, and using mixed covariates contrasting random versus traditional censoring

---

Coverage percentages were close to 95% for most combinations of the parameters simulated (0.89-0.97). Somewhat lower coverage – although still over 90% – was obtained when generating times using smaller limits with random censoring (e.g., coverage percentages for 2 Normal covariates assuming a target median of 35 were 0.92 to 0.95, 0.94 to 0.97, and 0.94 to 0.96 for bounds of 20-50, 20-150 and 20-300, respectively) (Figure S1A, middle columns, rows 4-6).

Independent of the bounds and censoring type assumed, when the percent of censored observations increased from 0% to 80% censored for the binary covariate, the MSE increased from 0.004 to 0.027 (Figure S2C, columns 2, 4 and 6, rows 1-3). In contrast, the MSE remained close to zero when covariates followed a Normal distribution (Figure S2C).

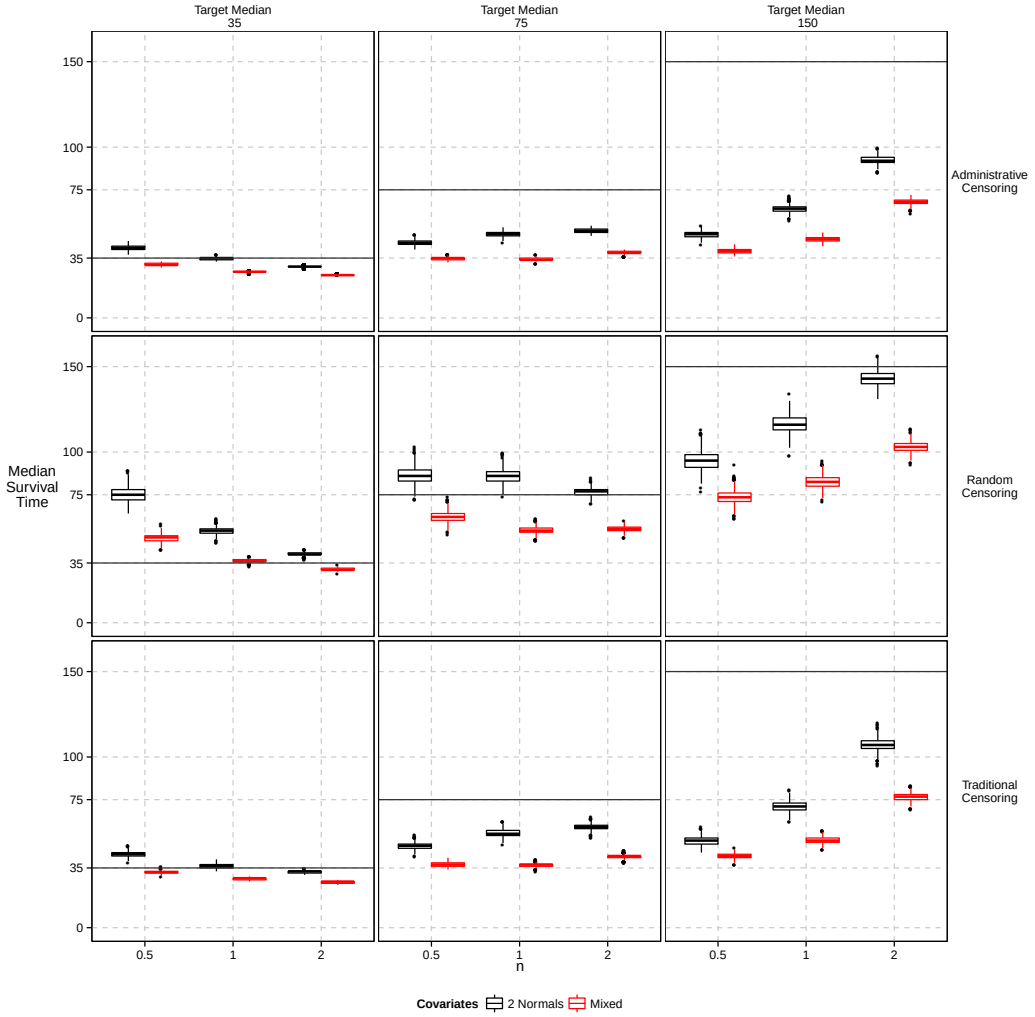

Figure 3. Median survival times for data generated bounded between 20 and 300; straight lines indicate the relevant target median

Impact of g Function Definition

Overall, under data generated with the least restrictive bound (20-300), when the shape parameter for the Weibull distribution was 2 (compared to 1 or 0.5), the median of the generated survival times came closer to the target (Figure 3).

More specifically, under random censoring, a shape parameter of 2 yielded survival time distributions with medians closer to the target median relative to the other shape parameter choices (e.g., for a target median equal to 150, under random censoring with 2 Normal covariates, median survival times ranged from 131.0 to 156.5 and 77 to 113.0 when $v = 2$ and 0.5, respectively). The value was almost on target when covariates were both generated from the Normal distribution but fell short when covariates were of mixed type (e.g., for a target median equal to 150 with $v = 2$, under random censoring, median survival times ranged from 131.0 to 156.5 and 93.0 to 114.0 when the two covariates were both normally distributed and mixed, respectively) (Figure 3, middle row).

Computational efficiency was affected by the choice of target median. The median run time increased as the target median increased (Figure 4). For random, administrative, and traditional censoring, respectively, run times ranged from 10.7 to 132.9 seconds, from 20.9 to 163.2 seconds, and from 49.7 to 213.2 seconds.

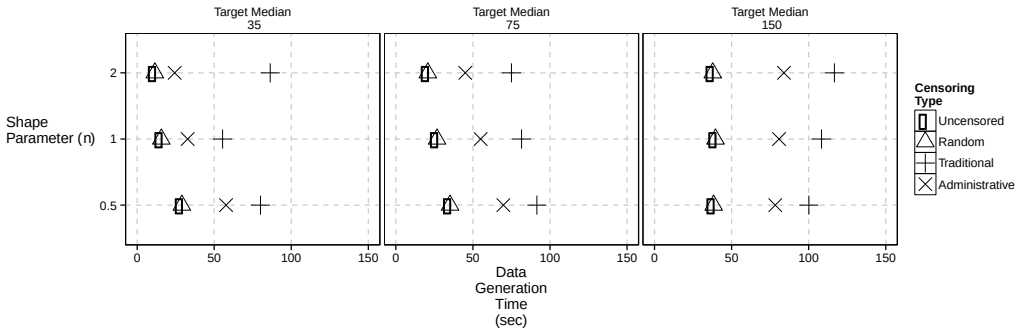

Figure 4. Median run times for censoring type when survival times are generated bounded between 20 and 300 and using mixed covariates

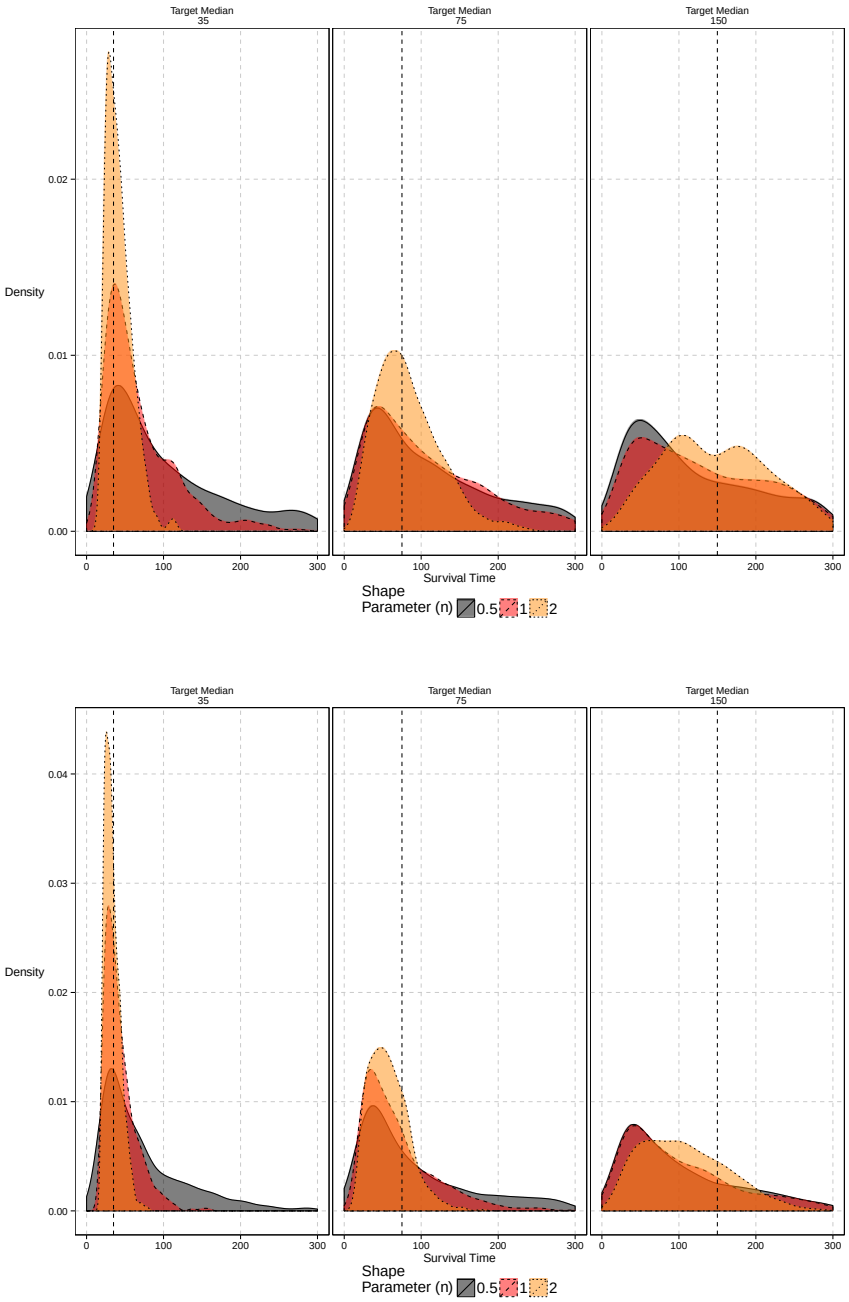

Figure 5. Example of density plots of generated survival times with random censoring for the various shape parameters when covariates are (a), above, two Normal random variables and (b), below, one Normal and one Bernoulli random variable

Overall, using a shape parameter equal to 2 provided distributions of survival times that have a median closer to the target value (Figure 5). We observed no differences in bias, MSE, or coverage probabilities between the choices of target median when the range of survival times is large (Figure S2).

Impact of Censoring Type

There was no difference in overall statistical performance of the algorithm by censoring type. No differences were found in standardized bias and coverage. For all censoring types, as the percent of censored observations increased, the MSE increased, ranging from 0.004 to 0.027 for the binary covariate and remaining low (on the order of 10-5) for Normal covariates (Figure S2).

However, it was found that computational run times were strongly affected by censoring type. Beyond the first step in the algorithm of generating uncensored survival times, random censoring took no additional time whereas traditional censoring more than doubled the run time (Figure 4). For example, for data generated with limits of 20 and 300, a target median of 75, and $v = 2$, median run times were 26.4, 26.6, and 89.2 seconds for uncensored, random and traditional censoring, respectively.

Impact of Type of Covariates and of Correlation

Negligible differences were found in performance by type of covariates or assumed correlation. In general, positive bias was found in the fitted coefficients corresponding to the binary covariate (e.g., for a target median equal to 35 and under random censoring, bias ranged from -4.0 to 9.7, -2.2 to 16.5, and 14.2 to 45.1 for bounds 20-300, 20-150, and 20-50, respectively) and negative bias for coefficients of the Normal covariates (e.g., for a target median equal to 35 and under random censoring, bias ranged from -12.3 to 2.1, -9.7 to -0.7, and -48.4 to -26.2 for bounds 20-300, 20-150, and 20-50, respectively). However, bias was negligible when the range of survival times generated (bounds) is large (Figure 1, left vs. right columns).

Median survival times generated were lower when using 1 Normal and 1 binary covariate compared to when both covariates were normally distributed (Figure 3). For example, for data generated between limits of 20 and 300 with a target median of 75 and $v = 2$, median survival times ranged from 36.0 to 60.0 and from 48.0 to 85.0 when the covariates were of mixed type and normally distributed, respectively (Figure 3, middle column).

In the subset of simulations performed to allow for varying correlation between the 2 covariates, we found a slight increase in the absolute MSE (e.g., from 2.0×10^{-5} to 4×10^{-5} for a Normal covariate with effect size equal to 0.02) as correlation increased. This is true for both combinations of the covariates (Figure 6). The MSE increased as the effect size increased, an effect that was more pronounced for the binary covariate (Figure 6b). For example, for data generated between 20 and 300, assuming a target median of 150, $\nu = 2$, and 50% observations randomly censored, when the two covariates were independently generated from a Normal distribution, the MSE for the covariate with effect size equal to 0.02 was 2.0×10^{-5}. When the effect size was instead 0.04, the MSE was 8.2×10^{-5}. The MSE for the coefficient of the uncorrelated binary covariate with an effect size of -0.5 equaled to 9.5×10^{-3}.

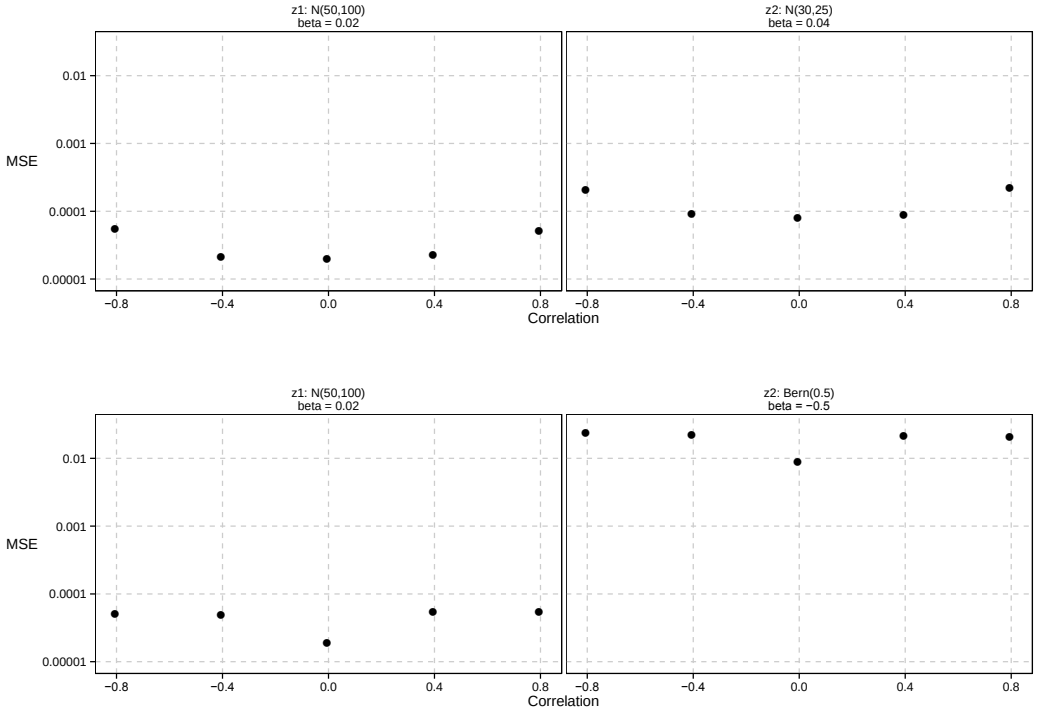

Figure 6. MSE of fitted coefficients by correlation amount between covariates when survival times are generated bounded between 20 and 300 with random censoring and 50% of observations censored when covariates are (a), above, two Normal random variables and (b), below, one Normal and one Bernoulli variable

Additional results in the Supplemental materials are shown with stratification on either censoring type or statistical performance metric to aid visual interpretation. Figures 1 and 2 are subsets of supplemental Figures S1 and S2. Tables with the information contained in Figures 1 to 4 are included in the supplemental material (Tables S1 to S4, respectively).

Discussion

Zhou's method (Zhou, 2001) of generating right-censored outcomes has been implemented by Hendry (2014) using the piecewise exponential framework and allows for an arbitrary number and functional form of the covariates. The main point of this study was to provide concrete recommendations for researchers interested in generating survival data with a specific structure in mind, as in mimicking a motivating data set from a real study. The algorithm proposed by Hendry offers flexibility, but the author did not provide guidance on how to choose parameters that will lead to data with desired features. In particular, one step of the algorithm requires the practitioner to choose an arbitrary monotone increasing function, g, such that $g(0) = 0$, and $g^{-1}(t)$ is differentiable. It was demonstrated that choosing a Weibull distribution for g(.) leads to a simple calculation that allows the practitioner to specify a target median survival time. This recommendation has important implications for practical use because it allows researchers to have much greater control over the generated data.

The simulation results show that, to minimize bias in fitted coefficients and achieve a realistic distribution of survival times, generating data with wider limits are better than keeping the range small even if the target median survival time is low. When generating data to achieve a target median survival time of 35, the standardized bias was high when survival times were generated between 20 and 50, but no meaningful bias was found if the range was expanded to 20-150 or 20-300. It was found, unexpectedly, that when using an overly-restrictive survival time interval with traditional censoring, bias was reduced as the amount of censoring increased (Figure 2). It is generally expected that higher percentages of censoring observations will either increase or have no effect on bias. Here, because the specified range of survival times was too restrictive, when applying traditional censoring we get an inverse relationship between the percentage of observations censored and bias.

This counterintuitive relationship is caused by the survival time generation algorithm's use of resampling to produce only survival times that fall within a specified interval. Consider the set of lower-risk individuals whose covariates

compel them to have an event later than the upper bound of the specified interval. The algorithm will use their data to repeatedly generate survival times until a time is produced that falls within the interval. From a modeling perspective, these lower-risk individuals are indistinguishable from the higher-risk individuals whose covariates compel them to have an event near the boundary of the specified time interval. The lower risk represented by these people's covariates is not reflected in their survival times, and the result is the bias we see with the 20-50 interval.

Traditional censoring affects the lower-risk subjects in a sample; subjects with survival times closer to the upper bound are more likely to be censored under traditional censoring than under random censoring. Subjects with an event near the upper bound of the pre-specified limits can be divided into two groups: 1) subjects with risk consistent with having an event near the boundary, or 2) subjects forced by the algorithm to have an event near the boundary despite their lower-risk covariates. Group 2's survival times are not indicative of the actual risk present in the covariates. Consequently, their inclusion in models estimating associations between covariates and risk results in bias. Traditional censoring removes members of group 2 at a proportionally higher rate than that of subjects whose survival times better reflect actual risk (group 1). The reduction in bias with increased percentages of subjects censored when using traditional censoring is caused by traditional censoring disproportionately removing the bias-causing portion of our sample. This bias in the observed sample produces the observed bias in the fitted coefficients when the range of possible survival times (bounds) is too restrictive. Random censoring targets all subjects equally, thus leaving the bias-producing component of our subjects proportionally intact, and so has a much less pronounced effect on bias.

Given these results, it is recommended that the range be wide enough to generate a distribution with the correct shape, but that it should not be too large to preserve reasonable computational efficiency. Because the algorithm generates survival times as a function of time-varying covariates that vary at integer-valued steps of the time scale, each subject will have as many records as the survival time generated. Thus, a large range means that some subjects will have many records. This is an important computational consideration, especially if generating data is just a first step in a much larger simulation.

There is a fine balance between the survival function, $\exp(\boldsymbol{\beta}\mathbf{Z}_j')$, and the baseline hazard, $h_0(t)$, that will influence the final survival times generated and the computational run times. Some of it can be controlled while defining the g function, which is a key component of the algorithm. It is suggested defining the g

function via a Weibull distribution with parameters informed by an empirical data set. One can look at the distribution of survival times to decide whether the g function should reflect an increasing, decreasing or constant baseline hazard, imposing a value for the shape parameter (v).

Also, the empirical median survival time can be used as the target median in the generated times and use this target median in a formula to compute the scale parameter (λ). It was found that the choice of g function worked well. Statistical performance did not vary by v or target median. However, we did find increased run times with increased target medians. Of note, this approach to generating survival times is still useful even if the observed survival data does not follow a Weibull distribution but the goal of the simulation study is to evaluate the performance of the Cox model. However, it might not be appropriate if, for example, the researcher's aim is to performance a power analysis.

The algorithm performed similarly for both combinations of covariates, but we found lower median survival times in the case of covariates of mixed type compared to the case of two Normal covariates. So, in order to achieve the target median when using covariates of mixed type, the values of the parameters needed to compute the scale parameter might need to be changed iteratively, mainly by inflating the target median, until the distribution of generated times adequately resembles the empirical target distribution.

There were no major issues when covariates were correlated. An increase was noted in the MSE, which was likely associated with an increase in the effect size and not necessarily with the type of covariates being used. Effect sizes play an important role as they have a direct impact on the distribution of the survival times. Further investigation is needed as well as exploring the performance of the algorithm in a scenario where correlations are observed within an individual.

Given the results shown by the simulation study, the following is suggested:

1. The use of the Weibull distribution to define the g function:
   $g = (\lambda^{-1}t)^{1/v}$ and $g^{-1} = \lambda t^{v}$
2. Parameters for the Weibull distribution can be informed from an empirical dataset:
   a. Use the distribution of survival times to decide if the g function should reflect an increasing, decreasing or constant baseline hazard to define the shape parameter (v);
   b. Use the observed median survival times to define a target median (M);

c. For a vector of effect estimates $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ and a vector of means of the covariates $\bar{\mathbf{Z}} = \left(\bar{Z}_1, \ldots, \bar{Z}_p\right)$, the scale parameter λ can be defined as follows:

$$\lambda = \frac{\log 2}{\boldsymbol{\beta}\bar{\mathbf{Z}}'} M^{-\nu}$$

3. Iterate until appropriate values can be found for the survival times. A wider range will yield a higher number of records per subject increasing the computational time. In contrast, a more limited range may introduce bias;
4. Utilize random censoring.

In conclusion, Hendry's algorithm for computing survival times that follow an extended Cox model with time-varying covariates were found to be a reasonable and practical solution when generating studies intended to closely resemble a motivating data set. Guidelines, substantiated by the simulation study, are provided to make this process easier.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Patient-Centered Outcomes Research Institute or the United States government.

Supplemental Material

Supplemental tables and figures are available in the Supplemental Material file, available at https://doi.org/10.22237/jmasm/1493597100.

References

Abrahamowicz, M., MacKenzie, T., & Esdaile, J. M. (1996). Time-dependent hazard ratio: Modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association, 91*(436), 1432-1439. doi: 10.2307/2291569

Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine, 31*(29), 3946-3958. doi: 10.1002/sim.5452

Bakoyannis, G., & Touloumi, G. (2012). Practical methods for competing risks data: A review. *Statistical Methods in Medical Research, 21*(3), 257-272. doi: 10.1177/0962280210394479

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine, 24*(11), 1713-1723. doi: 10.1002/sim.2059

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*(24), 4279-4292. doi: 10.1002/sim.2673

Collett, D. (2003). *Modelling survival data in medical research* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351. doi: 10.1037//1082-989x.6.4.330

Crowther, M. J., & Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine, 32*(23), 4118-4134. doi: 10.1002/sim.5823

Demirtas, H., & Doganay, B. (2012). Simultaneous generation of binary and normal data with specified marginal and association structures. *Journal of Biopharmaceutical Statistics, 22*(2), 223-236. doi: 10.1080/10543406.2010.521874

Desai, M., Bryson, S. W., & Robinson, T. (2013). On the use of robust estimators for standard errors in the presence of clustering when clustering membership is misspecified. *Contemporary Clinical Trials, 34*(2), 248-256. doi: 10.1016/j.cct.2012.11.006

Hendry, D. J. (2014). Data generation for the Cox proportional hazards model with time-dependent covariates: a method for medical researchers. *Statistics in Medicine, 33*(3), 436-454. doi: 10.1002/sim.5945

Leemis, L. M., Shih, L.-H., & Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statistics & Probability Letters, 10*(4), 335-339. doi: 10.1016/0167-7152(90)90052-9

Sylvestre, M. P., & Abrahamowicz, M. (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine, 27*(14), 2618-2634. doi: 10.1002/sim.3092

Zhou, M. (2001). Understanding the Cox regression models with time-change covariates. *The American Statistician, 55*(2), 153-155. doi: 10.1198/000313001750358491