



# HHS Public Access

Author manuscript

*Lifetime Data Anal.* Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

*Lifetime Data Anal.* 2020 January ; 26(1): 1–20. doi:10.1007/s10985-018-9452-5.

## Estimation for an Accelerated Failure Time Model with Intermediate States as Auxiliary Information

**Ritesh Ramchandani,**

Harvard T.H. Chan School of Public Health, FXB, 651 Huntington Ave. 5th floor, Boston, MA 02115, ritesh@mail.harvard.edu

**Dianne Finkelstein,**

Massachusetts General Hospital Biostatistics Center, 50 Staniford St. Suite 560. Boston, MA 02114, dfinkelstein@mgh.harvard.edu

**David Schoenfeld**

Massachusetts General Hospital Biostatistics Center, 50 Staniford St. Suite 560. Boston, MA 02114, dschoenfeld@mgh.harvard.edu

### Abstract

The accelerated failure time (AFT) model is a common method for estimating the effect of a covariate directly on a patient's survival time. In some cases, death is the final (absorbing) state of a progressive multi-state process, however when the survival time for a subject is censored, traditional AFT models ignore the intermediate information from the subject's most recent disease state despite its relevance to the mortality process. We propose a method to estimate an AFT model for survival time to the absorbing state that uses the additional data on intermediate state transition times as auxiliary information when a patient is right censored. The method extends the Gehan AFT estimating equation by conditioning on each patient's censoring time and their disease state at their censoring time. With simulation studies, we demonstrate that the estimator is empirically unbiased, and can improve efficiency over commonly used estimators that ignore the intermediate states.

### Keywords

Survival; AFT; Multi-state model; Semiparametric; Transition probabilities

## 1 Introduction

In many chronic diseases, patients move through a series of progressively worsening disease states until a primary failure such as death. Further, in clinical studies of progressive diseases, we often will not know every subject's failure time because many are lost to follow-up or do not fail within the time period of the study. We may, however, also have information on their disease course recorded up to their last follow-up time. For a clinical study of a progressive disease, we will provide an estimator for the effect on survival time that incorporates the information from these intermediate disease states as auxiliary information in a manner relevant to the primary failure. When there are relatively few observed primary failures in the study, it can be challenging to precisely estimate this effect.

The goal of this paper is to utilize the intermediate states to get a more precise and holistic estimate of the treatment effect.

The proportional hazards model has been used to obtain estimates of a survival treatment effect with auxiliary information, for example in Lu and Tsiatis (2008). While the PH model is useful for testing and hazard ratio estimation, the estimate does not have a direct interpretation in terms of the survival time for a subject. Alternatively, the estimate for an accelerated failure time (AFT) model has the straightforward interpretation of the treatment accelerating (or decelerating) the average time to failure. This makes it an appealing alternative to the proportional hazards model.

The standard semiparametric AFT model relates the covariates to the logarithm of the survival time through the following regression model:

$$\log(T_i) = \beta_0' X_i + \varepsilon_i \quad (1)$$

where  $T_i$  is the failure time for subject  $i$ ,  $\varepsilon_i$  are i.i.d. with unspecified distribution function  $F$ ,  $\beta_0$  is a vector of parameters, and  $X_i$  is a vector of covariates.

Several methods for estimating parameters of the semiparametric AFT model arose from treating the censored data linear rank tests as estimating equations (Prentice 1978; Tsiatis 1990). These linear rank tests include the popular log-rank (Mantel 1966; Cox 1972), Peto-Peto (Peto and Peto 1972), and Gehan (1965) tests. The weighted-log rank test with the Gehan weight has become a particularly attractive estimating function due to properties that make it more practical for model fitting than other methods. Fyngenson and Ritov (1994) showed that this estimating equation is monotone in each component of  $\beta$ , and Jin et al. (2003) developed an algorithm using linear programming to reliably estimate the parameters in multidimensional settings. Further, the Gehan function is amenable to smooth approximations, which allows for computationally simpler parameter and variance estimation (Brown and Wang 2007; Heller 2007).

These estimators for AFT models are based on univariate failure times, so they need to be modified to incorporate the intermediate states. Under the same premise of using linear rank tests as estimating equations, we propose estimating the AFT parameters based on a recent extension of the Gehan test statistic proposed by Ramchandani et al. (2015) that accounts for the observation of intermediate events, such as disease progression, among censored subjects. The test statistic modifies the Gehan test by estimating probabilities for each subject surviving longer than each of the other subjects conditional on their follow-up times and their last observed disease states. These probabilities are estimated using multi-state models, and allow us to compute the expected Wilcoxon ranks of survival for each subject conditional on what we observe. The idea is to obtain more precise parameter estimates by using intermediate disease status as additional information to the usual death and censoring times. This allows us to meaningfully include the intermediate transitions into parameter estimation while not allowing them to dominate the estimator. The key assumption that we have to make in order to obtain interpretable parameter estimates is that each transition of the process, in addition to the total time from origin to the absorbing failure, follows an

accelerated failure time model. The AFT model is a natural one to use in this case because of its straightforward interpretation in terms of linearly accelerating or decelerating a disease process.

In section 2, we will describe the model under which we are operating, and provide a formulation of the proposed estimating equation under the assumption that the probabilities were known. We will then describe the Aalen-Johansen estimator, which we use to estimate the probabilities. We follow by proposing a method for estimating the variance of the parameters based on a Monte Carlo smoothing method given by Jin et al. (2014). In section 3, we describe the simulation studies. We illustrate the method on a clinical trial for Amyotrophic Lateral Sclerosis (ALS) in section 4, and conclude with a discussion in section 5.

## 2 Methods

Suppose  $T_i$  is a failure time, and  $C_i$  the independent right censoring time for subject  $i$ ; let  $Y_i = \min(T_i, C_i)$ ,  $e_i^\beta = \log(Y_i) - \beta'X_i$  (the observed residual), and  $\delta_i = I(T_i \leq C_i), i=1, \dots, n$ . The Fyngenson-Ritov (Gehan) estimating equation for fitting the semiparametric accelerated failure time model is given by:

$$U_G(\beta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_i (X_i - X_j) I(e_i^\beta < e_j^\beta) = 0 \tag{2}$$

With a binary covariate, this equation is simply the Gehan-Wilcoxon test applied to the observed residuals, and counts all the pairs for which we know that  $\log(T_i) - \beta'X_i < \log(T_j) - \beta'X_j$ , i.e. that the failure time residual for one individual is less than the failure time residual for another. However, we can possibly get better precision if all pairs of residuals, whether uncensored or censored, contribute to the statistic in a meaningful way. The idea is to base an estimating equation on the expected scores of  $U_G(\beta)$  conditional on what we observe. Let  $\tilde{e}_i^\beta = \log(T_i) - X_i'\beta$  denote the possibly unobserved failure time residual for individual  $i$ . A straightforward modification to the above estimating equation (2) would be:

$$U_E(\beta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j) P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | e_i^\beta, e_j^\beta, \delta_i, \delta_j) = 0 \tag{3}$$

where  $P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | e_i^\beta, e_j^\beta, \delta_i, \delta_j)$  represents the probability that the failure time residual for subject  $i$  is less than that of subject  $j$ , conditional on each of their residual follow-up times and their failure status. This estimating equation is related to Efron's modification of the Gehan-Wilcoxon test (1967). Another way to think of the conditional probabilities in (3) is in terms of disease states. In the above setting, we are in the simple case of two disease states: alive and dead, with  $\delta_i$  the indicator for the latter. However, if we are in the setting of a chronic disease where individuals pass through multiple states on the way to failure, we can condition the above probabilities on the disease states of the individuals at each of their

last observed times to get a more precise estimate of the model parameters. Examples of this type of intermediate data include cancer stage progression, neurodegeneration from ALS, and Alzheimer’s disease transitioning from mild to severe. This can be an especially useful extension for studies with relatively low failure rates over long periods of time.

To develop this idea more precisely, suppose individuals move through a finite set of states  $S = \{0, 1, 2, \dots, D\}$  governed by a progressive multi-state process, where 0 is the initial state,  $D$  represents the single absorbing state, and that transitions to each state are observed exactly. We will consider progressive models of the forms given in Figure 1, and assume that the structure of the model is known and correctly specified. Let  $S_i(t)$  denote the state of individual  $i$  at time  $t$ , let  $T_{i,gh}$  denote the random variable for the transition time for individual  $i$  from state  $g$  to state  $h$ , where  $g \in \{0, 1, \dots, D-1\}$ ,  $h \in \{1, 2, \dots, D\}$ , and  $h > g$ . Let  $T_i$  be the absorbing failure time for individual  $i$  (i.e the time from origin to the absorbing state) We will assume that each transition from one state to another follows an AFT model, i.e.  $\log(T_{i,gh}) = \beta'_{gh} X_i + \varepsilon_{i,gh}$ , where the  $\varepsilon_{i,gh}$  are i.i.d. with unspecified distribution function and independent of  $X_i$ . If the absorbing failure time from origin (time 0) follows the AFT model  $\log(T_i) = \beta'_0 X_i + \varepsilon_i$ , where  $\varepsilon_i$  are i.i.d. with unspecified distribution function and independent of  $X_i$ , it follows that  $\beta_{gh} = \beta_0$  for models of the forms given in Figure 1 (proof in Appendix A. 1).

Additionally, we let  $C_i$  be a censoring random variable independent of the multistate process,  $Y_i = \min(T_i, C_i)$ , and  $\delta_i = I(T_i < C_i)$ . Let  $e_i^\beta = \log(Y_i) - \beta' X_i$ , and  $\tilde{e}_i^\beta = \log(T_i) - \beta' X_i$ . Under the model described, a reasonable estimating equation for  $\beta$  is:

$$U_P(\beta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j) P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) = 0 \tag{4}$$

where  $P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta))$  represents the probability that the failure time residual for individual  $i$  will be less than the failure time residual for individual  $j$  conditional on each of their observed disease states at their observed follow-up time residual. This extension of the Gehan estimating equation is based on the extension of Gehan’s test statistic proposed by Ramchandani et al. (2015) to account for intermediate disease state information. At  $\beta = \beta_0$ , this equation is centered around 0 under the true, unknown, probability measure. (See Appendix A.2). The estimating equation in (4) can also be written as:

$$U_P(\beta) = \frac{1}{n^2} \sum_i \sum_{j < i} (X_i - X_j) [P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - P(\tilde{e}_i^\beta > \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta))] = 0 \tag{5}$$

This formulation of  $U_P(\beta)$  can be identified as an order 2 U-statistic, thus giving us asymptotic normality of the score function at a fixed  $\beta$ , and providing a way of computing

the covariance matrix of  $\sqrt{n}U_P(\beta)$ . Let  $D(\beta)$  denote the asymptotic covariance matrix of  $\sqrt{n}U_P(\beta)$ . From standard U-statistics theory,  $D(\beta)$  has elements that can be estimated with:

$$D_{l,m}(\beta) = \frac{1}{n^3} \sum_i \sum_j \sum_{k \neq j} (X_{il} - X_{jl})(X_{im} - X_{km}) \phi_{ij}^\beta \phi_{ik}^\beta \tag{6}$$

where  $\phi_{ij}^\beta = P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - P(\tilde{e}_i^\beta > \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta))$  (Van der Vaart 2000).

When we know that  $\log(T_i) - \beta'X_i < \log(T_j) - \beta'X_j$ , the conditional probability in the summand of (4) is known and equal to 1, just as in the Gehan estimating equation,  $U_G(\beta)$ . This would happen in the scenario where both subjects  $i$  and  $j$  had reached the absorbing state, or if subject  $i$  reached the absorbing state and their residual time to the absorbing state was less than the residual time of subject  $j$ . It follows that we can rewrite  $U_P(\beta)$  as the sum of  $U_G(\beta)$  and an additional term of probabilities for censored subjects:

$$\begin{aligned} U_P(\beta) &= \sum_i \sum_j (X_i - X_j) I(e_i^\beta < e_j^\beta) \\ &\quad \left\{ \delta_i + (1 - \delta_i) \left[ P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - P(\tilde{e}_i^\beta > \tilde{e}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) \right] \right\} \tag{7} \\ &= 0 \end{aligned}$$

The summand of the estimating equation  $U_P(\beta)$  is based on the true probabilities, but in practice we have to estimate the probabilities. This can be done in a number of ways using event history models that account for incomplete observation. In this paper, we will estimate the probabilities nonparametrically using the Aalen-Johansen estimator.

### 2.1 The Aalen-Johansen Estimator

To estimate the failure probabilities, we propose using the empirical transition matrix developed by Aalen and Johansen (1978), fit on the residuals of each transition time. The Aalen-Johansen estimator is a natural generalization of the Kaplan-Meier estimator for non-homogeneous Markov chains with a finite number of states (Aalen et al 2008). Suppose we have a finite number of states  $S = \{0, 1, \dots, D\}$ . Let  $\alpha_{g,h}(t)$  denote the transition intensity from state  $g$  to state  $h$ , where  $g \neq h$ . This describes the instantaneous risk, or the hazard, of transitioning from state  $g$  to state  $h$  at time  $t$ . Now, let  $P_{g,h}(s,t)$  denote the probability of a subject being in state  $h$  at time  $t$  given that the subject was in state  $g$  at time  $s$ . This is called a transition probability, and it is the  $g,h$  entry of the  $d \times d$  transition probability matrix  $\mathbf{P}(s,t)$ . The transition probability matrix can be written as a function of the transition intensities through the product integral:

$$\mathbf{P}(s,t) = \prod_{(s,t]} [\mathbf{I} + d\mathbf{A}(u)]$$

where  $I$  is the identity matrix, and  $\mathbf{A}(u)$  is the cumulative transition intensity matrix with elements  $A_{gh}(t) = \int_0^t \alpha_{gh}(u) du$ . Let  $N_{gh}(t)$  be the number of individuals observed to

experience a transition from state  $g$  to  $h$  between time 0 and time  $t$ , and let  $Y_g(t)$  be the number of individuals in state  $g$  just before time  $t$ . For  $g = h$ , we can use the Nelson-Aalen estimator to estimate the cumulative hazard  $A_{gh}(t)$ , yielding  $\hat{A}_{gh}(t) = \int_0^t \frac{dN_{gh}(u)}{Y_g(u)}$ , which can also be written as  $\sum_{u \leq t} \frac{\Delta N_{gh}(u)}{Y_g(u)}$ , where  $N_{gh}(u)$  represents the number of transitions from state  $g$  to state  $h$  at time  $u$ . Also, we let  $\hat{A}_{gg}(t) = -\sum_{h \neq g} \hat{A}_{gh}(t)$ , so that the rows of the  $d \times d$  matrix  $\hat{A}(t)$  sum to 0. Suppose  $u_1 < u_2 < \dots$  are the exact times when a transition between any two states are observed. Then the estimate for  $P(s, t)$  is given by the matrix product:

$$\hat{P}(s, t) = \prod_{s < u_j \leq t} [I + d\hat{A}(u_j)].$$

where  $d\hat{A}(u_j) = \hat{A}(u_j) - \hat{A}(u_{j-1})$ , and we define  $u_0 = 0$ . For example, in a progressive three state model, this expression written out in full using the counting process notation described above would be:

$$\hat{P}(s, t) = \prod_{s < u_j \leq t} \begin{pmatrix} 1 - \frac{\Delta N_{01}(u_j)}{Y_0(u_j)} & \frac{\Delta N_{01}(u_j)}{Y_0(u_j)} & 0 \\ 0 & 1 - \frac{\Delta N_{12}(u_j)}{Y_1(u_j)} & \frac{\Delta N_{12}(u_j)}{Y_1(u_j)} \\ 0 & 0 & 1 \end{pmatrix}$$

In the presence of censoring, these transition probabilities will be used to estimate the probability of an individual's lifetime being less than another individual's, on the scale of the failure time residual,  $\tilde{e}^\beta = \log(T) - \beta'X$ . The rationale for estimating the transition probabilities based on the residuals is that, under our assumed model, the trajectory of each patient based on their residual transition times is identically distributed at the true  $\beta_0$ . There are several statistical packages that allow for the computation of the Aalen-Johansen estimator. One excellent option is the *etm* package in R (Allignol et al 2011; R Core Team 2014).

### 2.2 The Estimating Equation

For the estimating equation  $U_{\mu}(\beta)$ , when comparing two subjects, we have two scenarios where we would need to estimate a probability: when subject  $i$  is censored and  $j$  is uncensored, and when both are censored. In the first case, suppose subject  $i$  is censored in state  $k$  and  $j$  is uncensored. Then we can estimate

$$\left[ P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta) = k, S_j(e_j^\beta) = d) - P(\tilde{e}_i^\beta > \tilde{e}_j^\beta | S_i(e_i^\beta) = k, S_j(e_j^\beta) = d) \right] \text{ with}$$

$$\left[ \hat{P}_{kd}(e_i^\beta, e_j^\beta -) + \hat{P}_{kd}(e_i^\beta, e_j^\beta) - 1 \right] \text{ where } t- \text{ indicates a time just before time } t. \text{ Now suppose that}$$

subject  $i$  is censored in state  $k$ , subject  $j$  is censored in state  $k'$  and  $e_i^\beta < e_j^\beta$ . Then  $P(\tilde{e}_i^\beta < \tilde{e}_j^\beta | S_i(e_i^\beta) = k, S_j(e_j^\beta) = k') - P(\tilde{e}_i^\beta > \tilde{e}_j^\beta | S_i(e_i^\beta) = k, S_j(e_j^\beta) = k')$  can be estimated with:

$$\int_{e_i^\beta}^{\infty} [1 - \hat{P}_{k',d}(e_j^\beta, t)] d\hat{P}_{k,d}(e_i^\beta, t) - \int_{e_j^\beta}^{\infty} [1 - \hat{P}_{k,d}(e_i^\beta, t)] d\hat{P}_{k',d}(e_j^\beta, t) \tag{8}$$

where by convention we define  $\hat{P}(e_i, t) = 0$  for  $t < e_i$ . Note that these expressions are general in the sense that we can use them for any multi-state models where we estimate transition probabilities. In the case of the Aalen-Johansen estimator, the probabilities are step-functions, so in practice equation (8) is computed with sums. Denote the maximum follow-up residual time as  $e_{max}^\beta$  and let  $t_1, t_2, \dots$  be the jumps in  $\hat{P}(s, t)$  for any fixed  $s$ . We can compute (8) as:

$$\sum_{l: e_i^\beta < t_l \leq e_{max}^\beta} [1 - \hat{P}_{k',d}(e_j^\beta, t_l)] [\hat{P}_{k,d}(e_i, t_l) - \hat{P}_{k,d}(e_i, t_{l-1})] - \sum_{l: e_j^\beta < t_l \leq e_{max}^\beta} [1 - \hat{P}_{k,d}(e_i, t_l)] [\hat{P}_{k',d}(e_j, t_l) - \hat{P}_{k',d}(e_j, t_{l-1})]$$

We now denote the estimating equation  $U_P(\beta)$  as  $U_P(\beta; \hat{\mathbf{A}})$  to indicate that the equation depends on the estimated cumulative transition hazard matrices  $\hat{\mathbf{A}}(\cdot)$ . The estimating equation can now be written as:

$$U_P(\beta; \hat{\mathbf{A}}) = \frac{1}{n^2} \sum_i \sum_j (X_i - X_j) I(e_i^\beta < e_j^\beta) \left\{ \delta_i + (1 - \delta_i) \delta_j \left[ \hat{P}_{S_i(e_i^\beta), d}(e_i^\beta, e_j^\beta -) + \hat{P}_{S_i(e_i^\beta), d}(e_i^\beta, e_j^\beta) - 1 \right] + (1 - \delta_i)(1 - \delta_j) \left[ \int_{e_i^\beta}^{\infty} [1 - \hat{P}_{S_j(e_j^\beta), d}(e_j^\beta, t)] d\hat{P}_{S_i(e_i^\beta), d}(e_i^\beta, t) - \int_{e_j^\beta}^{\infty} [1 - \hat{P}_{S_i(e_i^\beta), d}(e_i^\beta, t)] d\hat{P}_{S_j(e_j^\beta), d}(e_j^\beta, t) \right] \right\} \tag{9}$$

The estimate  $\hat{\beta}$  is the value of  $\beta$  where  $U_P(\beta; \hat{\mathbf{A}})$  crosses 0.

In the simple two-state model, this estimator is similar to the Peto-Prentice version of the weighted log-rank estimator. Note that by using the Aalen-Johansen estimator for the probabilities, we are additionally making the assumption that the error terms for the multi-state process arise from a non-homogeneous Markov process. However, the method is more general as the probabilities can be estimated in other ways as well, including parametrically.

Alternative estimates for the probabilities may be used if one wants to relax the Markov assumption, such as those proposed by de Uña-Álvarez and Meira-Machado (2015), and Meira-Machado et al. (2006). Nevertheless, simulations suggest that the proposed estimator works well in some non-Markov settings as well.

### 2.3 Remark on estimation and asymptotic properties

As described in Section 2, the estimating function  $U_P(\beta)$  is a U-statistic when we know the true conditional probabilities in the summand of the statistic of equation (5). Those conditional probabilities are based on the true cumulative hazard process  $\mathbf{A}(\cdot)$  described in section 2.1. In practice we do not know the true hazard process, and we propose estimating it to obtain the necessary transition probabilities. By estimating this process, the estimating equation is no longer strictly a U-statistic, so the usual properties of U-statistics do not directly apply. While this may have implications on the convergence and asymptotic variance of the estimator, our simulation studies suggest that the estimator as presented is unbiased and that the variance estimator described in the next section is close to the empirical variance.

Additionally, it is clear that the proposed estimating equation is neither continuous nor monotone in  $\beta$ . This is not a major problem when there is a single covariate, but for multidimensional settings, it can make estimation of  $\beta$  difficult and admits the possibility of multiple solutions. In these settings, we could first find a consistent auxiliary estimator, such as the Gehan estimator that is obtained with linear programming as described by Jin et al. (2003). We would then solve for  $\beta$  as the minimizer of the norm  $\|U_P(\beta; \hat{A})\|$  using a derivative-free optimization algorithm, such as Nelder-Mead (1965), with the consistent auxiliary estimator as an initial value to arrive at a solution in the correct neighborhood of  $\beta_0$ . We encourage use of various starting values to ensure that the estimate obtained is the global minimizer.

While it is not ensured that a solution will exist in finite samples, at least one solution should exist provided that the following conditions apply:

1. If there are  $p$  covariates, there are at least  $p + 1$  failures to the absorbing state.
2. The distribution of covariates among those who have had an absorbing failure are not concentrated on a hyperplane of the dimension of the parameter space; i.e. if the covariate vectors among subjects who had an absorbing failure are linearly dependent, a  $p$ -dimensional solution may not exist.
3. Covariates are bounded and have finite variance.
4. Densities of the sojourn and absorbing failure times are bounded and have finite variance.

### 2.4 Inference Procedure for $\hat{\beta}$

It is well known that variance estimation for the parameters of the ordinary semiparametric accelerated failure time model is difficult. This is because the estimating equations are non-smooth, and the usual sandwich variance estimate involves the derivative of the unknown



hazard function of the error terms. For the general weightedlog rank estimating functions, it has been established that the covariance matrix for  $\sqrt{n}(\hat{\beta} - \beta_0)$  is given by  $\mathbf{V}=\mathbf{B}^{-1}\mathbf{DB}^{-1T}$ , where  $\mathbf{B}$  is the non-singular slope matrix of the estimating function  $U(\beta)$  and  $\mathbf{D}$  is the variance of the score function, each evaluated at  $\beta_0$  (Kalbfleisch and Prentice 2011). Estimation of  $\mathbf{D}$  is straightforward, but the discontinuities of the estimating equation do not allow for direct computation of  $\mathbf{B}$  using derivatives; further, direct numerical differentiation can be unstable in practice. This will similarly be the case for our estimating function, where we may estimate  $\mathbf{D}$  using the asymptotic variance formula for the U-statistic with the resubstituted probability estimates, but where an estimate of  $\mathbf{B}$  is difficult to obtain due to the discontinuity of the estimating equation in finite samples.

In light of these issues, some authors have pursued a smooth approximation of the Gehan estimating equation to allow for straightforward parameter and variance estimation (Brown and Wang 2007; Heller 2007). While this approach works well for the Gehan estimating equation, it is not straightforward to obtain smooth versions for other AFT estimators. To accommodate other types of estimators, Jin et al. (2014) proposed a Monte Carlo smoothing method based on the approach of Brown and Wang (2007) for estimating standard errors.

The idea is that in large samples, the distribution of  $\hat{\beta} - \beta_0$  is approximately equal to  $n^{-1/2}\mathbf{V}^{1/2}\mathbf{Z}$ , where  $\mathbf{Z}$  is a standard normal random vector and  $\mathbf{V}$  is the covariance matrix of  $\sqrt{n}(\hat{\beta} - \beta_0)$ . This induces the smooth estimating function  $\tilde{U}(V, \beta) = E_{\mathbf{Z}}[U(\beta + n^{-1/2}\mathbf{V}^{1/2}\mathbf{Z})]$ , where the expectation is taken with respect to the vector  $\mathbf{Z}$ . They then argue that the derivative of the smooth function  $\tilde{U}(V, \beta)$  is given by:

$$B(\Gamma; \beta) = E_{\mathbf{Z}}[U(\beta + \Gamma\mathbf{Z})\mathbf{Z}^T\Gamma^{-1}] \tag{10}$$

where  $\Gamma=n^{-1/2}\mathbf{V}^{1/2}$ . They propose an iterative method where they numerically approximate the integral in (10), update  $\Gamma$ (the current estimate for the standard error for  $\hat{\beta}$ ) and iterate through until convergence of  $\Gamma$ . One approach is a Monte Carlo method (MCM), with which perturbed score equations must be evaluated a very large number of times. This is an intuitive and simple way to estimate  $\mathbf{B}$ , but it requires a very large number of function evaluations at each step of the algorithm. The other approach is a Gaussian Quadrature Method (GQM), which we choose to implement here because it can be far more computationally efficient while giving similar results to the MCM. The idea is to choose a p-dimensional grid of nodes based on onedimensional Gauss-Hermite quadrature, obtain their corresponding weights, and use those to approximate the integral in (10). We describe the algorithm in Appendix A.3. Confidence intervals for  $\hat{\beta}$  can be obtained by the Wald method.

An alternative to this method would be to use a bootstrap approach for estimating the variance of  $\hat{\beta}$ . The classical bootstrap would entail resampling subjects' entire trajectory with replacement, reestimating the requisite probabilities using the Aalen-Johansen estimator, and obtaining an estimate of  $\hat{\beta}^*$  that solves the estimating equation based on the new sample. This process would be repeated a large number of times  $M$ , with standard errors computed from the empirical distribution of  $\hat{\beta}^* = (\hat{\beta}_1^*, \dots, \hat{\beta}_M^*)^T$ . Confidence intervals for  $\beta_0$

can be obtained either with the Wald method, or directly from the empirical distribution of  $\hat{\beta}^*$ .

### 3 Simulations

To test the performance of our estimator, we simulated data from a 3-state progressive multi-state model of the form  $0 \rightarrow 1 \rightarrow 2$  (where 2 is the absorbing state), such that the acceleration parameter acts on the entire process. Let  $T_{ik}$  represent the time taken to transition from state  $k-1$  to state  $k$ . We generated the sojourn times  $\log(T_{ik}) = 2 + \beta_0 X_i + \varepsilon_{ik}$ , for  $k=1,2, i=1, \dots, n$ . Clearly, the absorbing state failure time  $T_i = T_{i1} + T_{i2}$  satisfies  $\log(T_i) = 2 + \beta_0 X_i + \varepsilon_i$ . We set  $\beta_0 = 0.7$ , which corresponds approximately to a 2-fold acceleration of the failure time for a unit difference in the covariate  $X$ . This was done for various choices of  $\varepsilon_{ik}$ , including distributions for which the Markov assumption does not hold. In one setting, the  $\varepsilon_{ik}$  were independent of each other, and had either standard extreme-value (log-Weibull), standard normal, standard logistic distributions. In another setting we allowed the  $\varepsilon_{ik}$  to be correlated, with standard multivariate normal distributions with either correlation  $\rho = 0.5$  and  $\rho = 0.9$ . It should be noted that these are the distributions of the *state sojourn times* and not the distributions of the absorbing failure times. The covariate  $X_i$  was normally distributed with mean 0 and standard deviation 0.5 in all settings. Censoring values were generated from a Uniform(0,  $\tau$ ), with  $\tau$  chosen to yield a desired level of censoring. In each setting, we also allowed censoring to depend on the covariate, with  $C_i$  distributed as  $\exp(1.5 X_i) \cdot \text{Uniform}(0, \tau)$ .

We computed the bias, empirical standard error, and empirical MSE for the Fyngenson-Ritov (Gehan), the Peto-Prentice, the Log-rank, and the Proposed estimators. For the proposed estimator, we also computed standard error estimates, 95% coverage probabilities based on Wald confidence intervals, and relative efficiencies of the proposed estimator compared to the Gehan, Peto-Prentice, and Log-rank estimators. The variance of the score equation was obtained using equation (6) with the resubstituted probability estimates. Standard error estimates for the proposed estimator were obtained using the GQM method with 16 Gauss-Hermite quadrature nodes, and a tolerance level of  $10^{-4}$  for convergence of  $\Gamma$ . 1000 simulations were used in each setting, with sample sizes of 100 and 200.

Recall that the Gehan estimating function is given in equation (2). The Log-rank and Peto-Prentice estimating functions are given by:

$$\sum_i w_i \delta_i \left( x_i - \frac{\sum_j x_j I(e_i^\beta \leq e_j^\beta)}{\sum_j I(e_i^\beta \leq e_j^\beta)} \right)$$

where for the Log-rank estimator  $w_j = 1$ , and for the Peto-Prentice estimator  $\hat{F}(e_i^\beta)$ , where  $\hat{F}(\cdot)$  denotes the left-continuous Kaplan-Meier estimator based on the observed residuals.

The results are given in Tables 1 and 2. Table 1 refers to the setting where the censoring distributions are independent of the covariate, while Table 2 refers to the unequal censoring

case. Observe that in all settings, the proposed estimator is essentially unbiased, the average of the standard error estimator is close to the empirical standard error, and the coverage probabilities are close to the nominal level of 0.95. In addition, the proposed estimator is more efficient than the Gehan, Peto-Prentice, and Log-rank estimators in each of these settings, with the most efficiency gains coming in cases of high censoring. It is not expected that in finite samples the proposed estimator will always be more efficient, but these simulations demonstrate the potential efficiency gains we can get when the intermediate states are taken into account as auxiliary information.

## 4 Example

We will illustrate the proposed method on data from a clinical trial of patients with ALS (Berry et al 2013). Subjects in the trial were monitored for survival, and rate of decline in neurological function as measured by their ALSFRS-R scores. The ALSFRS-R is a functional rating scale by which physicians estimate the degree of functional impairment in ALS patients (Cedarbaum et al 1999). The scale ranges from 0–48, with a higher score indicating better function. We are interested in estimating the effect of treatment on survival, using ALSFRS-R score as the intermediate information. ALSFRS-R was measured periodically in patients until death, drop-out, or the end of the study. We discretized this score into 3 states: 33–48 (state 1), 17–32 (2), 0–16 (3). We assume the transition time occurs when a transition is observed, and we allowed all forward transitions that were seen in the data, but no backward transitions. This means that even if someone actually moved from state 2 to 1 for example, that they were kept in state 2 for the analysis. There were a total of 513 subjects in the analysis, an average follow-up time of 1.5 years, a maximum follow up time of 5.5 years, and 43% of all subjects were censored. It is known that site of disease onset is associated with survival, so we choose to include this covariate in the model.

We estimated coefficients for the model  $\log T_i = \beta_{trt} \times treatment + \beta_{site} \times site + \varepsilon_i$ , where  $treatment = 1$  for “active” and 0 for “placebo”, and  $site$  indicates site of disease onset (1 for bulbar-onset, 0 for limb-onset). We first estimated the coefficients using the Gehan estimating equations. The Gehan estimators were (.217, -.350) for treatment and site of onset, respectively. We then estimated the coefficients using the proposed estimating equation given in (9). This was done using the *optim* function in *R*, with the Nelder-Mead method and using the Gehan estimates as initial values. The coefficients for the proposed estimator were (.210, -.383). This implies that average progression and survival time among the treated group, adjusted for site of onset, was estimated to be  $\exp(.21) = 1.23$  times that of the placebo group. Similarly, adjusting for treatment, average progression and survival time in the bulbar-onset group was 0.68 times that in the limb-onset group.

Standard errors were estimated using the GQM method described in section 2.3 and Appendix A.3, and the bootstrap. Using the formula in (6), the covariance matrix of the score equation,  $D$ , was estimated to be:

$$\begin{pmatrix} .063 & .001 \\ .001 & .046 \end{pmatrix}.$$

For the GQM, we used 6 Gauss-Hermite quadrature nodes, given by the values  $z = \pm(2.35, 1.33, 0.436)$ , with requisite weights  $w = (.0045, .157, .725)$ . The transformed nodes  $z^* = \sqrt{2}z$  were used in order to approximate the desired integral in equation (10). An illustration of the grid of points over which we approximate the integral is given in Figure 2.

The algorithm converged in 4 iterations within a .0001 tolerance level for each entry of the estimated covariance matrix  $G$ . Standard error estimates of the coefficients for treatment and site of onset were .144 and .173, and p-values based on Wald test statistics were .145 and .026, respectively. We also estimated standard errors using the bootstrap, yielding standard error estimates of .145 and .159, with Wald p-values given by .148 and .016, respectively. We conclude that treatment, adjusted for site of onset, is not significantly associated with progression and survival when adjusted for disease site of onset, but that bulbar site of onset of is associated with earlier progression and failure, resulting in almost two-thirds the average survival time of patients whose site of onset was in a limb. These results were not unexpected given the difficulty of treating ALS, and that site of onset is established as prognostic of survival.

## 5 Discussion

While the asymptotic properties are not fully developed, simulations have demonstrated that the proposed estimator and the corresponding standard error estimator have good finite-sample properties in several settings. The estimators are close to their empirical values under semi-Markov sojourn time distributions, correlated sojourn time distributions (non-Markov), and when the censoring distribution depends on the covariates.

In most settings, the proposed estimator was more efficient than those obtained with the Gehan and Peto-Prentice estimators that ignore intermediate events. The improvement in efficiency will depend on the sojourn time distributions and the censoring distributions, with the most improvement in settings where there is very high censoring. Thus, the method of estimation can be particularly useful for shorter studies where the main event of interest is rarely observed, but subjects are monitored frequently for intermediate “benchmarks” as well. An example of this would be any relatively short clinical trial of a chronic disease such as ALS.

A key assumption for the proposed estimator is that the acceleration parameters act on every transition of the process. This is a stronger assumption than the ordinary accelerated failure time model for two states, but a necessary one to ensure that the AFT parameters we estimate are interpretable as such. Thus, it would be useful to devise a procedure to check if the AFT model holds in the manner specified. One potential way would be to treat the time from origin to state  $k$  as a failure time, and use the Gehan estimating equation to estimate  $\beta_k$ , for each non-initial state  $k = 1, \dots, D$ . We could then construct a test for  $H_0: \beta_j = \beta_k, j \neq k$ , using the method proposed by Lin and Wei (1992). If one was instead interested in estimating AFT parameters for each particular state's sojourn time, Huang's accelerated sojourn times model (2002) is the appropriate choice.

Additionally, under our assumed model, there may be other more efficient ways of estimating the desired parameters, such as in the framework of clustered or multivariate failure times (Johnson and Strawderman 2009; Chiou et al 2014). Other estimators could be proposed where each transition time for every participant contributes to the estimation, however, methods that may directly incorporate all intermediate transition times into estimation are somewhat different than what we are proposing. We are essentially treating the intermediate failures as auxiliary information that informs the primary failure of interest, the absorbing state. The absorbing failures still drive the proposed estimator, with some additional information gleaned from the intermediate disease states. Under the assumed model, more emphasis on the intermediate transitions can certainly make more efficient use of all of the observed data, but it was our desire to have an estimator driven primarily by survival that also incorporated the intermediate information in a manner directly relevant to the survival outcome. This also makes the proposed estimator more robust to intermediate transitions departing from the assumed AFT model.

To give more weight to the intermediate transitions, we can possibly estimate parameters from Huang's (2002) accelerated sojourn times model, and use a weighted combination of those model estimates and the ordinary semiparametric AFT estimate based on the survival time. For example, suppose we have a four-state progressive model with an absorbing state and two intermediate states. The weighted estimator could be given by  $w_1\beta_{H1} + w_2\beta_{H2} + (1-w_1-w_2)\beta_G$  where  $w_1 + w_2 = 1$ ,  $\beta_{H,k}$  represents Huang's model estimates for the  $k^{th}$  sojourn time, and  $\beta_G$  represents the Gehan estimator for overall survival. The weights could be specified based on some combination of clinical input, standard error of the estimates, and a measure of model fit. Huang's proposed goodness of fit test restricts the amount of follow-up data used by creating an artificial censoring time at various points of follow-up. Then new estimates for the parameter can be computed and compared with the estimate based on the entire follow-up time. Weights may be based on the relative level of variation between the restricted follow-up estimates and the original. Estimating the standard error for such an estimator will involve estimating the covariance between the components and could prove difficult, but a resampling procedure may be viable. Studying this weighted approach and the GEE approach proposed by Chiou (2014) in the multi-state context may be an avenue of further research.

The proposed estimating equation does not have the same desirable property of monotonicity as the Gehan estimating equation, but its close relationship with the Gehan function can make parameter estimation feasible in practical settings with sufficient sample size. In order to simplify parameter and standard error estimation, an induced smoothing approach may also work well with the proposed estimator, but such an approach would involve smoothing both the indicator functions and the transition probability estimates of the estimating function. Aalen and Johansen (1978) provide an asymptotically equivalent smooth version of their estimator that could be used for this purpose.

The asymptotic properties of the estimator need to be explored in greater detail. As with the traditional censored linear rank estimators, the key result is to establish asymptotic linearity of the score function in a neighborhood of  $\beta_0$ , from which consistency and asymptotic

normality of the estimate  $\hat{\beta}$  typically follow. Our simulation studies suggest this to be the case under certain assumptions, but it remains to be formally established.

## Acknowledgements

Thanks to Rebecca Betensky for her valuable feedback as a dissertation committee member. This work was supported by the Harvard Clinical and Translational Science Center under NIH National Center for Research Resources grant UL1 RR025758, and the Statistical and Data Management Center of the AIDS Clinical Trials Group under the NIH National Institute of Allergy and Infectious Diseases grant UM1 AI068634.

## A Appendix

### A.1 Proof that $\beta$ parameters for each transition are equal when individual transitions and overall survival time follow an AFT model

Consider a progressive model of  $D + 1$  states as in Figure 1a. Let  $T_{gh}$  indicate a continuous random variable for time to transition from state  $g$  to state  $h$ . Let  $T = T_{01} + T_{12} + \dots + T_{D-1,D}$ . Further, suppose  $T = \exp(\beta_0 X + \varepsilon)$ , where  $\varepsilon$  is independent of  $X$ . Let  $T_{g-1,g} = \exp(\beta_g X + \varepsilon_g)$  for  $g = 1, \dots, D$ , where  $\varepsilon_g$  is also independent of  $X$ . Then we have:

$$\begin{aligned} T &= \exp(\beta_0 X + \varepsilon) = \exp(\beta_1 X + \varepsilon_1) + \dots + \exp(\beta_D X + \varepsilon_D) \\ &= \exp(\beta_0 X) \frac{\exp(\beta_1 X)}{\exp(\beta_0 X)} \exp(\varepsilon_1) + \dots + \exp(\beta_0 X) \frac{\exp(\beta_D X)}{\exp(\beta_0 X)} \exp(\varepsilon_D) \\ &= \exp(\beta_0 X) \{ \exp[(\beta_1 - \beta_0)X + \varepsilon_1] + \dots + \exp[(\beta_D - \beta_0)X + \varepsilon_D] \} \end{aligned}$$

Since  $T = \exp(\beta_0 X) \exp(\varepsilon)$ , It follows that

$\varepsilon = \log \{ \exp[(\beta_1 - \beta_0)X + \varepsilon_1] + \dots + \exp[(\beta_D - \beta_0)X + \varepsilon_D] \}$ . Without loss of generality, suppose that  $\beta_1 = \beta_0$ . Then we have that  $\varepsilon = \log \{ \exp[cX + \varepsilon_1] + \dots + \exp[(\beta_D - \beta_0)X + \varepsilon_D] \}$ , where  $c$  is a non-zero constant. In this case,  $\varepsilon$  is not independent of  $X$  and cannot be independent of  $X$  unless  $\varepsilon_1$  is not independent of  $X$ . But by our model assumptions,  $\varepsilon$  and  $\varepsilon_g$  are independent of  $X$  for  $g = 1, \dots, D$ , so it must be the case that  $\beta_1 = 0$ . Similarly, we will have that  $\beta_2 = \dots = \beta_D = \beta_0$ .

A similar argument can be used for progressive models that have the form in Figure 1b. We will show the case of the progressive illness-death model (Figure 1c), but the proof is analogous for models with a larger state space. Suppose we have a 3-state model where subjects can transition from state  $0 \rightarrow 1$ ,  $1 \rightarrow 2$ , and  $0 \rightarrow 2$ , with state 2 as the absorbing state. As before, let  $T_{gh}$  denote the random variable for the direct transition from state  $g$  to state  $h$ , let  $T$  denote the absorbing failure time from origin, and assume  $T = \exp(\beta_0 X + \varepsilon)$ , where  $\varepsilon$  is independent of  $X$ . Let  $T_{gh} = \exp(\beta_{gh} X + \varepsilon_{gh})$  for  $g = 0, 1$ ,  $h = 1, 2$ , and  $h > g$ , where  $\varepsilon_{gh}$  is also independent of  $X$ . We have that

$$\begin{aligned}
 T &= 1[\exp(\beta_{01}X + \varepsilon_{01}) \leq \exp(\beta_{02}X + \varepsilon_{02})][\exp(\beta_{01}X + \varepsilon_{01}) + \exp(\beta_{12}X + \varepsilon_{12})] \\
 &\quad + I[\exp(\beta_{01}X + \varepsilon_{01}) > \exp(\beta_{02}X + \varepsilon_{02})][\exp(\beta_{02}X + \varepsilon_{02})] \\
 &= \exp(\beta_0 X) I[\exp(\beta_{01}X + \varepsilon_{01}) \leq \exp(\beta_{02}X + \varepsilon_{02})] \\
 &\quad \{ \exp[(\beta_{01} - \beta_0)X + \varepsilon_{01}] (\beta_{12} - \beta_0)X + \varepsilon_{12} \} \\
 &\quad + \exp(\beta_0 X) I[\exp(\beta_{01}X + \varepsilon_{01}) > \exp(\beta_{02}X + \varepsilon_{02})] \{ \exp[(\beta_{02} - \beta_0)X + \varepsilon_{02}] \}
 \end{aligned} \tag{11}$$

Since  $T = \exp(\beta_{01}X + \varepsilon)$ , it follows that

$$\begin{aligned}
 \varepsilon &= \log \{ I[\exp(\beta_{01}X + \varepsilon_{01}) < \exp(\beta_{02}X + \varepsilon_{02})] \{ \exp[(\beta_{01} - \beta_0)X + \varepsilon_{01}] + \exp[(\beta_{12} - \beta_0)X + \varepsilon_{12}] \} \\
 &\quad + I[\exp(\beta_{01}X + \varepsilon_{01}) > \exp(\beta_{02}X + \varepsilon_{02})] \{ \exp[(\beta_{02} + \beta_0)X + \varepsilon_{02}] \} \}
 \end{aligned}$$

In order for the indicator functions to be independent of  $X$ , we would need  $\beta_{01} = \beta_{02}$ , and in order for the non-indicator terms to be independent of  $X$ , we need  $\beta_{01} = \beta_{12} = \beta_{02} = \beta_0$ . If at least one of  $\beta_{01}, \beta_{12}, \beta_{02}$  are not equal to  $\beta_0$ , then  $\varepsilon$  is not independent of  $X$ , which contradicts our model assumption.

## A.2 Justification for Estimating Equation

Consider the formulation of the estimating equation given in (5):

$$U_P(\beta) = \frac{1}{n^2} \sum_i \sum_{j < i} (X_i - X_j) \{ P(\tilde{\varepsilon}_i^\beta < \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - P(\tilde{\varepsilon}_i^\beta > \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) \}$$

We can think of the probabilities as expectations of an indicator function conditional on what we observe:

$$\frac{1}{n^2} \sum_i \sum_{j < i} (X_i - X_j) E \{ I(\tilde{\varepsilon}_i^\beta < \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - I(\tilde{\varepsilon}_i^\beta > \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) \}$$

where the expectation is taken with respect to the distribution of the residual failure times conditional on the disease states at the residual follow-up times. This function can be seen to be centered at 0 when  $\beta = \beta_0$ , as its expectation is:

$$(X_i - X_j) E \{ E \{ I(\tilde{\varepsilon}_i^\beta < \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) - I(\tilde{\varepsilon}_i^\beta > \tilde{\varepsilon}_j^\beta | S_i(e_i^\beta), S_j(e_j^\beta)) \} \}$$

where the outside expectation is taken with respect to the distribution of the observed states at the residual follow-up times. By the law of iterated expectations, this is simply equal to:

$$(X_i - X_j) \{ P(\tilde{\varepsilon}_i^\beta < \tilde{\varepsilon}_j^\beta) - P(\tilde{\varepsilon}_i^\beta > \tilde{\varepsilon}_j^\beta) \}$$

Since  $\tilde{\varepsilon}_i^\beta$  and  $\tilde{\varepsilon}_j^\beta$  are i.i.d. and independent of  $X_i$  and  $X_j$  when  $\beta = \beta_0$ , it follows that the expectation is 0 under boundedness of the residual failure time and log censoring time densities, and the covariates.



### A.3 Variance estimation for $\hat{\beta}$ : Gaussian Quadrature Method

First, we give the assumptions in Jin et al. for validity of their Monte Carlo Method and Gaussian Quadrature Method of variance estimation (Jin et al 2014). Suppose we denote the estimating equation as  $U(\beta)$ , and  $\beta_0$  is the true parameter vector:

Assumption 1:  $\sqrt{n}U(\beta_0)$  is asymptotically normal with mean 0 and covariance matrix  $D$ .

Assumption 2: The estimator  $\hat{\beta}$  is root-n consistent, and  $\sqrt{n}(\hat{\beta} - \beta_0)$  is asymptotically normal with mean 0 and covariance matrix  $V$ .

Assumption 3:  $U(\beta)$  is locally asymptotically linear in a neighborhood of  $\beta_0$ .

Let  $B$  be the limiting slope matrix of  $U(\beta_0)$ .  $B$  is difficult to estimate because the estimating function  $U$  is not smooth in  $\beta$ . First, we define  $\Gamma = n^{-1/2} V^{1/2}$ , where  $V = B^{-1} D B^{-1}$ , i.e. the variance of  $\sqrt{n}(\hat{\beta} - \beta_0)$ . We are ultimately interested in estimating  $\Gamma$ , which depends on  $B$ . Jin et al. show that the derivative  $B$  of a smoothed version of the estimating equation satisfies the following expression:

$$B(\Gamma; \beta) = E_Z[U(\beta + \Gamma Z)Z^T \Gamma^{-1}]. \tag{12}$$

We can use Gaussian quadrature or Monte Carlo methods to approximate  $B(\Gamma; \beta)$  and evaluate  $\Gamma$ , but notice that  $B(\Gamma; \beta)$  also depends on  $\Gamma$ , resulting in an iterative algorithm. We describe our implementation of the algorithm for the Gaussian Quadrature Method below:

1. Calculate an estimate  $\hat{D}$  for  $D$ , the covariance matrix of  $\sqrt{n}U(\beta)$ . This can be done using the formula in (6), or a bootstrap procedure. Set  $\Gamma_0 = n^{-1/2}I$ .
2. Suppose the dimension of  $\beta$  is  $p$ . Choose  $m$  nodes  $x_j, j = 1, \dots, m$ , based on one-dimensional Gauss-Hermite quadrature, and let  $z_1, z_2, \dots, z_m$  each be a  $p \times 1$  vector for a unique single combination of the  $m$  nodes among  $p$  points. For example, if we choose 5 1-D Gauss-Hermite quadrature nodes, and we had 2  $\beta$ 's to estimate, we would have  $5^2$  unique vectors  $z_j$  of 2-dimensional nodes for estimating the (double) integral of interest; these two dimensional nodes would be  $(x_1, x_1), (x_1, x_2), \dots, (x_2, x_1), \dots, (x_5, x_5)$  (see Figure 2). Let  $w_j$  be the  $p \times 1$  vector of Gaussian quadrature weights corresponding to the nodes in  $z_j$ . Thus, we will have a grid of points over which we approximate the  $p$ -dimensional integral  $B(\Gamma; \beta)$ . We are interested in computing the integral,

$$\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^p e^{-x_p^2/2} \dots e^{-x_1^2/2} [U(\beta + \Gamma x)x^T \Gamma^{-1}] dx_1 \dots dx_p.$$

Since Gauss-Hermite quadrature computes integrals of the form  $\int_{-\infty}^{\infty} e^{-x^2} f(x) dx$ , we use a change of variable on  $x$  so that we can write the integral in this form. Set  $x^* = \sqrt{2}x =$  then



the integral becomes  $\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{\pi}}\right)^p e^{-x_1^2} \dots e^{-x_p^2} [U(\beta + \Gamma x^*) x^{*T} \Gamma^{-1}] dx_1^* \dots dx_p^*$ . Thus, let  $z_j^* = \sqrt{2}z_j$  for all  $j$ , and proceed.

3. Compute at the  $k^{\text{th}}$  step:

$$B_k = B(\Gamma_{k-1}; \hat{\beta}) = \frac{1}{\sqrt{\pi^p}} \sum_{j=1}^m U(\hat{\beta} + \Gamma_{k-1} z_j^*) z_j^{*T} \Gamma_{k-1}^{-1} \prod_{l=1}^p w_{jl} \quad (13)$$

where  $w_{jl}$  is the  $l^{\text{th}}$  element of the weight vector  $w_j$ .

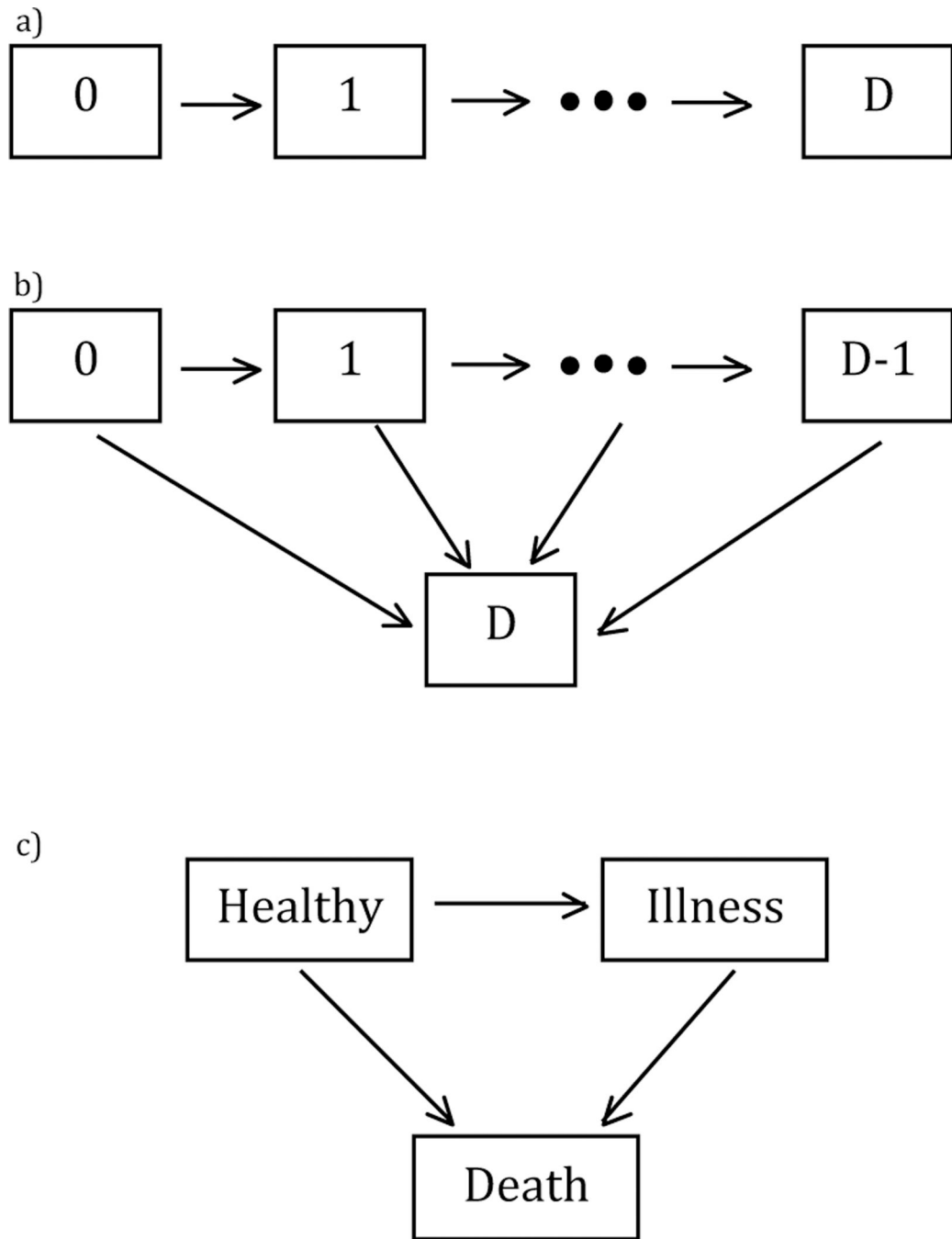
4. Calculate  $G_k = B_k^{-1} \hat{D} B_k^{-1}$  and let  $\Gamma_k = G_k^{1/2} n_k^{-1/2}$ .
5. Repeat steps 3 and 4 until  $\Gamma_k$  converges within a specified tolerance level.

The diagonal of the matrix  $\Gamma_k$  at the last iteration yields the standard error estimates for the vector  $\hat{\beta}$ . The MCM is the same as the above method, except that in step 2 the  $z_j$  vectors are randomly generated from a standard multivariate normal distribution, and in step 3  $B_k$  is estimated as  $B(\Gamma_{k-1}; \hat{\beta}) = \frac{1}{m} \sum_{j=1}^m U(\hat{\beta} + \Gamma_{k-1} z_j) z_j^T \Gamma_{k-1}^{-1}$ . In simulations, we found that as few as 8–10 Gauss-Hermite nodes worked reasonably well for the variance estimation when there is a single covariate.

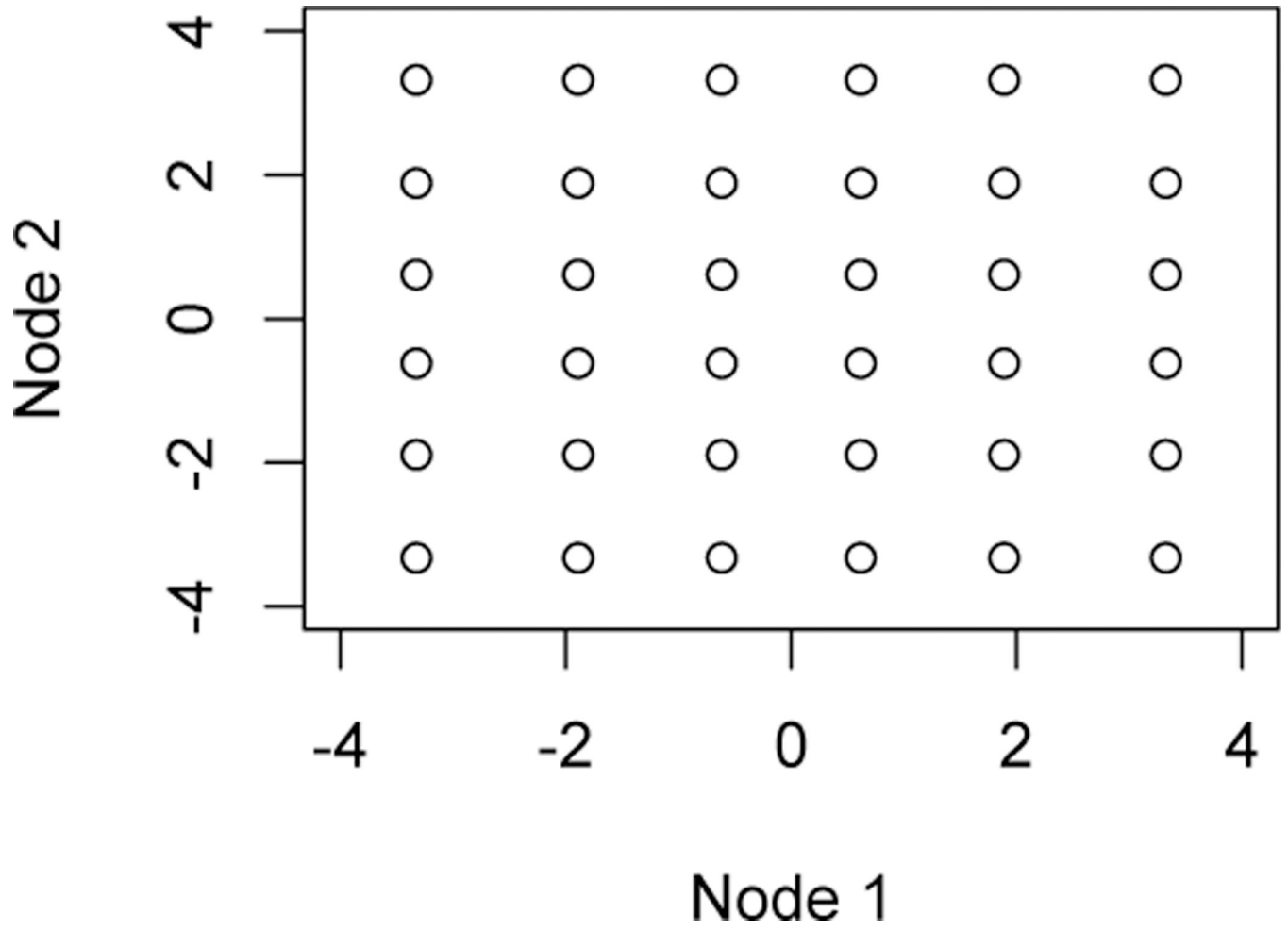
## References

- Aalen O, Borgan O, Gjessing H (2008) Survival and event history analysis: a process point of view Springer Science & Business Media
- Aalen OO, Johansen S (1978) An empirical transition matrix for non-homogeneous markov chains based on censored observations. Scandinavian Journal of Statistics pp 141–150
- Allignol A, Schumacher M, Beyersmann J, et al. (2011) Empirical transition matrix of multistate models: the etm package. Journal of Statistical Software 38(4):1–15
- Berry JD, Shefner JM, Conwit R, Schoenfeld D, Keroack M, Felsenstein D, Krivickas L, David WS, Vriesendorp F, Pestronk A, et al. (2013) Design and initial results of a multi-phase randomized trial of ceftriaxone in amyotrophic lateral sclerosis. PLoS One 8(4):
- Brown B, Wang YG (2007) Induced smoothing for rank regression with censored survival times. Statistics in medicine 26(4):828–836 [PubMed: 16646005]
- Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A (1999) The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function. Journal of the neurological sciences 169(1):13–21 [PubMed: 10540002]
- Chiou SH, Kang S, Kim J, Yan J (2014) Marginal semiparametric multivariate accelerated failure time model with generalized estimating equations. Lifetime data analysis 20(4):599–618 [PubMed: 24549607]
- Cox DR (1972) Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological) 34(2):187–220
- Efron B (1967) The two sample problem with censored data. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Prentice-Hall Engewood Cliffs, NJ, vol 4, pp 831–853
- Fygenon M, Ritov Y (1994) Monotone estimating equations for censored data. The Annals of Statistics pp 732–746
- Gehan EA (1965) A generalized wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52(1–2):203–223 [PubMed: 14341275]

- Heller G (2007) Smoothed rank regression with censored data. *Journal of the American Statistical Association* 102(478)
- Huang Y (2002) Censored regression with the multistate accelerated sojourn times model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(1):17–29
- Jin Z, Lin D, Wei L, Ying Z (2003) Rank-based inference for the accelerated failure time model. *Biometrika* 90(2):341–353
- Jin Z, Shao Y, Ying Z (2014) A monte carlo method for variance estimation for estimators based on induced smoothing. *Biostatistics* p kxu021
- Johnson LM, Strawderman RL (2009) Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika* p asp025
- Kalbfleisch JD, Prentice RL (2011) *The statistical analysis of failure time data*, vol 360 John Wiley & Sons
- Lin JS, Wei L (1992) Linear regression analysis for multivariate failure time observations. *Journal of the American Statistical Association* 87(420):1091–1097
- Lu X, Tsiatis AA (2008) Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika* 95(3):679–694
- Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports Part 1* 50(3):163–170 [PubMed: 5910392]
- Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C (2006) Nonparametric estimation of transition probabilities in a non-markov illness–death model. *Lifetime Data Analysis* 12(3):325–344 [PubMed: 16917736]
- Nelder JA, Mead R (1965) A simplex method for function minimization. *The computer journal* 7(4): 308– 313
- Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society Series A (General)* 135(2):185–207
- Prentice RL (1978) Linear rank tests with right censored data. *Biometrika* 65(1):167–179
- R Core Team (2014) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>
- Ramchandani R, Finkelstein DM, Schoenfeld DA (2015) A model-informed rank test for right-censored data with intermediate states. *Statistics in medicine*
- Tsiatis AA (1990) Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* pp 354–372
- de Uña-Álvarez J, Meira-Machado L (2015) Nonparametric estimation of transition probabilities in the non-markov illness-death model: A comparative study. *Biometrics*
- Van der Vaart AW (2000) *Asymptotic statistics*, vol 3 Cambridge university press



**Fig. 1.** a) Progressive multi-state model; b) Progressive multi-state model where absorbing state can be reached from any state; c) Illness-death model without recovery from illness



**Fig. 2.**  
 Points  $z_j^*$  used to evaluate the double integral in (10)

**Table 1**

Equal Censoring. Dist: Sojourn Time Distributions; PC: Percent Censoring; SE: empirical standard error; SEE: mean of standard error estimator; CP: 95% coverage probability; RE: Relative efficiency of Proposed estimator compared to Gehan/P-P estimator =  $MSE(Gehan/P-P)/MSE(Proposed)$ ; EV: Extreme-Value (log-Weibull); L: Logistic; N: Normal; CN1: Correlated Normal ( $\rho = 0.5$ ); CN2: Correlated Normal( $\rho=0.9$ ).

N	Dist.	PC	Proposed				Gehan			Peto-Prentice			Log-Rank			
			Bias	SE	SEE	CP	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE	
100	EV	50	0.012	0.196	0.198	0.935	0.016	0.216	1.213	0.016	0.206	1.101	0.012	0.200	1.034	
		75	0.013	0.263	0.261	0.947	0.024	0.321	1.496	0.026	0.306	1.356	0.026	0.306	1.358	
	L	50	0.012	0.303	0.295	0.938	0.014	0.314	1.073	0.013	0.318	1.101	0.013	0.335	1.219	
		75	0.008	0.351	0.348	0.946	0.009	0.390	1.239	0.013	0.388	1.222	0.012	0.395	1.268	
	N	50	-0.005	0.180	0.179	0.943	0.002	0.191	1.133	0.000	0.190	1.117	0.005	0.201	1.257	
		75	-0.007	0.221	0.221	0.936	-0.000	0.253	1.314	-0.001	0.255	1.330	0.003	0.265	1.447	
	CN1	50	0.005	0.211	0.205	0.941	0.005	0.225	1.144	0.007	0.226	1.153	0.003	0.241	1.308	
		75	0.011	0.245	0.243	0.931	0.012	0.284	1.340	0.013	0.282	1.318	0.014	0.296	1.455	
	CN2	50	-0.004	0.223	0.220	0.940	-0.005	0.238	1.136	-0.006	0.240	1.160	-0.003	0.252	1.282	
		75	0.004	0.259	0.264	0.946	0.014	0.307	1.410	0.010	0.305	1.389	0.010	0.315	1.481	
	200	EV	50	0.004	0.145	0.139	0.939	0.006	0.161	1.229	0.005	0.153	1.122	0.010	0.151	1.092
			75	-0.010	0.174	0.184	0.951	-0.004	0.214	1.504	-0.005	0.201	1.334	-0.004	0.206	1.401
L		50	0.005	0.212	0.210	0.943	0.007	0.218	1.058	0.008	0.219	1.069	0.005	0.233	1.202	
		75	-0.012	0.236	0.241	0.951	-0.002	0.268	1.291	-0.002	0.264	1.247	-0.000	0.268	1.286	
N		50	-0.000	0.130	0.127	0.944	-0.001	0.135	1.081	0.000	0.136	1.100	-0.002	0.145	1.244	
		75	0.003	0.145	0.156	0.953	0.010	0.166	1.316	0.009	0.166	1.307	0.012	0.174	1.435	
CN1		50	-0.006	0.143	0.146	0.959	-0.007	0.155	1.173	-0.006	0.154	1.151	-0.007	0.163	1.296	
		75	0.001	0.173	0.172	0.943	0.007	0.203	1.384	0.007	0.205	1.405	0.009	0.213	1.517	
CN2		50	0.000	0.155	0.156	0.949	-0.001	0.168	1.181	-0.000	0.167	1.169	0.003	0.175	1.279	
		75	-0.002	0.180	0.186	0.955	-0.005	0.212	1.394	-0.005	0.210	1.361	-0.005	0.217	1.459	

**Table 2**

Unequal Censoring. Dist: Sojourn Time Distributions; PC: Percent Censoring; SE: empirical standard error; SEE: mean of standard error estimator; CP: 95% coverage probability; RE: Relative efficiency of Proposed estimator compared to Gehan/P-P/LR estimator =  $MSE(Gehan/P-P/LR)/MSE(Proposed)$ ; EV: Extreme-Value (log-Weibull); L: Logistic; N: Normal; CN1: Correlated Normal ( $\rho = 0.5$ ); CN2: Correlated Normal( $\rho=0.9$ ).

N	Dist.	PC	Proposed				Gehan			Peto-Prentice			Log-Rank		
			Bias	SE	SEE	CP	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE
100	EV	50	-0.013	0.214	0.200	0.928	-0.018	0.234	1.193	-0.017	0.225	1.109	-0.018	0.218	1.040
		75	-0.018	0.257	0.273	0.953	-0.033	0.323	1.592	-0.029	0.302	1.390	-0.025	0.298	1.351
	L	50	-0.016	0.308	0.297	0.933	-0.017	0.311	1.016	-0.016	0.315	1.043	-0.013	0.331	1.155
		75	-0.026	0.337	0.352	0.950	-0.031	0.379	1.261	-0.032	0.377	1.251	-0.026	0.387	1.316
	N	50	0.001	0.179	0.180	0.946	-0.002	0.189	1.112	-0.003	0.189	1.120	0.002	0.206	1.320
		75	-0.000	0.216	0.223	0.940	-0.002	0.241	1.247	-0.004	0.246	1.298	-0.006	0.253	1.378
	CN1	50	-0.007	0.212	0.205	0.932	-0.013	0.229	1.162	-0.011	0.229	1.167	-0.014	0.241	1.292
		75	-0.010	0.243	0.249	0.943	-0.022	0.279	1.322	-0.018	0.283	1.361	-0.025	0.286	1.392
	CN2	50	-0.004	0.233	0.223	0.934	-0.004	0.249	1.138	-0.004	0.248	1.125	-0.003	0.259	1.229
		75	-0.005	0.262	0.267	0.949	-0.012	0.313	1.435	-0.007	0.316	1.462	-0.009	0.325	1.549
200	EV	50	-0.003	0.145	0.142	0.950	-0.007	0.161	1.232	-0.006	0.154	1.134	-0.010	0.151	1.089
		75	-0.001	0.182	0.190	0.951	-0.016	0.226	1.561	-0.012	0.209	1.333	-0.012	0.211	1.355
	L	50	0.012	0.213	0.208	0.943	0.009	0.214	1.007	0.009	0.217	1.041	0.012	0.232	1.189
		75	-0.002	0.232	0.246	0.955	-0.001	0.261	1.265	-0.002	0.265	1.307	-0.003	0.272	1.378
	N	50	-0.003	0.127	0.129	0.949	-0.005	0.133	1.094	-0.005	0.132	1.084	-0.005	0.142	1.253
		75	0.000	0.149	0.160	0.962	-0.006	0.169	1.288	-0.006	0.170	1.303	-0.005	0.177	1.421
	CN1	50	0.001	0.146	0.146	0.945	-0.001	0.156	1.151	-0.001	0.157	1.155	-0.002	0.165	1.285
		75	-0.010	0.165	0.177	0.959	-0.011	0.193	1.375	-0.014	0.192	1.368	-0.016	0.201	1.490
	CN2	50	-0.011	0.161	0.158	0.937	-0.015	0.173	1.158	-0.015	0.173	1.157	-0.017	0.179	1.256
		75	-0.009	0.179	0.191	0.963	-0.017	0.212	1.399	-0.015	0.210	1.369	-0.016	0.215	1.441