Scientific
Research

Survival Model Inference Using Functions of Brownian Motion

John O'Quigley
Universite Pierre et Marie Curie, Paris, France
Email: john.oquigley@upmc.edu

ABSTRACT

A family of tests for the presence of regression effect under proportional and non-proportional hazards models is described. The non-proportional hazards model, although not completely general, is very broad and includes a large number of possibilities. In the absence of restrictions, the regression coefficient, $\beta(t)$, can be any real function of time. When $\beta(t) = \beta$, we recover the proportional hazards model which can then be taken as a special case of a non-proportional hazards model. We study tests of the null hypothesis; $H_0 : \beta(t) = 0$ for all t against alternatives such as; $H_1 : \int \beta(t) \, dF(t) \neq 0$ or $H_1 : \beta(t) \neq 0$ for some t. In contrast to now classical approaches based on partial likelihood and martingale theory, the development here is based on Brownian motion, Donsker's theorem and theorems from O'Quigley [1] and Xu and O'Quigley [2]. The usual partial likelihood score test arises as a special case. Large sample theory follows without special arguments, such as the martingale central limit theorem, and is relatively straightforward.

Keywords: Brownian Motion; Brownian Bridge; Cox Model; Integrated Brownian Motion; Kaplan-Meier Estimate; Non-Proportional Hazards; Reflected Brownian Motion; Time-Varying Effects; Weighted Score Equation

1. Introduction

1.1. Background

The complex nature of data arising in the context of survival studies is such that it is common to make use of a multivariate regression model. Cox's semi-parametric proportional hazards model [3] has enjoyed wide use in view of its broad applicability. The model makes the key assumption that the regression coefficients do not change with time and much study has gone into investigating and correcting for potential departures from these assumptions [4-10]. Sometimes we can anticipate in advance that the proportional hazards model may be too restrictive. The example which gave rise to our own interest in this question concerned 2174 breast cancer patients, followed over a period of 15 years at the Institut Curie in Paris, France. For these data, as well as a number of other studies in breast cancer, the presence of non-proportional hazards effects has been observed by several authors. Often this is ignored but this can seriously impact inferences.

The model used to make inferences will then often differ from that which can be assumed to have generated the observations. In situations of non proportional hazards, unless dealing with very large data sets relative to the number of studied covariates, it will often not be feasible to study the whole, possibly of infinite dimension, $\beta(t)$. Xu and O'Quigley [11] argue that an estimate of average effect can be used in a preliminary analysis of a data set with time varying regression effects. For a given sample, a single average effect can be estimated more accurately (and more easily) than the whole $\beta(t)$. Xu and O'Quigley [11] derive an estimate $\tilde{\beta}$ of an average regression effect β^*. They provide an interpretation of β^* as a population average effect. It is approximated by $\int \beta(t) \, dF(t) = E\{\beta(T)\}$ under certain conditions, where F is the marginal distribution function of the failure time random variable T. The purpose here is to test the null hypothesis; $H_0 : \beta(t) = 0$ for all t against the alternatives $H_1 : \int \beta(t) \, dF(t) \neq 0$ or $H_1 : \beta(t) \neq 0$ for some t. The development is based on Brownian motion, Donsker's theorem and theorems from O'Quigley [1] and Xu and O'Quigley [2]. We show that the usual partial likelihood score test arises as a special case. Large sample theory is straightforward.

1.2. Notation

The probability structure, although quite simple, is not the immediate one which would come to mind. The ran-

dom variables of interest are the failure time, T_i, the censoring time, C_i, and the possibly time dependent covariate, $Z_i(\cdot)$, $i = 1, \cdots, n$. We view these as a random sample from the distribution of T, C and $Z(\cdot)$. It will not be particularly restrictive and is helpful to our development to assume that T and C have support on some finite interval. The time-dependent covariate $Z(\cdot)$ is assumed to be a predictable stochastic process and, for ease of exposition, taken to be of dimension one whenever possible. Let $F(t) = P(T < t)$, $A(t) = P(C < t)$ and $H(t) = F(t) - \int_0^t \{F(t) - F(u)\} dA(u)$. For each subject i we observe $X_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \le C_i)$. The "at risk" indicator $Y_i(t)$ is defined as, $Y_i(t) = I(X_i \ge t)$ The counting process $N_i(t)$ is defined as, $N_i(t) = I\{T_i \le t, T_i \le C_i\}$ and we also define $\bar{N}(t) = \sum_1^n N_i(t)$. The inverse function, $\bar{N}^{-1}(j)$ corresponds to the value t_j where

$$t_j = \bar{N}^{-1}(j) = \inf\{t : \bar{N}(t) = j\}.$$ It is of notational convenience to define $Z(t) = \sum_1^n Z_i(t) I(X_i = t, \delta_i = 1)$, in words a continuous function equal to zero apart from at the observed failures in which it assumes the covariate value of the subject that fails. The number of observed failures k is given by $k = \bar{N}(\infty)$. If there are ties in the data our suggestion is to split them randomly although there are a number of other suggested ways of dealing with ties. All of the techniques described here require only superficial modification in order to accommodate any of these other approaches for dealing with ties.

1.3. Models

Insight is helped when we group the models together under as general a heading a possible. The most general model is then the non proportional hazards model written,

$$\lambda(t|Z(t)) = \lambda_0(t) \exp\{\beta(t) Z(t)\}, \tag{1.1}$$

where $\lambda(t|\cdot)$ is the conditional hazard function, $\lambda_0(t)$ the baseline hazard and $\beta(t)$ the time-varying regression effect. Whenever $Z(t)$ has dimension greater than one we view $\beta(t) Z(t)$ as an inner product, $\beta(t)$ having the same dimension as $Z(t)$. In order to avoid problems of identifiability we assume that $Z(t)$, if indeed time-dependent, has a clear interpretation such as the value of a prognostic factor measured over time, so that $\beta(t)$ is precisely the regression effect of $Z(t)$ on the log hazard ratio at time t. The above model becomes a proportional hazards model under the restriction that $\beta(t) = \beta$, a constant *i.e.*

$$\lambda(t|Z(t)) = \lambda_0(t) \exp\{\beta Z(t)\}. \tag{1.2}$$

O'Quigley and Stare [12] introduced the name "partially proportional hazards models" to describe models in which at least one component of the function $\beta(t)$ is constrained to be constant. Such models can be shown to include the stratified proportional hazards model [13] whereby;

$$\lambda(t|Z(t), w) = \lambda_{0w}(t) \exp\{\beta Z(t)\} \tag{1.3}$$

as well as random effects models [12].

2. Model Based Probabilities

The probability structure of the model, needed in our development, is described in O'Quigley [1]. We recall the main results in this section. Most often time is viewed as a set of indices to certain stochastic processes, so that, for example, we consider $Z(t)$ to be a random variable having different distributions for different t. Also, the failure time variable T can be viewed as a non-negative random variable with distribution $F(t)$ and, whenever the set of indices t to the stochastic process coincide with the support for T, then, not only can we talk about the random variables $Z(t)$ for which the distribution corresponds to $P(Z \le z | T = t)$ but also marginal quantities such as the random variable $Z(T)$ having distribution $G(z) = P(Z \le z)$. An important result concerning the conditional distribution of $Z(t)$ given $T = t$ follows. First we need the following:

Definition 1. *The discrete probabilities* $\pi_i(\beta(t), t)$ *are given by*;

$$\pi_i(\beta(t), t) = \frac{Y_i(t) \exp\{\beta(t) Z_i(t)\}}{\sum_{j=1}^n Y_j(t) \exp\{\beta(t) Z_j(t)\}}. \tag{2.1}$$

Under (1.2), *i.e.* the constraint $\beta(t) = \beta$, the product of the π's over the observed failure times gives the partial likelihood [3]. When $\beta = 0$, $\{\pi_i(0, t)\}_i$ is the empirical distribution that assigns equal weight to each sample subject in the risk set. Based upon the $\pi_i(\beta(t), t)$ we have:

Definition 2. *Moments of Z with respect to* $\pi_i(\beta(t), t)$ *are given by*;

$$\mathcal{E}_{\beta(t)}(Z^k | t) = \sum_{i=1}^n Z_i^k(t) \pi_i(\beta(t), t), k = 1, 2, \cdots \tag{2.2}$$

Definition 3. *In order to distinguish conditionally independent censoring from independent censoring we define* $\phi(z, t)$ *where*;

$$\phi(z, t) = \frac{\int P(C \ge t | w) g(w) dw}{P(C \ge t | z)}, g(z) = G'(z).$$

Note that when censoring does not depend upon z then $\phi(z, t)$ will depend upon neither z or t and is, in fact, equal to one. Otherwise, under a conditionally inde-

AM

pendent censoring assumption, we can consistently estimate $\phi(z,t)$ and we call this $\hat{\phi}(z,t)$. The following theorem underlies our development.

Theorem 1. *Under model* (1) *and assuming* $\beta(t)$ *known, the conditional distribution function of* $Z(t)$ *given* $T = t$ *is consistently estimated by*

$$\hat{P}\{Z(t) \le z | T = t\} = \frac{\sum_{z_i \le z} Y_i(t) \exp\{\beta(t) z_i(t)\} \hat{\phi}(z_i, t)}{\sum_{j=1}^{n} Y_j(t) \exp\{\beta(t) z_j(t)\} \hat{\phi}(z_j, t)}$$
(2.3)

Proof. (see [1])

Straightforward applications of Slutsky's theorem enable us to claim the result continues to hold whenever $\beta(t)$ is replaced by any consistent estimator $\hat{\beta}(t)$, in particular the partial likelihood estimator when we assume the more restricted model (1.2). □

The theorem has many important consequences includeing;

Corollary 1. *Under model* (1) *and an independent censorship, assuming* $\beta(t)$ *known, the conditional distribution function of* $Z(t)$ *given* $T = t$ *is consistently estimated by*

$$\hat{P}\left(Z(t) \le z | T = t\right) = \sum_{j=1}^{n} \pi_j\left(\beta(t), t\right) I\left(Z_j(t) \le z\right). \quad (2.4)$$

Corollary 2. *For a conditionally independent censoring mechanism we have*

$$\hat{P}\left(Z(t) \le z | T = t, C > t\right) = \sum_{j=1}^{n} \pi_j\left(\beta(t), t\right) I\left(Z_j(t) \le z\right).$$
(2.5)

Again simple applications of Slutsky's theorem shows that the result still holds for $\beta(t)$ replaced by any consistent estimate. When the hypothesis of proportionality of risks is correct then the result holds for the estimate $\hat{\beta}$. Having first defined $\mu = \int \beta(t) \mathrm{d}F(t)$, it is also of interest to consider the approximation;

$$\hat{P}\left(Z(t) \le z | T = t, C > t\right) \approx \sum_{j=1}^{n} \pi_j\left(\mu, t\right) I\left(Z_j(t) \le z\right) (2.6)$$

and, for the case of an independent censoring mechanism,

$$\hat{P}\left(Z(t) \le z | T = t\right) \approx \sum_{j=1}^{n} \pi_j\left(\mu, t\right) I\left(Z_j(t) \le z\right). \quad (2.7)$$

For small samples it will be unrealistic to hope to obtain reliable estimates of $\beta(t)$ for all of t so that, often, we take an estimate of some summary measure, in particular μ. It is in fact possible to estimate μ without estimating $\beta(t)$ [11] although the usual partial likelihood estimate does not accomplish this. In fact the partial likelihood estimate turns out to be equivalent to obtaining the solution of an estimating equation based upon

$H(t)$ and using $\hat{H}(t)$ as an estimate whereas, to consistently estimate μ, it is necessary to work with some consistent estimate of $F(t)$, in particular the Kaplan-Meier estimate. We firstly need some definition of what is being estimated when the data are generated by model (1.1) and we are working with model (1.2). This is contained in the following definition for β^*.

Definition 4. *Let* β^* *be the constant value satisfying*

$$\int_T \mathcal{E}_{\beta^*}\left(Z|t\right) \mathrm{d}F(t) = \int_T \mathcal{E}_{\beta(t)}\left(Z|t\right) \mathrm{d}F(t). \quad (2.8)$$

The definition enables us to make sense out of using estimates based on (1.2) when the data are in fact generated by (1.1). Since we can view T as being random, whenever $\beta(t)$ is not constant, we can think of having sampled from $\beta(T)$. The right hand side of the above equation is then a double expectation and β^* the best fitting value under the constraint that $\beta(t) = \beta$ We can show the existence and uniqueness of solutions to Equation (8) [11] More importantly, β^* can be shown to have the following three properties; 1) under model (1.2) $\beta^* = \beta$; 2) under a subclass of the Harrington-Fleming models, $\beta^* = \int_T \beta(t) \mathrm{d}F(t)$ and 3) for general situations $\beta^* \approx \int_T \beta(t) \mathrm{d}F(t)$ Estimates of β^* are discussed in [2,11] and, in the light of the foregoing, we can take these as estimates of μ. We also have the further two corollaries to Therorem 1:

Corollary 3. *For* $k = 1, 2, \cdots$, $E_{\hat{\beta}(t)}\left(Z^k | t\right)$ *provides consistent estimates of* $E_{\beta(t)}\left(Z^k(t) | T = t\right)$, *under model* (1). *In particular* $E_{\hat{\beta}}\left(Z^k | t\right)$ *provides consistent estimates of* $E_{\beta}\left(Z^k(t) | T = t\right)$, *under model* (1.2).

Furthermore, once again under the model, if we let $V_{\beta(t)}\left(Z|t\right) = E_{\beta(t)}\left(Z^2 | t\right) - E_{\beta(t)}^2\left(Z|t\right)$; then,

Corollary 4. *Under model* (1.2), $Var\left(Z|t\right)$ *is consistently estimated by* $V_{\hat{\beta}}\left(Z|t\right)$.

Theorem 1 and its corollaries provide the ingredients necessary to a construction from which several tests can be derived.

3. Important Empirical Processes

Consider the partial scores introduced by Wei [14];

$$U\left(\beta, t\right) = \int_0^t \left\{Z(s) - \mathcal{E}_\beta\left(Z|s\right)\right\} \mathrm{d}\bar{N}(s) \quad (3.1)$$

Wei was interested in goodness of fit for the two group problem and based a test on $\sup_t U\left(\hat{\beta}, t\right)$, large values indicating departures away from proportional hazards in the direction of non proportional hazards. Considerable exploration of this idea, and substantial generalization via the use of martingale based residuals, has been carried out by Lin, Wei and Ying [15]. These investigations

AM

showed that we could work with a much broader class of statistics that those based on the score so that a wide choice of functions, potentially describing different kinds of departures from the model, are available. Apart from the two group case, limiting distributions are complicated and usually approximated via simulation. Although the driving idea is that of goodness of fit, the same techniques can be applied to testing for the presence of regression effects against a null hypothesis that $\beta(t) = 0$. Furthermore, working directly with the increments of the process rather than the process itself, we can derive related processes for which the limiting distributions are available analytically. From the previous section the increments of the process $\int_0^t Z(s) d\bar{N}(s)$ at $t = X_i$ have mean $E(Z|X_i)$ and variance $V(Z|X_i)$. We can view these increments as being independent [16,17]. Thus only the existence of the variance is necessary in order to carry out appropriate standardization and to be able to appeal to the functional central limit theorem. We can then treat our observed process as though arising from a Brownian motion process. Simple calculations allow us to also work with the Brownian bridge, integrated Brownian motion and reflected Brownian motion, processes which will be useful under particular alternative to the model specified under the null hypothesis. Consider the process $U^*(\alpha, \beta, u)$, $(0 < u < 1)$, in which

$$U^*\left(\alpha, \beta, \frac{j}{k}\right) = \frac{1}{\sqrt{k}} \int_0^{t_j} V_\alpha(Z|s)^{-1/2} dU(\beta, s), \quad j = 1, \cdots, k.$$

(3.2)

where $t_j = \bar{N}^{-1}(j)$. This process is only defined on k equispaced points of the interval (0, 1] but we extend our definition to the whole interval via linear interpolation so that, for u in the interval j/k to $(j+1)/k$, we write;

$$U^*(\alpha, \beta, u) = U^*\left(\alpha, \beta, \frac{j}{k}\right)$$
$$+ \{uk - j\} \left\{ U^*\left(\alpha, \beta, \frac{j+1}{k}\right) - U^*\left(\alpha, \beta, \frac{j}{k}\right) \right\}$$

As n goes to infinity, under the usual Breslow and Crowley conditions, then we have that, for each j $(j = 1, \cdots, k-1)$, $U^*(\alpha, \beta, j/k)$ converges in distribution to a Gaussian process with mean zero and variance equal to j/k. This follows directly from Donsker's theorem. Replacing β by a consistent estimate leaves asymptotic properties unaltered.

3.1. Some Remarks on the Notation $U^*(\alpha, \beta, u)$

Various aspects of the statistic $U^*(\alpha, \beta, u)$ will be used to construct different tests. We choose the * symbol to indicate some kind of standardization as opposed to

the non standardized U. The variance and the number of failure points are used to carry out the standardization. Added flexibility in test construction can be achieved by using the two parameters, α and β, rather than a single parameter β. In practice these are replaced by quantities which are either fixed or estimated under some hypothesis. For goodness of fit procedures which we consider later we will only use a single parameter, typically $\hat{\beta}$. Goodness of fit tests are most usefully viewed as tests of hypotheses of the form $H_0: \beta = \hat{\beta}$. A test then of a hypothesis $H_0: \beta = 0$ may not seem very different. This is true in principle. However, for a test of $H_0: \beta = 0$, we need keep in mind not only behaviour under the null but also under the alternative. Because of this it is often advantageous, under a null hypothesis of $\beta = 0$, to work with $\alpha = \hat{\beta}$ and $\beta = 0$ in the expression $U^*(\alpha, \beta, u)$. Under the null, $\hat{\beta}$ remains consistent for the value 0 and, in the light of Slutsky's theorem, the large sample distribution of the test statistics will not be affected. Under the alternative however things look different. The increments of the process $\int_0^t Z(s) d\bar{N}(s)$ at $t = X_i$ no longer have mean $E(Z|X_i)$ and adding them up will indicate departures from the null. But the denominator is also affected and, in order to keep the variance estimate not only correct but also as small as we can, it is preferable to use the value $\hat{\beta}$ rather than zero.

3.2. Some Properties of $U^*(\alpha, \beta, u)$

A very wide range of possible tests can be based upon the statistic $U^*(\alpha, \beta, u)$ and we consider a number of these below. Well known tests such as the partial likelyhood score test obtain as a special cases. First we need to make some observations on the properties of $U^*(\alpha, \beta, u)$ under different values of α, β and u.

Lemma 1. *The process* $U^*(\alpha, \beta, u)$, *for all finite* α *and* β *is continuous on* [0, 1]. *Also* $EU^*(\beta, \beta, 0) = 0$.

Lemma 2. *Under model* 1.2 $EU^*(\hat{\beta}, \beta, u)$ *converges in probability to zero.*

Lemma 3. *Suppose that* $v < u$, *then* $Cov\{U^*(\hat{\beta}, \beta, v); U^*(\hat{\beta}, \beta, u)\}$ *converges in probability to v.*

Proofs of the above lemmas are all immediate. Since the increments of the process are asymptotically independent we can treat $U^*(\hat{\beta}, \beta, u)$ (as well as $U^*(\beta, \beta, u)$ under some hypothesized β) as though it were Brownian motion. From Corollary 3 we have that $E_{\hat{\beta}}(Z^k|t)$ provides consistent estimates of $E_\beta(Z^k(t)|T = t)$, under model (1.2) and that $V_{\beta(t)}(Z|t)$ is consistent for $Var(Z|t)$. Therefore, at $t = j/k$ $(j = 1, \cdots, k)$, the variance of

$V_{\beta(t)}^{-1/2}\left(Z|t\right)\left[Z-E_{\hat{\beta}}\left(Z^k|t\right)\right]$ goes to the value one, as a simple application of Slutsky's theorem. A further application of Slutsky's theorem, together with theorems of Cox and Andersen and Gill [16,17] provide that the increments $dU\left(\hat{\beta},j/k\right),(j=1,\cdots,k)$ are asymptotically uncorrelated. Let

$$\varepsilon_k\left(t\right)=\max\left\{\frac{\left|dU\left(\beta,1/k\right)\right|}{\sqrt{k}},\frac{\left|dU\left(\beta,2/k\right)\right|}{\sqrt{k}},...,\frac{\left|dU\left(\beta,1\right)\right|}{\sqrt{k}}\right\}$$

Then

$$P\left(\varepsilon_k\left(t\right)>\delta\right)=1-P\left(\varepsilon_k\left(t\right)\leq\delta\right)$$

$$=1-P\left(\frac{\left|dU\left(\beta,j/k\right)\right|}{\sqrt{k}}\leq\delta,\forall j,j=1,\cdots,k\right)$$

$$=1-\left(P\left(\left|dU\left(\beta,1/k\right)\right|\leq\delta\sqrt{k}\right)\right)^k$$

$$=1-\left(1-P\left(\left|dU\left(\beta,1/k\right)\right|>\delta\sqrt{k}\right)\right)^k.$$

Applying the Chebyshev inequality,

$$P\left(\left|dU\left(\beta,1/k\right)\right|>\delta\sqrt{k}\right)\leq E\left|dU\left(\beta,1/k\right)\right|^3\Big/\delta^3k^{3/2}$$

from which

$$P\left(\varepsilon_k\left(t\right)>\delta\right)$$

$$=1-\left(1-E\left|dU\left(\beta,1/k\right)\right|^3\Big/\delta^3k^{3/2}\right)^k$$

$$=O\left\{1-\exp\left(-E\left|dU\left(\beta,1/k\right)\right|^3\Big/\delta^3k^{1/2}\right)\right\}\to0$$

as k becomes large. Apart from the necessity for the existence of the third moment of Z we also require that, as k increases, the fluctuations of the process $U\left(\hat{\beta},u\right)$ between successive failures become sufficiently small in probability, the so called tightness of the process [18]. We can assume this holds in real applications. We then conclude from Donsker's theorem that $U\left(\hat{\beta},u\right)$ converges in distribution to Brownian motion. Replacing β by a consistent estimate leaves asymptotic properties unaltered. Suppose that the assumption of zero effect, *i.e.*, $\beta(t)=0$ is incorrect, and, in particular, that $\beta(t)$ is a smoothly changing monotonic function of time. Without losing generality we will assume this monotonicity to be an increasing one. Now, at each time point t, instead of subtracting off $E_{\beta(t)}\left(Z|t\right)$ from the observed value of Z at that point, we subtract instead $E_0\left(Z|t\right)$. The variance is also impacted but the variance is always positive (thereby impacting the average size but not the average sign of the increments). So, we will observe a positive trend in the standardized residuals, the early ones tending to be too large and the later ones tending to be too large

also but negatively. A good model for this, at least as a first approximation, would be Brownian motion with drift [1]. For our purposes we note that, moving away from the model of zero regression effects, we anticipate observing some trend rather than zero mean Brownian motion. For decreasing $\beta(t)$, the same argument holds leading to an approximation of Brownian motion with drift but with the sign changed. These assertions follow when $\beta(t)$ is monotone. Now suppose that $\beta(t)$ is not monotone. There are two cases of interest. The first is where $\beta(t)$ is broadly monotone, by which we mean the following. Divide the time interval into non overlapping time segments and take the average value of $\beta(t)$ for each segment. We then suppose that the average value of $\beta(t)$, over the different time segments, is monotone. We could make this intuitive idea more precise if needed. For this case we would expect, again, the procedure to work well. The second case of interest is where $\beta(t)$ changes over the time period in question in a way that has no obvious pattern or trend. We would not expect to be able to detect such behavior and the power of the test procedures would be low. We would most likely conclude that there is no effect, a conclusion that, even though not correct, would be reasonable, at least as an approximation.

4. Non- and Partially Proportional Hazards Models

The Brownian motion approximations of the above section extend immediately to the case of non proportional hazards and partially proportional hazards models. The generalization of Equation (1) is natural and would lead to an unstandardized score;

$$U\left(\beta(t),t\right)=\int_0^t\left\{Z(s)-\mathcal{E}_{\beta(t)}\left(Z|s\right)\right\}d\bar{N}(s) \qquad (4.1)$$

and, as before, under the null hypothesis that $\beta(t)$ is correctly specified the function $U\left(\beta(t),t\right)$ will be a sum of zero mean random variables. The range of possible alternative hypotheses is large and, mostly, we will not wish to consider anything too complex. Often the alternative hypothesis will specify an ordering, or a non zero value, for just one of the components of a vector values $\beta(t)$. In the exact same way as in the previous section, all of the calculations lean upon the main theorem and its corollaries. The increments of the process $\int_0^t Z(s)d\bar{N}(s)$ at $t=X_i$ have mean $E\left(Z|X_i\right)$ and variance $V\left(Z|X_i\right)$. A little bit of extra care is needed, in practice, in order to maintain the view of the independence of these increments. When $\beta(t)$ is known there is no problem but if, as usually happens, we wish to use estimates, then, for asymptotic theory to still hold, we require the sample size (number of failures) to become

infinite relative to the dimension of $\beta(t)$. Thus, if we wish to estimate the whole function $\beta(t)$, then some restrictions will be needed because, full generality implies an infinite dimensional parameter $\beta(t)$. For the stratified model and, generally, partially proportional hazards models, the problem does not arise because we do not estimate $\beta(t)$.

The sequentially standardized process will now be written $U^*(\alpha(t), \beta(t), u), (0 < u < 1)$, in which

$$U^*\left(\alpha(t), \beta(t), \frac{j}{k}\right)$$
$$= \frac{1}{\sqrt{k}} \int_0^{t_j} V_{\alpha(t)}(Z|s)^{-1/2} \, dU(\beta(t), s), j = 1, \cdots, k.$$

where $t_j = \bar{N}^{-1}(j)$. This process can be made to cover the whole interval (0, 1] continuously by interpolating in exactly the same way as in the previous section. For this process we reach the same conclusion, *i.e.*, that as n goes to infinity, under the usual Breslow and Crowley conditions [17], then we have that, for each j ($j = 1, \cdots, k-1$), $U^*(\alpha(t), \beta(t), j/k)$ converges in distribution to a Gaussian process with mean zero and variance equal to j/k. The only potential difficulty is making use of Slutsky's theorem whereby, if we replace $\alpha(t)$ and $\beta(t)$ by consistent estimates the result still holds. The issue is that of having consistent estimates, which for an infinite dimensional unrestricted parameter we can not achieve. The solution is simply to either restrict these functions or to work with the stratified models in which we do not need to estimate them. Subsection 3.1 applies equally well here if we replace α and β by $\alpha(t)$ and $\beta(t)$. The lemmas of the above section describing the properties of $U^*(\alpha, \beta, t)$ apply equally we if we are working with $U^*(\alpha(t), \beta(t), t)$, specifically, the process $U^*(\alpha(t), \beta(t), t)$, for all finite $\alpha(t)$ and $\beta(t)$ is continuous on [0, 1] and $EU^*(\beta(t), \beta(t), 0) = 0$, under model 1.2 $EU^*(\hat{\beta}(t), \beta, u)$ converges in probability to zero and for $v < u$,

$Cov\left\{U^*(\hat{\beta}(t), \beta(t), v); U^*(\hat{\beta}(t), \beta(t), u)\right\}$ converges in

probability to v. Since the increments of the process are asymptotically independent we will treat

$U^*(\hat{\beta}(t), \beta(t), t)$ (as well as $U^*(\beta(t), \beta(t), t)$ under some hypothesized $\beta(t)$) as though it were Brownian motion.

5. Test Statistics

Several tests of point hypotheses can be constructed based on the theory of the previous section. These tests can also be used to construct test based confidence intervals of parameter estimates, obtained as solutions to an estimating equation. Among these tests are the following.

5.1. Distance Travelled at Time *t*

At time *t*, under the null hypothesis that $\beta = \beta_0$, often a hypothesis of absence of effect in which case $\beta_0 = 0$, we have that $U^*(\hat{\beta}, \beta_0, t)$ can be approximated by a normal distribution with mean zero and variance *t*. A *p*-value corresponding to the null hypothesis is then obtained from

$$\Pr\left\{U^*(\hat{\beta}, \beta_0, t)\big/\sqrt{t} > z\right\} = 1 - \Phi(z).$$

This *p*-value is for a one-sided test in the direction of the alternative $\beta < \beta_0$. For a one-sided alternative in the opposite direction we would use;

$$\Pr\left\{U^*(\hat{\beta}, \beta_0, t)\big/\sqrt{t} < z\right\} = \Phi(z)$$

and, for a two sided alternative, we would, as usual, consider the absolute value of the test statistic and multiply $1 - \Phi(z)$ by two. Under the alternative, say $\beta > \beta_0$, if we take the first two terms of a Taylor series expansion of $U^*(\hat{\beta}, \beta_0, t)$ about β, we can deduce that a good approximation for this would be Brownian motion with drift. At time *t* this is then a good test for absence of effect (Brownian motion) against a proportional hazards alternative (Brownian motion with drift), good in the sense that type I error is controlled for and, under these alternatives, the test has good power properties. Power will be maximized by using the whole time interval, *i.e.*, taking $t = 1$. Nonetheless there may be situations in which we may opt to take a value of *t* less than one. If we know for instance that, under both the null and the alternative we can exclude the possibility of effects being persistent beyond some time τ say, *i.e.*, the hazard ratios beyond that point should be one or very close to that, then we will achieve greater power by taking *t* to be less than one, specifically some value around τ. A confidence interval for β_0 can be obtained using normal approximations or by constructing the interval $\{\hat{\beta}^-\hat{\beta}^+\}$ such that for any point *b* contained in the interval a test of the null hypothesis, $H_0 : \beta = b$, is not rejected.

5.2. Greatest Distance from Origin at Time *t*

In cases where we wish to consider values of *t* less than one, we may have knowledge of some τ of interest. Otherwise we might want to consider several possible values of τ. Control on Type I error will be lost unless specific account is made of the multiplicity of tests. One simple way to address this issue is to consider the maximum value achieved by the process during the interval $(0, \tau)$. Again we can appeal to known results for some well known functions of Brownian motion. In particular we have;

$$\Pr\left\{\sup_{t\in(0,\tau)} U^*\left(\hat{\beta},\beta_0,t\right)\Big/\sqrt{t} > z\right\}$$
$$= 2\Pr\left\{U^*\left(\hat{\beta},\beta_0,\tau\right)\Big/\sqrt{\tau} > z\right\}.$$

Under the null and proportional hazards alternatives this test, as opposed to the usual score test, would lose power comparable to carrying out a two sided rather than a one-sided test. Under non-proportional hazards alternatives this test could be of use, an extreme example being crossing hazards where the usual score test may have power close to zero. As the absolute value of the hazard ratio increases so would the maximum distance from the origin.

5.3. Brownian Bridge Test

Since we are viewing the process $U^*\left(\hat{\beta},\beta_0,t\right)$ as though it were a realization of a Brownian motion, we can consider some other well known functions of Brownian motion. Consider then the bridged process $U_0^*\left(\hat{\beta},\beta_0,t\right)$;

Definition 5. *The bridged process is defined by the transformation*

$$U_0^*\left(\hat{\beta},\beta_0,t\right) = U^*\left(\hat{\beta},\beta_0,t\right) - tU^*\left(\hat{\beta},\beta_0,1\right)$$

Lemma 4. *The process $U_0^*\left(\hat{\beta},\beta_0,t\right)$ converges in distribution to the Brownian bridge, in particular, for large samples, $EU_0^*\left(\hat{\beta},\beta_0,t\right) = 0$ and*

$$Cov\left\{U_0^*\left(\hat{\beta},\beta_0,s\right),U_0^*\left(\hat{\beta},\beta_0,t\right)\right\} = s(1-t).$$

The Brownian bridge, also referred to as tied down Brownian motion for the obvious reason that at $t=0$ and $t=1$ the process takes the value 0, will not be particularly useful for carrying out a test at $t=1$. It is more useful to consider, as a test statistic, the greatest distance of the bridged process from the time axis. We can then appeal to;

Lemma 5.

$$\Pr\left\{\sup_u\left|U_0^*\left(\hat{\beta},\beta_0,u\right)\right| \geq a\right\} \approx 2\exp\left(-2a^2\right) \qquad (5.1)$$

which follows as a large sample result since;

$$\Pr\left\{\sup_u\left|U_0^*\left(\hat{\beta},\beta_0,u\right)\right| \leq a\right\}$$
$$\rightarrow 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp\left(-2k^2a^2\right), a \geq 0.$$

This is an alternating sign series and therefore, if we stop the series at $k=2$ the error is bounded by $2\exp\left(-8a^2\right)$ which for most values of a that we will be interested in will be small enough to ignore. For alternatives to the null hypothesis $(\beta=0)$ belonging to the proportional hazards class, the Brownian bridge test will be less pow-

erful than the distance from origin test. It is more useful under alternatives of a non-proportional hazards nature, in particular an alternative in which $U^*\left(\hat{\beta},\beta_0,1\right)$ is close to zero, a situation we might anticipate when the hazard functions cross over. Its main use, in our view, is in testing goodness of fit, *i.e.*, a hypothesis test of the form $H_0:\beta=\hat{\beta}$ [1].

5.4. Reflected Brownian Motion

An interesting property of Brownian motion is the following. Let $W(t)$ be Brownian motion, choose some positive value r and define the process $W_r(t)$ in the following way: If $W(t) < r$ then $W_r(t) = W(t)$. If $W(t) \geq r$ then $W_r(t) = 2r - W(t)$. It is easily shown that the reflected process $W_r(t)$ is also Brownian motion. Choosing r to be negative and defining $W_r(t)$ accordingly we have the same result. The process $W_r(t)$ coincides exactly with $W(t)$ until such a time as a barrier is reached. We can imagine this barrier as a mirror and beyond the barrier the process $W_r(t)$ is a simple reflection of $W(t)$. So, consider the process $U^r\left(\hat{\beta},\beta_0,t\right)$ defined to be $U^*\left(\hat{\beta},\beta_0,t\right)$ if $\left|U^*\left(\hat{\beta},\beta_0,t\right)\right| < r$ and to be equal to $2r - U^*\left(\hat{\beta},\beta_0,t\right)$ if $\left|U^*\left(\hat{\beta},\beta_0,t\right)\right| \geq r$.

Lemma 6. *The process $U^r\left(\hat{\beta},\beta_0,t\right)$ converges in distribution to Brownian motion, in particular, for large samples, $EU^r\left(\hat{\beta},\beta_0,t\right) = 0$ and*

$$Cov\left\{U^r\left(\hat{\beta},\beta_0,s\right),U^r\left(\hat{\beta},\beta_0,t\right)\right\} = s.$$

Under proportional hazards there is no obvious role to be played by U^r. However, imagine a non-proportional hazards alternative where the effect reverses at some point, the so-called crossing hazards problem. The statistic $U^*\left(\hat{\beta},0,t\right)$ would increase up to some point and then decrease back to a value close to zero. If we knew this point, or had some reasons for guessing it in advance, then we could work with $U^r\left(\hat{\beta},\beta_0,t\right)$ instead of $U^*\left(\hat{\beta},\beta_0,t\right)$. A judicious choice of the point of reflection would result in a test statistic that continues to increase under such an alternative so that a distance from the origin test might have reasonable power. In practice we may not have any ideas on a potential point of reflection. We could then consider trying a whole class of points of reflection and choosing that point which results in the greatest test statistic. A bound for a supremum type test can be derived by applying the results of Davies [19,20]. Under the alternative hypothesis we could imagine increments of the same sign being added together until the value r is reached, at which point the sign of the increments changes. Under the alternative hypothesis the ab-

solute value of the increments is strictly greater than zero. Under the null, r is not defined and, following the usual standardization, this set up fits in with that of Davies [19, 20]. We can define γ_r to be the time point satisfying $U^*(\hat{\beta}, \beta_0, \gamma_r) = r$. A two-sided test can then be based on the statistic $M = \sup_r \left\{ \left| U^r(\hat{\beta}, \beta_0, 1) \right| : 0 \le \gamma_r \le 1 \right\}$. Inference can then be based upon;

$$\Pr\left\{ \sup \left| U^r(\hat{\beta}, \beta_0, 1) \right| > c : 0 \le \gamma_r \le 1 \right\}$$
$$\le \Phi(-c) + \frac{\exp\left(-c^2/2\right)}{2\pi} \int_0^1 \left\{ -\rho_{11}(\gamma) \right\}^{\frac{1}{2}} \mathrm{d}\gamma \quad (5.2)$$

where Φ denotes the cumulative normal distribution function, $\rho_{11}(\gamma) = \left\{ \partial^2 \rho(\phi, \gamma) / \partial \phi^2 \right\}_{\phi = \gamma}$ and where $\rho(\gamma_r, \gamma_s)$ is the autocorrelation function between $U^r(\hat{\beta}, \beta_0, 1)$ and $U^s(\hat{\beta}, \beta_0, 1)$. In general the autocorrelation function $\rho(\phi, \gamma)$, needed to evaluate the test statistic is unknown. However, it can be consistently estimated using bootstrap resampling methods. For γ_r and γ_s taken as fixed, we can take bootstrap samples from which several pairs of $U^r(\hat{\beta}, \beta_0, 1)$ and $U^s(\hat{\beta}, \beta_0, 1)$ can be obtained. Using these pairs, an empirical, *i.e.* product moment, correlation coefficient can be calculated. Under the usual conditions [21,22], the empirical estimate provides a consistent estimate of the true value. This sampling strategy is investigated in related work by O'Quigley and Natarajan [23].

Davies also suggests an approximation in which the autocorrelation is not needed. This may be written down as

$$\Pr\left\{ \sup \left| U^r(\hat{\beta}, \beta_0, 1) \right| > M \right\} \approx \Phi(-M) + \frac{V_\rho \exp\left(-M^2/2\right)}{\sqrt{8\pi}}$$
$$(5.3)$$

where $V_\rho = \sum_i \left| U^r(\hat{\beta}, \beta_0, 1) - U^s(\hat{\beta}, \beta_0, 1) \right|$, the γ_i, ranging over (L, U), are the turning points of $T(0, \hat{\beta}; \cdot)$ and M is the observed maximum of $T(0, \hat{\beta}; \cdot)$.

Turning points only occur at the k distinct failure times and, to keep the notation consistent with that of the next section, it suffices to take $\gamma_i (i = 2, \cdots, k)$ as being located half way between adjacent failures, $\gamma_1 = 0$ and γ_{k+1} any value greater than the largest failure time. We would though require different inferential procedures for this.

5.5. Partial Likelihood Score Test

Suppose that we wish to test $H_0 : \beta = 0$ and instead of

$U^*(\hat{\beta}, 0, t)$ we choose to work with $U^*(0, 0, t)$. In the light of Slutsky's theorem it is readily seen that the large sample null distributions of the two test statistics are the same. Next, instead of standardizing by $V_0(Z|X_i)$ at each X_i we take a simple average of such quantities, over the observed failures. To see this, note that $\int V_{\hat{\beta}}(Z|t) \mathrm{d}\hat{F}(t)$ is consistent for $E\left[Var(Z|t) \right]$. Rather than integrate with respect to $\hat{F}(t)$ it is more common, in the counting process context, to integrate with respect to $\bar{N}(t)$, the two coinciding in the absence of censoring. It is also more common to fix $\hat{\beta}$ in $V_{\hat{\beta}}(Z|t)$ at its null value zero. This gives us:

Definition 6. *The empirical average conditional variance, V_0 is defined as*

$$V_0 = \int_0^1 V_0(Z|t) \mathrm{d}\bar{N}(t).$$

If, in Equation (3.2), we replace $V_\alpha(Z|s)$ by V_0 then the distance from origin test at time $t = 1$ coincides exactly with the partial likelihood score test. Indeed this observation could be used to construct a characterization of the partial likelihood score test. In epidemiological applications it is often assumed that the conditional variance, $V_\beta(Z|t)$ is constant through time. Otherwise time independence is often a good approximation to the true situation and gives further motivation to the partial likelihood test.

6. Multivariate Model

In practice it is the multivariate setting that we are most interested in; testing for the existence of effects in the presence of related covariates, or possibly testing the combined effects of several covariates. In this work we give very little specific attention to the multivariate setting, not because we do not feel it to be important but because the univariate extensions are almost always rather obvious and the main concepts come through more clearly in the relatively notationally uncluttered univariate case. Nonetheless, some thought is on occasion required. The basic theorem giving a consistent estimate of the distribution of the covariate at each time point t applies equally well when the covariate $Z(t)$ is multi-dimensional. Everything follows through in the same way and there is no need for additional theorems. In the multivariate case the product $\beta(t)Z_i(t)$ becomes a vector or inner product, a simple linear sum of the components of $Z_i(t)$ and the corresponding components of $\beta(t)$. Suppose, for simplicity, that $Z_i(t)$ is two dimensional so that $Z_i(t)^T = \left\{ Z_{1i}(t), Z_{2i}(t) \right\}$. Then the $\pi_i(\beta(t), t)$ give our estimate for the joint distribution of $\left\{ Z_{1i}(t), Z_{2i}(t) \right\}$ at time t. As for any multi-dimen-

sional distribution if we wish to consider only the marginal distribution of, say, $Z_1(t)$ then we simply sum the $\pi_i(\beta(t),t)$ over the variable $Z_2(t)$. In practice we work with the $\pi_i(\beta(t),t)$, defined to be of the highest dimension that we are interested in, for the problem in hand, and simply sum over the subsets of vector Z needed. To be completely concrete let us return to the partial scores,

$$U(\beta,t) = \int_0^t \{Z(s) - E_\beta(Z|s)\} d\bar{N}(s), \qquad (6.1)$$

defined previously for the univariate case. Both $Z(s)$ and $E_\beta(Z|s)$ are vectors of the same dimension. So also is then $U(\beta,t)$. The vector $U(\beta,t)$ is made up of the component marginal processes any of which we may be interested in. For each marginal covariate, let's say Z_1 for instance, we also calculate $E_\beta(Z_1|s)$ and we can do this either by first working out the marginal distribution of Z_1 or just by summing over the joint probabilities. The result is the same and it is no doubt easier to work out all expectations with the respect to the joint distribution. Let us then write;

$$U(\beta,Z_1,t) = \int_0^t \{Z_1(s) - \mathcal{E}_\beta(Z_1|s)\} d\bar{N}(s) \qquad (6.2)$$

where the subscript "1" denotes the first component of the vector. The interesting thing is that E_β does not require any such additional notation, depending only on the joint $\pi_i(\beta(t),t)$. As for the univariate case we can work with any function of the random vector Z, the expectation of the function being readily estimated by an application of an immediate generalization of Corollary 3. Note that the process we are constructing is not the same one that we would obtain were we to simply work with only Z_1. This is because the $\pi_i(\beta(t),t)$ involve a univariate Z in the former case and a multivariate Z in the latter. The increments of the process $\int_0^t Z_1(s) d\bar{N}(s)$ at $t = X_i$ have mean $E(Z_1|X_i)$ and variance $V(Z_1|X_i)$. As before, these increments can be taken to be independent [16,17] so that only the existence of the variance is necessary to be able to appeal to the functional central limit theorem. This observed process will also be treated as though arising from a Brownian motion process. The same calculations as above allow us to also work with the Brownian bridge, integrated Brownian motion and reflected Brownian motion. Our development is entirely analogous to that for the univariate case and we consider now the process $U^*(\alpha,\beta,Z_1,u),(0<u<1)$, in which

$$U^*\left(\alpha,\beta,Z_1,\frac{j}{k}\right) = \frac{1}{\sqrt{k}} \int_0^{t_j} V_\alpha(Z_1|s)^{-1/2} dU(\beta,Z_1,s),$$

$$j = 1,\cdots,k.$$

where $t_j = \bar{N}^{-1}(j)$. This process is only defined on k equispaced points of the interval $(0, 1]$ and, again, we

extend our definition to the whole interval so that, for $u \in \left[j/k,(j+1)/k\right]$ we can write $U^*(\alpha,\beta,Z_1,u)$ as;

$$U^*\left(\alpha,\beta,Z_1,\frac{j}{k}\right)$$

$$+ \{uk-j\}\left\{U^*\left(\alpha,\beta,Z_1,\frac{j+1}{k}\right) - U^*\left(\alpha,\beta,Z_1,\frac{j}{k}\right)\right\}$$

As n goes to infinity, under the usual Breslow and Crowley conditions [17], then we have that, for each j ($j = 1,\cdots,k-1$), $U^*(\alpha,\beta,Z_1,j/k)$ converges in distribution to a Gaussian process with mean zero and variance equal to j/k. This follows in the same way as for the univariate case, directly from Donsker's theorem. Replacing β by a consistent estimate leaves asymptotic properties unaltered.

6.1. Some Further Remarks on the Notation

The notation $U^*(\alpha,\beta,Z_1,u)$ is a little heavy but becomes even heavier if we wish to treat the situation in great generality. The first component of Z is Z_1 but of course this could be any component. Indeed Z_1 can itself be a vector, some collection of components of Z and, once we see the basic idea, it is clear what to do even though the notation starts to become slightly cumbersome. As for the notation, $U^*(\alpha,\beta,u)$, in which there is only one Z and no need to specify it, the * symbol continues to indicate standardization by the variance and number of failure points. For the multivariate situation, the two parameters, α and β, are themselves, both vectors. The parameter α which indexes the variance will be, in practice, the estimated full vector β, i.e., $\hat{\beta}$. Note that, as for the process $U^*(\hat{\beta},\beta,u)$ we use, for the first argument to this function, the unrestricted estimate. Exactly the same applies here. In the numerator however, under some hypothesis for β_1, say $\beta_1 = \beta_{10}$ then, for the increments $dU(\beta,Z_1,s)$, we would have β_1 fixed at β_{10} and the other components of the vector β replaced by their restricted estimates, i.e., zeros of the estimating equations in which $\beta_1 = \beta_{10}$.

6.2. Some Properties of $U^*(\alpha,\beta,Z_1,u)$

The same range of possible tests as before can be based upon the statistic $U^*(\alpha,\beta,Z_1,u)$. To support this it is worth noting:

Lemma 7. *The process* $U^*(\alpha,\beta,Z_1,u)$, *for all finite* α *and* β *is continuous on* [0, 1]. *Also* $EU^*(\beta,\beta,Z_1,0) = 0$.

Lemma 8. *Under model* 2 $EU^*(\hat{\beta},\beta,Z_1,u)$ *converges in probability to zero.*

Lemma 9. *Suppose that* $v < u$, *then*

$$Cov\left\{U^*(\hat{\beta},\beta,Z_1,v);U^*(\hat{\beta},\beta,Z_1,u)\right\}$$

converges in probability to v.

Since the increments of the process are asymptotically independent we can treat $U^*\left(\hat{\beta},\beta,Z_1,u\right)$ (as well as $U^*\left(\beta,\beta,Z_1,u\right)$ under some hypothesized β) as though it were Brownian motion.

6.3. Tests in the Multivariate Setting

When we carry out a test of $H_0:\beta_1=\beta_{10}$ it is important to keep in mind the alternative hypothesis which is, usually, $H_1:\beta_1\neq\beta_{10}$ together with β_j, $j\neq 1$ unspecified. Such a test can be carried out using $U^*\left(\hat{\beta},\beta,Z_1,u\right)$ where, for the second argument β, the component β_1 is replaced by β_{10} and the other components by estimates with the constraint that β_1 is fixed at β_{10}. Assuming our model is correct, or a good enough approximation, then we are testing for the effects of Z_1 having "accounted for" the effects of the other covariates. The somewhat imprecise notion "having accounted for" is made precise in the context of a model. It is not of course the same test as that based on a model with only Z_1 included as a covariate.

Another situation of interest in the multivariate setting is one where we wish to test simultaneously for more than one effect. This situation can come under one of two headings. The first, analogous to an analysis of variance, is where we wish to see if there exists any effect without being particularly concerned about which component or components of the vector Z may be causing the effect. As for an analysis of variance if we reject the global null we would probably wish to investigate further to determine which of the components appears to be the cause. The second is where we use, for the sake of argument, two covariates to represent a single entity, for instance 3 levels of treatment. Testing for whether or not treatment has an impact would require us to simultaneously consider the two covariates defining the groups. We would then consider, for a two variable model, $Z(t)$ is a vector with components $Z_1(t)$ and $Z_2(t)$, step functions with discontinuities at the points X_i, $i=1,\cdots,n$, where they take the values $Z_{1i}(X_i)$ and $Z_{2i}(X_i)$ respectively. For this two dimensional case we consider the increments in the process

$$\int_0^t \left\{\beta_1 Z_1(s)+\beta_2 Z_2(s)\right\}\mathrm{d}\bar{N}(s)$$

at $t=X_i$, having mean

$$\beta_1 E\left(Z_1|X_i\right)+\beta_2 E\left(Z_2|X_i\right)$$

and variance

$$\beta_1^2 V\left(Z_1|X_i\right)+\beta_2^2 V\left(Z_2|X_i\right)+2\beta_1\beta_2 Cov\left(Z_1,Z_2|X_i\right).$$

The remaining steps now follow through just as in the one dimensional case, β_1 and β_2 being replaced by $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively, and the conditional expecta-

tions, variances and covariances being replaced using analogous results to Corollaries 3 and 4.

7. An Example

The classical example studied in Cox's famous 1972 paper [3] concerning the two sample problem in a clinical trial in leukemia has been re-examined by several authors and we reconsider those data in the light of the work here. Our development sidesteps the issue of ties, a problem of sufficient importance for the Freireich data that it warranted lengthy discussion in Cox's paper. Here we simply used a random split, although all of the suggested approaches for dealing with ties (see for example Kalbfleisch and Prentice [13] are accommodated without any additional difficulty. The distance from the origin at the maximum follow up time was equal to 3.92 (p < 0.001), a result which is to be anticipated since the proportional hazards assumption is known to be a good one for these data, and effects are strong. The partial likelyhood test produced the figure 3.94, confirming, at least here, the agreement we would expect given that the estimated conditional variance of the covariate given time is fairly constant with time itself. If the process is reflected at the time point $t=12$, roughly the marginal median, then we obtain the value 0.58 which is what we might well expect. On the other hand if we use the reflected process, maximized across all potential times, then we obtain, again, a p-value less than 0.001 suggesting that little has been sacrificed by this more general approach even when the proportional hazards assumption appears well founded.

REFERENCES

[1] J. O'Quigley, "Khmaladze-Type Graphical Evaluation of the Proportional Hazards Assumption," *Biometrika*, Vol. 90, No. 3, 2003, pp. 577-584. doi:10.1093/biomet/90.3.577

[2] R. Xu and J. O'Quigley, "Proportional Hazards Estimate of the Conditional Survival Function," *Journal of the Royal Statistical Society*: *Series B*, Vol. 62, No. 4, 2000, pp. 667-680. doi:10.1111/1467-9868.00256

[3] D. R. Cox, "Regression Models and Life Tables (with discussion)," *Journal of the Royal Statistical Society*: *Series B*, Vol. 34, 1972, pp. 187-220.

[4] T. Lancaster and S. Nickell, "The Analysis of Re-Employment Probabilities for the Unemployed," *Journal of the Royal Statistical Society*, *Series A*, Vol. 143, No. 2, 1980, pp. 141-165. doi:10.2307/2981986

[5] M. H. Gail, S. Wieand and S. Piantadosi, "Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates," *Biometrika*, Vol. 71, No. 3, 1984, pp. 431-444. doi:10.1093/biomet/71.3.431

[6] J. Bretagnolle and C. Huber-Carol, "Effects of Omitting

Covariates in Cox's Model for Survival Data," *Scandinavian Journal of Statistics*, Vol. 15, 1988, pp. 125-138.

[7] J. O'Quigley and F. Pessione, "Score Tests for Homogeneity of Regression Effect in the Proportional Hazards Model," *Biometrics*, Vol. 45, 1989, pp. 135-144. doi:10.2307/2532040

[8] J. O'Quigley and F. Pessione, "The Problem of a Covariate-Time Qualitative Interaction in a Survival Study," *Biometrics*, Vol. 47, 1991, pp. 101-115. doi:10.2307/2532499

[9] G. L. Anderson and T. R. Fleming, "Model Misspecification in Proportional Hazards Regression," *Biometrika*, Vol. 82, No. 3, 1995, pp. 527-541.

[10] I. Ford, J. Norrie and S. Ahmadi, "Model Inconsistency, Illustrated by the Cox Proportional Hazards Model," *Statistics in Medicine*, Vol. 14, No. 8, 1995, pp. 735-746. doi:10.1002/sim.4780140804

[11] R. Xu and J. O'Quigley, "Estimating Average Regression Effect under Non Proportional Hazards," *Biostatistics*, Vol. 1, 2000, pp. 23-39. doi:10.1093/biostatistics/1.4.423

[12] J. O'Quigley and J. Stare, "Proportional Hazard Models with Frailties and Random Effects," *Statistics in Medicine*, Vol. 21, 2003, pp. 3219-3233. doi:10.1002/sim.1259

[13] J. Kalbfleisch and R. L. Prentice, "The Statistical Analysis of Failure Time Data," Wiley, New York, 1980.

[14] L. J. Wei, "Testing Goodness of Fit for Proportional Hazards Model with Censored Observations," *Journal of the American Statistical Association*, Vol. 79, 1984, pp. 649-652.

[15] D. Y. Lin, L. J. Wei and Z. Ying, "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals," *Biometrika*, Vol. 80, No. 3, 1993, pp. 557-572. doi:10.1093/biomet/80.3.557

[16] D. R. Cox, "Partial Likelihood," *Biometrika*, Vol. 63, 1975, pp. 269-276. doi:10.1093/biomet/62.2.269

[17] P. K. Andersen and R. D. Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study," *Annals of Statistics*, Vol. 10, No. 4, 1982, pp. 1100-1121. doi:10.1214/aos/1176345976

[18] P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.

[19] R. B. Davies, "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, Vol. 64, No. 2, 1977, pp. 247-254. doi:10.2307/2335690

[20] R. B. Davies, "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, Vol. 74, 1987, pp. 33-43.

[21] B. Efron, "Censored Data and the Bootstrap," *Journal of the American Statistical Association*, Vol. 76, No. 374, 1981, pp. 312-319.

[22] B. Efron, "Nonparametric Estimates of Standard Error: The Jacknife, the Bootstrap and Other Resampling Methods," *Biometrika*, Vol. 68, 1981, pp. 589-599. doi:10.1093/biomet/68.3.589

[23] J. O'Quigley and L. Natarajan, "Erosion of Regression Effect in a Survival Study," *Biometrics*, Vol. 60, No. 2, 2004, pp. 344-351. doi:10.1111/j.0006-341X.2004.00178.x