

## Biometrika Trust

---

Confidence Bands for Survival Curves Under the Proportional Hazards Model

Author(s): D. Y. Lin, T. R. Fleming and L. J. Wei

Source: *Biometrika*, Vol. 81, No. 1 (Mar., 1994), pp. 73-81

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2337051>

Accessed: 13/10/2014 13:46

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Biometrika Trust* is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

# Confidence bands for survival curves under the proportional hazards model

By D. Y. LIN, T. R. FLEMING

*Department of Biostatistics, SC-32, University of Washington, Seattle, Washington 98195, U.S.A.*

AND L. J. WEI

*Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.*

## SUMMARY

In this paper, we show how to construct simultaneous confidence bands for the subject-specific survival curve under the Cox proportional hazards model. The idea is to approximate the distribution of the normalized cumulative hazard estimator by a zero-mean Gaussian process whose distribution can be easily generated through simulation. Numerical studies indicate that the proposed bands are appropriate for practical use. A liver disease example is presented.

*Some key words:* Censoring; Counting process; Cox regression; Equal-precision band; Gaussian process; Hall–Wellner band; Martingale; Simultaneous inference.

## 1. INTRODUCTION

The Cox (1972) proportional hazards model specifies that the hazard function  $\lambda(t)$  for the failure time  $T$  under  $Z(t) = z(t)$  has the following form:

$$\lambda(t; z) = \lambda_0(t)e^{\beta_0'z(t)}, \quad (1.1)$$

where  $Z(\cdot)$  is a  $p$ -vector of possibly time-varying covariates,  $\beta_0$  is a  $p$ -vector of unknown regression parameters, and  $\lambda_0(\cdot)$  is an unspecified baseline hazard function. Statistical inference on  $\beta_0$  is usually based on the partial likelihood function (Cox, 1972, 1975)

$$\prod_{i=1}^n \left\{ \frac{e^{\beta'Z_i(X_i)}}{\sum_{j=1}^n Y_j(X_i)e^{\beta'Z_j(X_i)}} \right\}^{\Delta_i}, \quad (1.2)$$

where  $X_i = \min(T_i, C_i)$ ,  $C_i$  is the censoring time variable,  $\Delta_i = I(T_i \leq C_i)$ ,  $Y_i(t) = I(X_i \geq t)$ , and  $I(\cdot)$  is the indicator random variable of the specified event.

One of the main goals in fitting a survival model is to predict survival experience for future subjects. Under model (1.1), we are interested in estimating the cumulative hazard function

$$\Lambda(t; z_0) = \int_0^t \lambda(u; z_0) du$$

and survival function  $S(t; z_0) = e^{-\Lambda(t; z_0)}$  for a subject with a particular set of covariate values  $z_0(\cdot)$ . Note that  $z_0(\cdot)$  should consist only of external covariates (Kalbfleisch & Prentice, 1980, § 5.3) in order for  $S(t; z_0)$  to be interpretable. A commonly used estimator

for

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du$$

is

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \frac{I(X_i \leq t) \Delta_i}{\sum_{j=1}^n Y_j(X_i) e^{\hat{\beta}' Z_j(X_i)}},$$

where  $\hat{\beta}$  is the value of  $\beta$  that maximizes (1.2) (Breslow, 1972). Correspondingly,  $\Lambda(t; z_0)$  is estimated by

$$\hat{\Lambda}(t; z_0) = \sum_{i=1}^n \frac{I(X_i \leq t) \Delta_i e^{\hat{\beta}' z_0(X_i)}}{\sum_{j=1}^n Y_j(X_i) e^{\hat{\beta}' Z_j(X_i)}}, \quad (1.3)$$

and  $S(t; z_0)$  by  $\hat{S}(t; z_0) = e^{-\hat{\Lambda}(t; z_0)}$ .

For any given time point  $t$ , one may construct confidence intervals for  $\Lambda(t; z_0)$  and  $S(t; z_0)$  using the asymptotic properties of  $\hat{\Lambda}(t; z_0)$  and  $\hat{S}(t; z_0)$ . In many applications, however, it is desirable to obtain simultaneous confidence bands for  $\Lambda(\cdot; z_0)$  or  $S(\cdot; z_0)$  so that a single probability statement can be made regarding the survival experience over the entire time span of interest. In the one-sample case, where no covariates are involved, such bands have been extensively studied by various authors, including Hall & Wellner (1980), Nair (1984) and Bie, Borgan & Liestøl (1987), and described at great length in the recent texts of Fleming & Harrington (1991, § 6.3) and Andersen et al. (1993, §§ IV.1.3, IV.3.3). The existing one-sample bands rely on the fact that the normalized Nelson–Aalen estimator for the cumulative hazard function converges weakly to a zero-mean Gaussian martingale, which can be transformed to the standard Brownian bridge. As will be seen in the next section, under model (1.1), the process  $n^{\frac{1}{2}} \{\hat{\Lambda}(\cdot; z_0) - \Lambda(\cdot; z_0)\}$  also converges weakly to a zero-mean Gaussian process. Unlike the one-sample case, however, this limiting process does not possess an independent increment structure and therefore cannot be transformed to the standard Brownian bridge.

In this paper, we develop confidence bands for  $\Lambda(\cdot; z_0)$  and  $S(\cdot; z_0)$  under the Cox model by simulating a Gaussian process which approximates the distribution of the process  $n^{\frac{1}{2}} \{\hat{\Lambda}(\cdot; z_0) - \Lambda(\cdot; z_0)\}$ . This approach is described in the next section. In § 3, we assess the performance of the proposed bands for practical sample sizes and then apply them to the Mayo Clinic data on primary biliary cirrhosis patients.

## 2. CONSTRUCTION OF CONFIDENCE BANDS

Suppose that  $\{T_i, C_i, Z_i(\cdot)\}$  ( $i = 1, \dots, n$ ) are independent and identically distributed. Assume that  $T_i$  and  $C_i$  are independent conditional on  $Z_i(\cdot)$  and that  $Z_i(\cdot)$  is bounded. It is convenient to introduce the notation

$$S^{(r)}(\beta, t) = n^{-1} \sum_{j=1}^n Y_j(t) e^{\beta' Z_j(t)} Z_j^{\otimes r}(t), \quad s^{(r)}(\beta, t) = E\{S^{(r)}(\beta, t)\},$$

$$\bar{Z}(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}, \quad \bar{z}(\beta, t) = \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)},$$

for  $r = 0, 1, 2$ , where, for a column vector  $a$ ,  $a^{\otimes 0} = 1$ ,  $a^{\otimes 1} = a$  and  $a^{\otimes 2} = aa'$ . Note that the

asymptotic covariance matrix of  $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$  is

$$\Omega = \int_0^\infty \{s^{(2)}(\beta_0, t)/s^{(0)}(\beta_0, t) - \bar{z}(\beta_0, t)^{\otimes 2}\} s^{(0)}(\beta_0, t) \lambda_0(t) dt,$$

which as usual is assumed to be positive definite.

Define the counting processes  $N_i(t) = \Delta_i I(X_i \leq t)$  and martingales

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0' Z_i(u)} d\Lambda_0(u) \quad (i = 1, \dots, n).$$

In the counting process notation, the estimator (1.3) becomes

$$\hat{\Lambda}(t; z_0) = \sum_{i=1}^n \int_0^t \frac{e^{\hat{\beta}' z_0(u)} dN_i(u)}{nS^{(0)}(\hat{\beta}, u)}.$$

Now, let

$$W(t; z_0) = n^{\frac{1}{2}} \{ \hat{\Lambda}(t; z_0) - \Lambda(t; z_0) \}, \quad \tau < \inf \{t : EY_1(t) = 0\}.$$

By the arguments given by Andersen & Gill (1982), the process  $W(t; z_0)$  ( $0 \leq t \leq \tau$ ) is asymptotically equivalent to

$$\tilde{W}(t; z_0) = n^{-\frac{1}{2}} \sum_{i=1}^n \left[ \int_0^t \frac{e^{\beta_0' z_0(u)} dM_i(u)}{S^{(0)}(\beta_0, u)} + h'(t; z_0) \Omega^{-1} \int_0^\infty \{Z_i(u) - \bar{Z}(\beta_0, u)\} dM_i(u) \right], \quad (2.1)$$

where

$$h(t; z_0) = \int_0^t e^{\beta_0' z_0(u)} \{z_0(u) - \bar{z}(\beta_0, u)\} d\Lambda_0(u).$$

It then follows from the martingale central limit theorem that the process  $\tilde{W}(\cdot; z_0)$  and thereby  $W(\cdot; z_0)$  converges weakly to a zero-mean Gaussian process on  $[0, \tau]$  with covariance function

$$\xi(t, s; z_0) = \int_0^{t \wedge s} \frac{e^{2\beta_0' z_0(u)} d\Lambda_0(u)}{s^{(0)}(\beta_0, u)} + h'(t; z_0) \Omega^{-1} h(s; z_0), \quad (2.2)$$

where  $t \wedge s = \min(t, s)$ . Denote  $\xi(t, t; z_0)$  by  $\sigma^2(t; z_0)$ , which can be consistently estimated by

$$\hat{\sigma}^2(t; z_0) = n^{-1} \sum_{i=1}^n \frac{I(X_i \leq t) \Delta_i e^{2\hat{\beta}' z_0(X_i)}}{S^{(0)}(\hat{\beta}, X_i)^2} + \hat{h}'(t; z_0) \hat{\Omega}^{-1} \hat{h}(t; z_0),$$

where

$$\begin{aligned} \hat{\Omega} &= n^{-1} \sum_{i=1}^n \Delta_i \{S^{(2)}(\hat{\beta}, X_i)/S^{(0)}(\hat{\beta}, X_i) - \bar{Z}(\hat{\beta}, X_i)^{\otimes 2}\}, \\ \hat{h}(t; z_0) &= n^{-1} \sum_{i=1}^n I(X_i \leq t) \Delta_i e^{\hat{\beta}' z_0(X_i)} \{z_0(X_i) - \bar{Z}(\hat{\beta}, X_i)\} / S^{(0)}(\hat{\beta}, X_i). \end{aligned}$$

As is evident from (2.2), the process  $W(\cdot; z_0)$  does not have an independent increment

structure asymptotically even if covariates are time-invariant; therefore, it is difficult to evaluate the limiting distribution analytically.

We now show how to approximate the distribution of  $W(\cdot; z_0)$  by a zero-mean Gaussian process, denoted by  $\hat{W}(\cdot; z_0)$ , whose distribution can be easily generated through simulation. This approximation is based on the facts that

$$E\{M_i(u)\} = 0, \quad \text{var}\{M_i(u)\} = E\{N_i(u)\} \quad (i = 1, \dots, n).$$

Specifically, we replace  $\{M_i(u)\}$  ( $i = 1, \dots, n$ ) in (2.1) with  $\{N_i(u)G_i\}$  ( $i = 1, \dots, n$ ), where  $\{G_i\}$  ( $i = 1, \dots, n$ ) are independent standard normal variables, and also replace other unknown quantities in (2.1) with their respective sample estimators to yield

$$\hat{W}(t; z_0) = n^{-\frac{1}{2}} \sum_{i=1}^n \left[ \frac{I(X_i \leq t) \Delta_i e^{\hat{\beta}' z_0(X_i)} G_i}{S^{(0)}(\hat{\beta}, X_i)} + \hat{h}'(t; z_0) \hat{\Omega}^{-1} \Delta_i \{Z_i(X_i) - \bar{Z}(\hat{\beta}, X_i)\} G_i \right].$$

We regard  $\{G_i\}$  ( $i = 1, \dots, n$ ) as random and  $\{X_i, \Delta_i, Z_i(\cdot)\}$  ( $i = 1, \dots, n$ ) as fixed in  $\hat{W}(\cdot; z_0)$ . The process  $\hat{W}(\cdot; z_0)$ , which consists of a sum of  $n$  independent random variables at each fixed time point, can be shown to converge weakly to a zero-mean Gaussian process by applying the Lindeberg–Feller theorem and by verifying a tightness criterion (Billingsley, 1968, p. 128). Furthermore,

$$\begin{aligned} E[\hat{W}(t; z_0) \hat{W}(s; z_0) | \{X_i, \Delta_i, Z_i(\cdot)\}] \\ = n^{-1} \sum_{i=1}^n \frac{I(X_i \leq t \wedge s) \Delta_i e^{2\hat{\beta}' z_0(X_i)}}{S^{(0)}(\hat{\beta}, X_i)^2} \\ + \hat{h}'(t; z_0) \hat{\Omega}^{-1} n^{-1} \sum_{i=1}^n \Delta_i \{Z_i(X_i) - \bar{Z}(\hat{\beta}, X_i)\} \otimes^2 \hat{\Omega}^{-1} \hat{h}(s; z_0) \\ + \hat{h}'(s; z_0) \hat{\Omega}^{-1} n^{-1} \sum_{i=1}^n \frac{I(X_i \leq t) \Delta_i e^{\hat{\beta}' z_0(X_i)}}{S^{(0)}(\hat{\beta}, X_i)} \{Z_i(X_i) - \bar{Z}(\hat{\beta}, X_i)\} \\ + \hat{h}'(t; z_0) \hat{\Omega}^{-1} n^{-1} \sum_{i=1}^n \frac{I(X_i \leq s) \Delta_i e^{\hat{\beta}' z_0(X_i)}}{S^{(0)}(\hat{\beta}, X_i)} \{Z_i(X_i) - \bar{Z}(\hat{\beta}, X_i)\}. \end{aligned}$$

The first two terms on the right-hand side converge almost surely to (2.2) and the remaining two vanish due to the strong consistency of  $\hat{\beta}$  and  $\hat{\Lambda}_0(\cdot)$  and the fact that  $Y_i(t) e^{\beta_0' Z_i(t)} \lambda_0(t)$  is the intensity function of  $N_i(t)$ . Hence,  $\hat{W}(\cdot; z_0)$  and  $W(\cdot; z_0)$  have the same limiting distribution. To approximate the distribution of  $W(\cdot; z_0)$ , we simply obtain a large number of realizations from  $\hat{W}(\cdot; z_0)$  by repeatedly generating normal random samples  $\{G_i; i = 1, \dots, n\}$  while fixing the data  $\{X_i, \Delta_i, Z_i(\cdot)\}$  ( $i = 1, \dots, n$ ) at their observed values.

The aforementioned Monte Carlo scheme is similar to the bootstrap in terms of sample space, convergence, etc. Specifically, the sample space for  $\hat{W}(\cdot; z_0)$  is conditional on the data whereas that of  $W(\cdot; z_0)$  is unconditional; the distribution of  $\hat{W}(\cdot; z_0)$  converges to the same limit as that of  $W(\cdot; z_0)$  for almost all realizations of the data  $\{X_i, \Delta_i, Z_i(\cdot)\}$  ( $i = 1, \dots, n$ ). As one referee pointed out, bootstrap would be a possible alternative to the proposed approach. However, it is not clear how to justify that the bootstrap is valid for the present problem.

For constructing confidence bands for  $\Lambda_0(\cdot; z_0)$  and  $S(\cdot; z_0)$ , it is fruitful to consider the class of transformed processes

$$B(t; z_0) = n^{\frac{1}{2}} g(t; z_0) [\phi\{\hat{\Lambda}(t; z_0)\} - \phi\{\Lambda(t; z_0)\}],$$

where  $\phi$  is a known function whose derivative  $\phi'$  is continuous and nonzero in the time interval  $[t_1, t_2]$  ( $0 \leq t_1 \leq t_2 \leq \tau$ ), and the weight function  $g(\cdot; z_0)$  converges in probability to a nonnegative bounded function uniformly on  $[t_1, t_2]$ . By the functional delta-method (Andersen et al., 1993, § II.8), the process  $B(t; z_0)$  is asymptotically equivalent to  $g(t; z_0)\phi'\{\hat{\Lambda}(t; z_0)\}W(t; z_0)$ , whose distribution can be approximated by that of

$$\hat{B}(t; z_0) = g(t; z_0)\phi'\{\hat{\Lambda}(t; z_0)\}\hat{W}(t; z_0).$$

Let  $q_\alpha$  be the boundary value satisfying  $\text{pr}\{\max_{t_1 \leq x_j \leq t_2} |\hat{B}(X_j; z_0)| > q_\alpha\} = \alpha$ , the probability being estimated through simulation. Then an approximate  $(1 - \alpha)$  confidence band for  $\phi\{\Lambda(t; z_0)\}$  on  $[t_1, t_2]$  is  $\phi\{\hat{\Lambda}(t; z_0)\} \mp n^{-\frac{1}{2}}q_\alpha/g(t; z_0)$ .

One may calculate the confidence bands for  $\Lambda(\cdot; z_0)$  and  $S(\cdot; z_0)$  directly by letting  $\phi(x) = x$  and  $\phi(x) = e^{-x}$ , respectively. The resulting band for  $\Lambda(\cdot; z_0)$  may include negative values, and that of  $S(\cdot; z_0)$  may contain values outside  $[0, 1]$ . As indicated by Kalbfleisch & Prentice (1980, pp. 14–5), this problem can be avoided by choosing the log transformation  $\phi(x) = \log x$ , which not only restricts the bands for  $\Lambda(\cdot; z_0)$  and  $S(\cdot; z_0)$  to meaningful ranges but also improves the attained coverage probabilities in small samples. This log transformation will be used implicitly in the sequel unless indicated otherwise.

The choice of the weight function  $g(\cdot; z_0)$  affects the relative widths of the band at different time points. We shall confine our attention to two weight functions

$$g_1(t; z_0) = \hat{\Lambda}(t; z_0)/\hat{\sigma}(t; z_0), \quad g_2(t; z_0) = \hat{\Lambda}(t; z_0)/\{1 + \hat{\sigma}^2(t; z_0)\}.$$

The resulting  $(1 - \alpha)$  bands for  $S(t; z_0)$  are

$$\hat{S}(t; z_0)^{\exp\{\pm n^{-1/2}q_{1,\alpha}\hat{\sigma}(t; z_0)/\hat{\Lambda}(t; z_0)\}}, \quad (2.3)$$

$$\hat{S}(t; z_0)^{\exp[\pm n^{-1/2}q_{2,\alpha}\{1 + \hat{\sigma}^2(t; z_0)\}/\hat{\Lambda}(t; z_0)]}, \quad (2.4)$$

where  $q_{1,\alpha}$  and  $q_{2,\alpha}$  are the boundary values associated with  $g_1$  and  $g_2$ . Expression (2.3) is the so-called equal-precision band (Nair, 1984), whose bounds are proportional to the pointwise confidence limits on the log  $\Lambda$  scale since  $n^{-\frac{1}{2}}\hat{\sigma}(t; z_0)/\hat{\Lambda}(t; z_0)$  is the standard error estimator for  $\log \hat{\Lambda}(t; z_0)$ . Note that (2.3) is in the same form as (1.12) of Kalbfleisch & Prentice (1980, p. 15), the key difference being that the standard normal critical value is now replaced by a larger number in order to achieve a proper simultaneous coverage probability. Expression (2.4) corresponds to the one-sample transformed Hall–Wellner band advocated by Bie et al. (1987). Based on the direct approximation for  $\hat{S}(\cdot; z_0)$ , the Hall–Wellner type band takes the form

$$\hat{S}(t; z_0) \mp n^{-\frac{1}{2}}q_{2,\alpha}\hat{S}(t; z_0)\{1 + \hat{\sigma}^2(t; z_0)\}. \quad (2.5)$$

In the absence of covariates, (2.5) is asymptotically equivalent to the original Hall–Wellner band, the latter reducing to the well-known Kolmogorov band for completely observed survival data. We shall refer to (2.4) and (2.5) as the transformed and untransformed Hall–Wellner bands. Because the approximations tend to be poor for  $t$  close to 0 or  $\tau$ , we shall confine all our bands between the first and last uncensored failure time points. Additional restrictions may be necessary for the equal-precision band. Following the recommendations of Nair (1984) and Bie et al. (1987) for the one-sample case, we shall further restrict the equal-precision band to the time interval  $[t_1^*, t_2^*]$  such that  $\hat{c}_1 = 1 - \hat{c}_2 = 0.05$  or 0.1, where

$$\hat{c}_k = \hat{\sigma}^2(t_k^*; z_0)/\{1 + \hat{\sigma}^2(t_k^*; z_0)\} \quad (k = 1, 2).$$

The confidence bands (2.3)–(2.5) can be easily incorporated into an existing statistical package. The FORTRAN and S codes are available from the first author.

3. NUMERICAL RESULTS

3.1. Simulation studies

A series of simulation experiments was conducted to evaluate the confidence bands (2.3)–(2.5). Failure times were generated from an exponential model with a standard normal covariate truncated at  $\pm 5$ , and censoring times from uniform distributions. The empirical coverage probabilities were estimated from 2000 simulation samples; for each simulation sample, the boundary values  $q_{1,\alpha}$  and  $q_{2,\alpha}$  were calculated from 1000 realizations of  $\hat{B}(\cdot; z_0)$ . The results are summarized in Table 1. The proposed bands, especially the equal-precision band with  $\hat{c}_1 = 0.1$  and the transformed Hall–Wellner band, maintain their coverage probabilities near the nominal level, even for sample size of 50 with heavy censoring. As one might expect,  $\hat{c}_1 = 0.1$  results in better coverage probabilities than  $\hat{c}_1 = 0.05$  in small samples. Results not shown in the table indicate that the collection of pointwise confidence intervals for  $S(\cdot; z_0)$  has simultaneous coverage probabilities of about 0.6 over the time interval for which  $\hat{c}_1 = 0.1$ .

Table 1. Empirical coverage probabilities of confidence bands for  $S(\cdot; z_0)$  with 0.95 nominal confidence coefficient under the model  $\lambda(t; z) = e^{0.3z}$

n	z <sub>0</sub>	25% censoring				50% censoring				75% censoring			
		EP		HW		EP		HW		EP		HW	
		(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
50	−1	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.97	0.96	0.97	0.97	0.96
	0	0.90	0.91	0.92	0.92	0.90	0.92	0.93	0.91	0.90	0.92	0.93	0.90
	1	0.92	0.95	0.95	0.93	0.92	0.94	0.94	0.92	0.91	0.93	0.94	0.90
100	−1	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97
	0	0.92	0.93	0.94	0.94	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.90
	1	0.95	0.96	0.96	0.94	0.94	0.96	0.96	0.96	0.95	0.95	0.95	0.91
200	−1	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	0	0.91	0.92	0.94	0.93	0.93	0.93	0.94	0.92	0.93	0.94	0.94	0.92
	1	0.94	0.95	0.95	0.95	0.95	0.96	0.96	0.95	0.95	0.95	0.96	0.94

For EP bands, namely equal-precision bands,  $\hat{c}_1 = 1 - \hat{c}_2 = 0.05$  and 0.1 for columns (a) and (b) respectively. For HW bands, namely Hall–Wellner bands, columns (a) and (b) are the transformed and untransformed ones, respectively.

3.2. Mayo primary biliary cirrhosis model

The Mayo Clinic developed a database for the primary biliary cirrhosis patients who were referred to Mayo between January 1974 and May 1984. Primary biliary cirrhosis is a fatal chronic liver disease of unknown cause. Because this is a rare disease, the Mayo database has been a valuable resource to liver specialists. These data are listed in Appendix D of Fleming & Harrington (1991).

The Cox regression method and comprehensive data from the Mayo patients were used successfully by Dickson et al. (1989) to derive a natural history model for primary biliary cirrhosis based on patients’ age, total serum bilirubin and serum albumin concentrations, prothrombin time and the severity of oedema. The original model was built from 312



patients who participated in a clinical trial where the test treatment was found to be ineffective. When it was cross-validated on the independent set of 106 patients concurrently treated at Mayo who did not participate in the clinical trial, the model predicted survival accurately. Table 2 provides the variable transformations and parameter estimates for the final model with the 418 Mayo patients, which has been validated on data generated outside the Mayo Clinic (Grambsch et al., 1989). These results have been inserted into expression (1.3) to predict survival for individual patients.

The Mayo model has been extremely useful not only in counselling patients and in understanding the course of the disease in untreated patients, but also in providing historical control information to evaluate the efficacy of new therapeutic interventions such as liver transplantation. Dickson and other hepatologists have expressed great interest in attaching confidence bands to the survival curve estimates derived from the Mayo model so that proper probability statements could be made.

Table 2. Regression parameter estimates for the final Mayo primary biliary cirrhosis model

Variable	Parameter estimate	Standard error	Est./SE
Age	0.0394	0.0077	5.1508
log (Albumin)	-2.5328	0.6482	-3.9074
log (Bilirubin)	0.8707	0.0826	10.5372
Oedema	0.8592	0.2711	3.1688
log (Prothrombin time)	2.3797	0.7666	3.1043

For illustration, Fig. 1 displays the equal-precision and Hall–Wellner bands with  $\alpha = 0.05$  for a patient with 51 years of age, 3.4 gm/dl serum albumin, 1.8 mg/pl serum bilirubin, 10.74 seconds of prothrombin time and no oedema. The estimated risk score  $e^{\hat{\beta}'z_0}$  for this patient equals 5.07, which is the sample mean of the estimated risk scores for the 418 Mayo patients. As shown in Fig. 1(a), the choice of  $\hat{c}_1 = 0.05$  results in the restriction of the equal-precision band to the time interval of 2 to 11.5 years. Figure 1(a) also presents the ‘unrestricted’ equal-precision band, which covers the interval between the first death time and the last observation time. The boundary values for the restricted and unrestricted equal-precision bands were found to be 2.879 and 3.024 based on 10 000 realizations of  $\hat{B}(\cdot; z_0)$ , indicating that the equal-precision bands are about 50% wider than the pointwise confidence limits, as can be seen in Fig. 1(a). The lower bound of the transformed Hall–Wellner band is very low for the first two years due to extremely small values of  $\hat{\Lambda}$ ; however, in Fig. 1(b) we use the lower bound at  $t = 2$  years for  $t < 2$  years since  $S(\cdot; z_0)$  is monotone. The untransformed Hall–Wellner band is slightly above the transformed one until year 10, and its upper bound extends above 1 for the first two years, which would be set to 1 in a final presentation. The boundary value for the Hall–Wellner bands was estimated at 1.276, which is smaller than the Brownian bridge critical value 1.358. By overlaying Fig. 1(b) over Fig. 1(a), one can observe that the Hall–Wellner bands are narrower than the equal-precision bands over the middle region of the data and wider at the two tails. The widths of the four confidence bands shown in the figure are reasonably small and fairly close to one another between the second and the tenth years.

4. REMARKS

As mentioned earlier, the choice of the weight function  $g(\cdot; z_0)$  affects the relative widths of the band at different time points. From a practical point of view, one would choose a



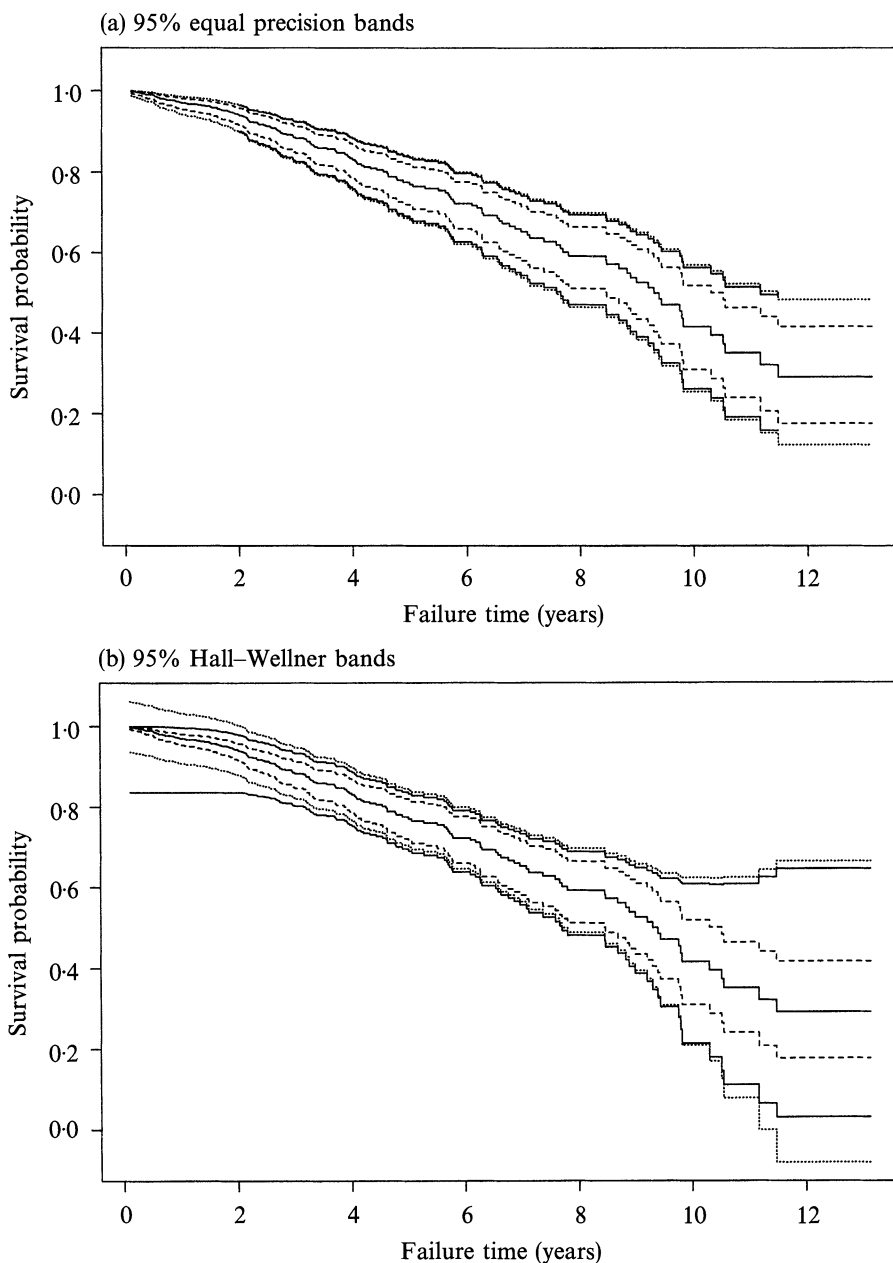


Fig. 1. Survival probability estimation for a primary biliary cirrhosis patient with 51 years of age, 3.4 gm/dl albumin, 1.8 mg/pl bilirubin, 10.74 seconds of prothrombin time and no oedema. Point estimate, shown by the middle solid curve; pointwise 95% confidence intervals, the dashed curves. (a) 95% equal-precision band with  $\hat{c}_1 = 1 - \hat{c}_2 = 0.05$ , the outside solid curves; and 95% unrestricted equal-precision band, the dotted curves. (b) 95% transformed Hall-Wellner band, the outside solid curves; and 95% untransformed Hall-Wellner band, the dotted curves.

weight function which yields the narrowest band for the part of the time interval that is of the most interest. In general, the Hall-Wellner bands tend to be narrower than the equal-precision bands over the middle region of the data and wider at the two tails.

Obviously, one may tighten the equal-precision band by shortening the time interval. For both types of bands, the log transformation is recommended.

The basic ideas developed in this paper may also be used to construct confidence bands for other survival models. A related topic is the goodness-of-fit test for the distributional assumption of a fully parametric regression model. A natural goodness-of-fit process is the difference between the Aalen–Breslow type estimator and the parametric maximum likelihood estimator for the baseline survival function. Such a process has a somewhat more complicated limiting covariance structure than  $W(.; z_0)$ , but can be expressed as sums of martingale integrals so that the basic principle described in § 2 still applies. The details will be presented elsewhere.

#### ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health. The authors are grateful to Drs E. Rolland Dickson and Zhiliang Ying for helpful discussions, and to three referees for their timely reviews.

#### REFERENCES

- ANDERSEN, P. K., BORGAN, Ø, GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–20.
- BIE, O., BORGAN, Ø. & LIESTØL, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scand. J. Statist.* **14**, 221–33.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: Wiley.
- BRESLOW, N. (1972). Discussion of paper by D. R. Cox. *J. R. Statist. Soc. B* **34**, 216–7.
- COX, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B* **34**, 187–220.
- COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.
- DICKSON, E. R., GRAMBSCH, P. M., FLEMING, T. R., FISHER, L. D. & LANGWORTHY, A. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* **10**, 1–7.
- FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- GRAMBSCH, P. M., DICKSON, E. R., KAPLAN, M., LESAGE, G., FLEMING, T. R. & LANGWORTHY, A. (1989). Extramural cross-validation of the Mayo primary biliary cirrhosis survival model establishes its generalizability. *Hepatology* **10**, 846–50.
- HALL, W. J. & WELLNER, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–43.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- NAIR, V. N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265–75.

[Received February 1993. Revised August 1993]