Methods for Comparing Cumulative Hazard Functions in a Semi-Proportional Hazard Model

Dorota M. Dabrowska[1,2]
Department of Biostatistics, School of Public Health
University of California, Los Angeles, CA 90024-1772
(213)-825-5250

Kjell A. Doksum[1,3,4]
Department of Statistics
University of California, Berkeley, CA 94720
(510)-642-2718

Nicholas J. Feduska
Pacific Transplant Institute at Presbyterian Medical Center,
2340 Clay Street - Suite 417, San Francisco, CA 94115
(415)-923-3450

Robert Husing
Renal Preservation Technology
University of California, San Francisco, CA 94143
(415) 476-1551

Padraic Neville[3]
Department of Statistics
University of California, Berkeley, CA 94720
(510)-787-7255

[1]Research partially supported by NSF Grant DMS-8901603
[2]Research partially supported by NIH Grant USHHS-CA 16042
[3]Research partially supported by NIH through a
Biomedical Research Support Grant at U.C., Berkeley
[4]Research partially carried out in the
Department of Statistics, Harvard University

Technical Report No. 321
August 1991
(revised Feb. 1992)

Department of Statistics
University of California
Berkeley, California 94720

Methods for Comparing Cumulative Hazard Functions in a Semi-Proportional Hazard Model

Abstract. Graphical methods based on the analysis of differences between log cumulative hazard functions are considered for a two-group semi-proportional hazard model which allows for interaction between treatments and covariates. Confidence procedures and test statistics that can be used to test for interaction, for main effects, and for proportional hazards, are developed. Their use is illustrated by applying them to the analysis of kidney transplant data from the University of California, San Francisco.

Key words and phrases: Cox regression, stratification, semi-proportional hazard model, interaction, cumulative hazard function, graphical methods.

1. INTRODUCTION

1.1 The model

Consider a study where patients are grouped according to a certain characteristic. We discuss two cases: (i) grouping based on treatment group membership and (ii) grouping based on stratification of a covariate. We consider a *semi-proportional* hazard model where a subject in group i with covariate vector **z** has cumulative hazard function

$$\Lambda_i(t|\mathbf{z}) = \Lambda_{0i}(t)\exp\{\beta_i^T\mathbf{z}\}, \quad i = 1,2 \tag{1.1}$$

where β_1 and β_2 are vectors of unknown regression coefficients and $\Lambda_{01}(t)$ and $\Lambda_{02}(t)$ are unknown baseline cumulative hazard functions for the two groups. In this paper we discuss graphical methods based on the analysis of differences between log cumulative hazard functions, and give confidence procedures and test statistics that can be used to test for interaction and main effects. As an illustration we apply some of the results to the analysis of kidney transplant data from the University of California, San Francisco.

Model (1.1) allows for $\beta_1 \neq \beta_2$ and thereby makes it possible to analyse interaction between treatments and covariates as well as between covariates. Thall and Lachin[1] and Andersen, Borgan, Gill and Keiding[2] present models which include (1.1) as special cases. Kay[3], Kalbfleisch and Prentice[4], Therneau, Grambsch and Fleming[5], and Kronborg and Aaby[6], among others, consider model (1.1) with $\beta_1 = \beta_2 = \beta$.

Model (1.1) can be used to address three related questions:

(a) *Interaction.* Suppose that the two groups correspond to treatment groups 1 and 2. In this case, model (1.1) corresponds to proportional hazards with respect to the covariates, but (possibly) non-proportional hazards with respect to the treatments. Note that $\beta_1 \neq \beta_2$ corresponds to interaction between treatment and covariates. This is important when treatment efficacy depends on the covariate values of the patient. We return to the question of how to measure this interaction in subsection 1.2.

(b) *Treatment Comparison.* The four parameters β_1, β_2, $\Lambda_{01}(t)$ and $\Lambda_{02}(t)$ on the right hand side of (1.1) can be combined to give a description of the relative performance of two treatments at time t for a patient with given covariate values. We return to this in the next subsection.

(c) *Proportional hazards.* Suppose that *within* a treatment group we define groups by stratifying one of the covariates, say z_1. In this case, in model (1.1), **z** stands for the vector of the remaining covariates and model (1.1) with $\Lambda_{01}(t) = \Lambda_{02}(t)$ corresponds to a proportional hazard model with respect to the covariate used to create the strata. Similarly, we can check for proportional hazards between two treatments by letting the

groups correspond to the two treatments and letting $\mathbf{z}$ be the vector of all covariates, as in (a) above. In many studies, hazards are proportional, but here are two possible types of deviations from the proportional hazard model:

(i) *Decaying treatment effect* refers to the case where initially the new treatment is superior to the old treatment but after a period of say several months, the ratio of the cumulative hazards corresponding to the two treatments tends towards one. A decaying treatment effect was found in our analysis of the kidney-transplant data (see Section 3) and in the analysis of malignant melanoma data by Andersen, Borgan, Gill and Keiding[2] (1992, Examples 7.3.1 and 7.3.4). These authors showed that an extension of the proportional hazard model based on introducing a frailty parameter resulted in a good model fit. A counting process approach to the analysis of the resulting frailty model is given by Nielsen, Gill, Andersen and Sorensen[7] as well as Andersen, Borgan, Gill and Keiding[2].

(ii) *Diverging treatment effect* refers to the case where initially there is no difference between two groups, but after a period of say several months, the hazard experiences of the two groups start to diverge. An instance of gradual "treatment" effect is found in the analysis of ex-smoker (treatment 1) and current smoker (treatment 2) data Doll[8], Freedman and Navidi[9]). Here the lung cancer hazard rate for a group of smokers who quit smoking was initially the same as that of a group that continued smoking, but gradually the rate for the ex-smokers started decreasing. It then turned around and began increasing, but stayed well below that of the group that continued smoking. This is the opposite of the decaying treatment effect in that the ratio of the cumulative hazards starts at one and then gradually decreases until it stabilizes at a constant below one.

Examples (i) and (ii) are treatment group examples. For groups determined by covariates such as age and gender, the effects described in (i) and (ii) above are perhaps less likely and the proportional hazards model may be more plausable.

1.2 The log cumulative hazard difference

A well known and useful approach for checking the proportionality of hazards between strata is plotting the log cumulative hazards.[2,3,4,6,10] If the estimated log cumulative hazards for two strata appear nearly parallel, this is evidence in favor of the proportional hazard model. Here we consider a modification of this procedure consisting of plotting the difference between these two curves, that is, for fixed $\mathbf{z}$ we plot an estimate of the curve $\rho(t|\mathbf{z})$ defined by

$$\rho(t|\mathbf{z}) = \log \Lambda_1(t|\mathbf{z}) - \log \Lambda_2(t|\mathbf{z})$$

where $\Lambda_1(t|\mathbf{z})$ and $\Lambda_2(t|\mathbf{z})$ are as in model (1.1). There are two advantages to this

modification: (i) if the proportional hazard model holds, $\rho(t|\mathbf{z})$ is constant in t, and it is easier to see whether a curve is constant than to check whether the two nonlinear log cumulative hazard curves are parallel; (ii) by giving confidence procedures for $\rho(t|\mathbf{z})$ we can do statistical inference. In particular, by using test statistics based on an estimate $\hat{\rho}(t|\mathbf{z})$ of $\rho(t|\mathbf{z})$, we can test for interaction and main effects.

Next we consider the application of $\rho(t|\mathbf{z})$ to the three questions in section 1.1.

(a) *Interaction.* Consider model (1.1) with groups 1 and 2 corresponding to the treatment groups 1 and 2. Note that for a patient with covariate vector $\mathbf{z}$

$$\rho(t|\mathbf{z}) = \rho_0(t) + (\beta_1 - \beta_2)^T \mathbf{z} \tag{1.2}$$

where $\rho_0(t) = \log \Lambda_{01}(t) - \log \Lambda_{02}(t)$ is a baseline log cumulative hazard difference and $(\beta_1 - \beta_2)^T \mathbf{z}$ is an interaction term. If the hazards are proportional also with respect to treatments, say $\Lambda_{01}(t) = \exp\{\theta\}\Lambda_{02}(t)$, then $\rho(t|\mathbf{z}) = \theta + (\beta_2 - \beta_1)\mathbf{z}$, which is constant in t. By plotting an estimate of $\rho(t|\mathbf{z})$ as a function of t we get a check of proportional hazards with respect to treatments, and by looking at such plots for two or more $\mathbf{z}$, we get a check of whether there is an interaction effect. By testing $H_0: \beta_1 = \beta_2$ we get a statistical test of whether there is interaction. Thall and Lachin[1] use partial likelihood ratio methods to analyse interaction effects in semi-proportional hazard models.

(b) *Main effects.* When comparing two treatment groups we can get a confidence interval for the difference $\rho(t|\mathbf{z})$ of log cumulative hazards by considering a pivot of the form $|\hat{\rho}(t|\mathbf{z}) - \rho(t|\mathbf{z})| / \hat{\sigma}(t|\mathbf{z})$ where $\hat{\rho}(t|\mathbf{z})$ is an estimate of $\rho(t|\mathbf{z})$ obtained by substituting estimates of β_1, β_2 and $\rho_0(t)$ into the right hand side of (1.2), and $\hat{\sigma}(t|\mathbf{z})$ is an estimate of the standard error of $\hat{\rho}(t|\mathbf{z})$. Thus we can test whether for a person with covariate vector $\mathbf{z}$, one treatment is better than another.

(c) *Proportional hazards.* If we have strata determined by stratifying on one covariate, say z_1, and $\mathbf{z}$ represents the other covariates, then the proportional hazard assumption with respect to covariate z_1 corresponds to $\rho(t|\mathbf{z})$ being constant in t. Another approach to checking the proportional hazard assumption in the $\beta_1 = \beta_2$ case is based on residuals. See Kay[3], Therneau, Grambsch and Fleming[5], and Fleming and Harrington[10].

In Section 2 we develop the statistical inference methods to accompany the graphical approach. In Section 3 we illustrate some of our methods using survival kidney time data from a kidney transplant study at the University of California, San Francisco. We consider several immunosuppressive treatment regimes and a covariate z_B which gives the number of transfused units of blood the transplant recipients had received prior to transplant. By applying the log cumulative hazard difference we illustrate significant interaction between treatment choice and the covariate z_B. This has

important implications for treatment assignment since for a given patient the value of z_B is known prior to transplant and this value determines which treatment is preferable. In Section 3 we also illustrate proportional hazards within treatment groups with respect to the covariate z_B, and we illustrate the analysis of treatment effects in the semi-proportional hazard model (1.1).

Besides $\rho(t|\mathbf{z})$, another closely related function which also can be used to check for interaction, main effects and proportionality of hazards is the relative difference $\Delta(t|\mathbf{z})$ of cumulative hazards, that is

$$\Delta(t|\mathbf{z}) = [\Lambda_1(t|\mathbf{z}) - \Lambda_2(t|\mathbf{z})]/\Lambda_2(t|\mathbf{z}).$$

Note that $\Delta(t|\mathbf{z}) = \exp\{\rho(t|\mathbf{z})\} - 1$, so that the $\Delta(t|\mathbf{z})$ and $\rho(t|\mathbf{z})$ carry the same information. $\Delta(t|\mathbf{z})$ is an extension to the case of covariates of the function $\Delta(t)$ considered by Dabrowska, Doksum and Song[11]. In this paper we consider $\rho(t|\mathbf{z})$ rather than $\Delta(t|\mathbf{z})$ since it is "symmetric" in the sense of remaining the same, except for the sign, if the labeling of the two groups is reversed. Moreover, the results of Bie, Borgan and Liestól[12] indicate that the asymptotic approximations are more accurate when using log cumulative hazards. The results for $\Delta(t|\mathbf{z})$ are very similar to the results for $\rho(t|\mathbf{z})$.

2. STATISTICAL INFERENCE FOR THE LOG CUMULATIVE HAZARD DIFFERENCE(LCHD) IN THE SEMI-PROPORTIONAL HAZARD MODEL.

We consider an estimate of the LCHD $\rho(t|\mathbf{z})$ in model (1.1), and find the asymptotic distribution of a standardized version of this estimate and use it to construct approximate confidence procedures for $\rho(t|\mathbf{z})$.

Let $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$ denote the survival times of group 1 and group 2 subjects, respectively. These are assumed to be independent random samples from populations with continuous distributions. In the case of kidney transplant studies, patients enter the hospital at different times and for some patients the time to failure X_{ij} of the graft is known, while for others the transplanted kidney has not yet failed but the time C_{ij} that the kidney was last observed as functioning is known. Such a time C_{ij} is referred to as a censoring time. We write the observable data as T_{ij}, the failure or censoring time, and δ_{ij}, a variable indicating failure or censoring for the jth subject in treatment group i. We model our study in the usual way by representing T_{ij} and δ_{ij} as

$$T_{ij} = \min(X_{ij}, C_{ij}), \quad \delta_{ij} = I[C_{ij} \geq X_{ij}]$$

where $\{C_{ij}\}$ are censoring times that are independent of the $\{X_{ij}\}$, and $I[C_{ij} \geq X_{ij}]$ equals 1 when the failure time X_{ij} is observed, and equals zero when the graft has not

yet failed and only the time C_{ij} is observed.

For the jth subject in group i, we also observe a vector $\mathbf{Z}_{ij}^T = (Z_{1ij}, \ldots, Z_{dij})$ of covariates such as age, genetic matching score, or amount of blood transfused prior to transplant. On the basis of these data, we estimate the LCHD $\rho(t|\mathbf{z})$ by

$$\hat{\rho}(t|\mathbf{z}) = (\hat{\beta}_1 - \hat{\beta}_2)\mathbf{z} + \hat{\rho}_0(t), \text{ where } \hat{\rho}_0(t) = \log \hat{\Lambda}_{01}(t) - \log \hat{\Lambda}_{02}(t).$$

Here $\hat{\beta}_1$ and $\hat{\beta}_2$ are the Cox[13,14] partial likelihood estimates of the proportional hazard model parameters $\beta_1^T = (\beta_{11}, \ldots, \beta_{1d})$ and $\beta_2^T = (\beta_{21}, \ldots, \beta_{2d})$, and $\hat{\Lambda}_{01}(t)$ and $\hat{\Lambda}_{02}(t)$ are the Breslow[15] estimates of $\Lambda_{01}(t)$ and $\Lambda_{02}(t)$.

Under regularity conditions such as those in Andersen and Gill[16] or Tsiatis[17], standardized versions of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\Lambda}_{01}(t)$ and $\hat{\Lambda}_{02}(t)$ converge in distribution to Gaussian variables (processes). This leads to approximate statistical inference procedures that can be used to address questions (a) and (b) of Section 1

(a) *Interaction.* To test whether there is interaction we test $H_0: \beta_1 = \beta_2$ vs $H_1: \beta_1 \neq \beta_2$ by rejecting H_0 for large values of the test statistic

$$T = M(\hat{\beta}_2 - \hat{\beta}_1)^T \hat{\Sigma}_{12}(\hat{\beta}_2 - \hat{\beta}_1), \text{ where } M = n_1 n_2/(n_1 + n_2).$$

Here $\hat{\Sigma}_{12}$, the estimated inverse asymptotic covariance matrix of $\sqrt{M}(\hat{\beta}_2 - \hat{\beta}_1)$, is given in the appendix. T has an asymptotic χ^2 distribution with d degrees of freedom, so we can find approximate critical values in the χ^2 table. An asymptotically equivalent approach would be to use partial likelihood ratio statistics as in Thall and Lachin[1].

(b) *Main effects.* Let t be some fixed time point of interest, e.g. t = 150 could correspond to survival for 150 days. Moreover, let $\mathbf{z}$ be a covariate vector of interest, e.g., $\mathbf{z}$ could be the covariate vector for a current or future patient. If we denote by $\hat{\sigma}(t|z)$ the estimate of the asymptotic standard error of $\hat{\rho}(t|\mathbf{z})$, as given in the appendix, then a $100(1-\alpha)\%$ pointwise confidence interval for the LCHD $\rho(t|\mathbf{z})$ is given by

$$\hat{\rho}(t|\mathbf{z}) \pm c_{\alpha/2}\hat{\sigma}(t|\mathbf{z})/M^{1/2} \tag{2.1}$$

where $c_{\alpha/2}$ is the upper $\alpha/2$ critical value in the standard normal distribution.

In some cases it is useful to have confidence intervals that are valid simultaneously for several t, say for t = 75, 150, 225, 300 and 375 days. Conservative intervals can be based on the Bonferroni's inequality. For the case of 5 time points as above, this amounts to replacing the critical constant $c_{\alpha/2}$ in the confidence interval (2.1) by $c_{\alpha/10}$. Thus if we want 90% confidence that (2.1) is valid for five given time points, we would use critical constant $c_{.1/10} = c_{.01} = 2.326$, while for 10 given time points we would use $c_{.1/20} = c_{.005} = 2.576$.

3. ANALYSIS OF THE UCSF KIDNEY TRANSPLANT STUDY.

3.1 Data description

The data considered in this paper relate to cadaveric kidney transplant performed at the University of California San Francisco. In most cases, members of the UCSF transplant team were directly involved in the processes of donor selection, donor management, recovery of the kidneys for transplantation, and their preservation prior to transplantation. Recipients for the transplants were selected on the basis of compatibility of ABO blood type and HLA and DR tissue types. However, they were rarely selected on the basis of the quality of tissue match, except in those unusual cases where the degree of match between the donors and recipients appeared to be perfect.

Immunosuppressive therapy was initiated at the time of transplantation, and was continued indefinitely thereafter for as long as the patients had a functioning transplant. Episodes of rejection were diagnosed in the conventional manner, based upon clinical findings supported by appropriate laboratory results and other diagnostic studies, which sometimes included biopsies of the transplants for microscopic examination. Rejection episodes were typically treated with higher doses of corticosteriods, and the use of other anti-rejection therapies was generally reserved for those cases in which the rejection episodes appeared refractory to corticosteriods.

Graft survival rates were calculated assuming that ''functioning'' transplants were those able to support the patients' needs without dialytic therapy, while graft ''losses'' represented those requiring surgical removal or the resumption of chronic dialytic therapy.

Two immunosuppressive treatments are considered. The first one consisted of a combination of prednisone and azathiprine administered at the time of transplant. The dosage of azathiprine was reduced in the case of toxicity. We refer to this treatment group as the *prednisone* group.

Patients in the second treatment group received cyclosporine and prednisone as the dominant mode of therapy. This group is referred to as the *cyclosporine* group. We further consider two regimens. Nonsequential therapy relates to patients in whom the use of cyclosporine and prednisone was initiated at the time of transplant. Cyclosporine is known to be potentially nephrotoxic during the first post transplant period, when kidney function is impaired,[18] so that following the earlier period of the study, sequential therapy was used preferentially. In sequential therapy the initial treatment consisted of the Minnesota Anti-Lymphoblast Globulin and prednisone and changed to cyclosporine and prednisone some days following transplant when the quality of kidney function was improved.

We consider two covariates: age (AGE)and number of units of blood transfused prior to transplant (BLOOD). Pretransplant transfusions have been shown in a number of studies to have a beneficial effect on graft survival rates. The mechanism by which this occurs is not well understood. In the case of cadaveric kidney transplants, Terasaki et al.[19] showed that transfusions have their greatest effect in the first months following transplantation by reducing accelerated graft failure occuring primarily in the first month. Transfusions appear to have almost no effect after the third month after transplant. Further, the beneficial effect of blood transfusion appears to be achieved equally with one to five transfusions, while more transfusions afford no additional benefit[20,21] A major problem with a higher number of transfusions is the risk of hepatitis followed by chronic liver disease, and also the greater likelihood of achieving sensitization which in turn reduces the possibility of readily identifying a compatible cadaver kidney by direct crossmatch testing.

3.2 Data analysis

For each patient the failure time considered is the time from transplant until failure or rejection of the graft. Only first transplant recipients are included in the analysis.

We first use the LCHD to compare hazard experiences of patients that are stratified into two groups according to the covariate BLOOD. Figure 1 gives a comparison of the patients that have had more than five (group 1) and the group that has had 5 or fewer prior units of transfused blood (group 2). In each case we set the level of the covariate z = age at 40. The graphs for age = 25 and age = 55 were very similar. In Figure 1(a) all the patients were in the cyclosporine group while in Figure 1(b) they all were in the prednisone group. The graphs of LCHD show that in this study the group with more than five units of transfused blood did better, but looking at the simultaneous confidence bands, this is not statistically significant since the lower boundary does not exceed zero. This confidence band is the Bonferroni confidence band based on confidence coefficient $1 - \alpha = .90$ and ten equally spaced time points.

Note that the LCHD is fairly close to a horizontal line in both cases in Figure 1. This supports the proportional hazard model since for this model the estimand $\rho(t|40)$ is a horizontal line. In other words, when patients receiving the same treatment are put in different groups according to different values of the covariate blood, a proportional hazard model is indicated for the different groups.

Since proportional hazards is indicated by the LCHD, we perform a Cox likelihood analysis of the data and find that the hypothesis of difference in survival experience between the two blood groups is not rejected. The two-sided P-values for the cyclosporine and prednisone patients were 0.13 and 0.66, respectively, which is consistent with our above findings based on the LCHD.

Figure 1 here

In Figure 2 we use the LCHD to compare the nonsequential (group 1) and sequential (group 2) cyclosporine regimes for patients with BLOOD covariate values 0, 5 and 10. The figure illustrates strong treatment-covariate interaction. In this study for blood level 0 (A) the sequential treatment is beneficial since $\hat{\rho}(t|0)$ is negative. For blood level 5 (B), the group $\hat{\rho}(t|5)$ is close to the horizontal line passing through 0 and both treatments have approximately the same effect on graft survival. Finally, for blood level 10 (C) the LCHD estimate is positive so that the nonsequential treatment is the better one. This interaction has important implications for treatment allocation since the value of the covariate BLOOD depends only on the patients pre-transplant history.

To test whether the treatment-covariate interaction is significant we use the Cox model for each group since an analysis of the LCHD within groups, as illustrated in Figure 1, shows proportional hazards with respect to the covariate BLOOD. Using the P2L routin in the BMPD statistical package we find the value $(-2.59)^2$ for the interaction test statistic T of Section 2. This yields a P-value of 0.0096 and we reject the hypothesis of no interaction.

Figure 2 here

We have now used the LCHD to illustrate proportional hazards within treatment groups with respect to the covariate BLOOD as well as interaction between the treatment regimens nonsequential-sequential and the covariate BLOOD. Next, in Figure 3, we illustrate the analysis of treatment effects when the Cox model does not hold by using the LCHD to compare the prednisone (group 1) and cyclosporine (group 2) regimens for patients with BLOOD covariate value 5. The graph of $\hat{\rho}(t|5)$ in Figure 3 shows that in this study there is a decaying treatment effect in the sense that cyclosporine does much better initially but this advantage diminishes with time. Thus it is not safe to assume a Cox model. On the other hand, we have seen in the discussion of Figure 1 that proportional hazards within the prednisone and cyclosporine groups are indicated and thus the semi-proportional hazard model (1.1) will be applied. Since the upper 90% simultaneous confidence boundary is below zero we can conclude at the 10% level of significance that for patients with BLOOD covariate value 5, cyclosporine is the more beneficial treatment.

The estimates and confidence bands were computed using locally written FORTRAN subroutines incorporated into S.

Figure 3 here

Appendix. Asymptotics. Estimation, confidence procedures and testing. Suppose that $n_i/(n_1+n_2) \to \eta_i$, $0 < \eta_i < 1$, $i = 1,2$, and set $M = n_1 n_2/(n_1+n_2)$. Further, let us assume that given $\mathbf{z}$, the conditions of Andersen and Gill (1982) are satisfied by both groups in model (1.1) in a time interval $[0, t_1]$. Finally, for $i = 1,2$, let $Y_{ij}(t) = I[T_{ij} \geq t]$ and set

$$S_i^{(0)}(t,\beta_i) = \sum_{j=1}^{n_i} Y_{ij}(t)\exp(\beta_i^T Z_{ij}), \quad S_i^{(1)}(t,\beta_i) = \sum_{j=1}^{n_i} Y_{ij}(t)\,Z_{ij}\exp(\beta_i^T Z_{ij})$$

$$S_i^{(2)}(t,\beta_i) = \sum_{j=1}^{n_i} Y_{ij}(t)\,Z_{ij}^{\otimes 2}\exp(\beta_i^T Z_{ij}), \quad E_i(t,\beta_i) = S_i^{(1)}(t,\beta_i)/S_i^{(0)}(t,\beta_i),$$

$$V_i(t,\beta_i) = [S_i^{(2)}(t,\beta_i)/S^{(0)}(t,\beta_i)] - E_i^{\otimes 2}(t,\beta_i)$$

where for any vector $(a_1, \ldots, a_d) = \mathbf{a}$, $\mathbf{a}^{\otimes 2}$ is a $d \times d$ matrix with (k, l) entry equal to $a_k a_l$. For the ith group, $S_i^{(0)}(t,\beta_i)$ represents a weighted number of individuals at risk at time t. Further $S_i^{(1)}(t,\beta_i)$ is the vector of partial derivatives of $S_i^{(0)}(t,\beta_i)$ with respect to $\beta_i = (\beta_1, \ldots, \beta_d)$ whereas $S_i^{(2)}(t,\beta_i)$ is the matrix of second partial derivatives. We let $s_i^{(j)}(t,\beta_{0i})$, $j = 0,1,2$, $e_i(t,\beta_{0i})$ and $v_i(t,\beta_{0i})$ be the "in probability" limits of the above variables in a neighbourhood of the true parameter vectors β_{0i}. We follow here Andersen and Gill (1982, Condition B), and assume that these limits exist. Let

$$\Sigma_i = \Sigma_i(\beta_{0i}) = \int_0^{t_1} v_i(u,\beta_{0i})\,s_i^{(0)}(u,\beta_{0i})\,\lambda_{0i}(u)\,du$$

$$\xi_i(t,\beta_{0i}) = \int_0^{t} e_i(u,\beta_{0i})\,\lambda_{0i}(u)\,du$$

$$\mu_i(t,\beta_{0i}) = \int_0^{t} s_i^{(0)}(u,\beta_{0i})^{-1}\,\lambda_{0i}(u)\,du$$

where $\lambda_{0i}(u)$ is the derivative of Λ_{0i} In the following distributions are conditional given $\mathbf{Z} = \mathbf{z}$. We assume model (1.1).

Proposition A.1. Let t_0 and t_1 be two fixed time points with $0 \leq t_0 < t_1 < \infty$. Under the conditions in Andersen and Gill (1982) the process $M^{1/2}[\hat{\rho}(t|\mathbf{z}) - \rho(t|\mathbf{z})]$ converges for almost all $\mathbf{z}$ weakly in $D[t_0, t_1]$ to a mean zero Gaussian process $W(t|\mathbf{z})$ with covariance function given by

$$\operatorname{cov}(W(s|\mathbf{z}),\ W(t|\mathbf{z})) =$$

$$= [\eta_2 C_1(s,t|\mathbf{z})/\Lambda_{01}(s)\Lambda_{01}(t)] + [\eta_1 C_2(s,t|\mathbf{z})/\Lambda_{02}(s)\Lambda_{02}(t)]$$

where

$$C_i(s,t|z) = \mu_i(s \wedge t, \beta_{0i}) + [\xi_i(s,\beta_{0i}) - \Lambda_{0i}(s)z]^T \Sigma_i^{-1} [\xi_i(t,\beta_{0i}) - \Lambda_{0i}(t)z]$$

is the covariance function of

$$n_i^{1/2}[\log\hat{\Lambda}_i(t|z) - \log\Lambda_i(t|z)] = [\log\hat{\Lambda}_{0i}(t) - \log\Lambda_{0i}(t)] + (\hat{\beta}_i - \beta_{0i})^T z,\ i = 1,2$$

Proof. We write

$$M^{1/2}[\log\hat{\Lambda}_i(t|z) - \log\Lambda_i(t|z)] = M^{1/2}[\log\hat{\Lambda}_{0i}(t) - \log\Lambda_{0i}(t)] + M^{1/2}(\hat{\beta}_i - \beta_{0i})^T z.$$

By the δ-method for processes, $M^{1/2}[\log\hat{\Lambda}_{0i}(t) - \log\Lambda_{0i}(t)]$ can, for the purposes of obtaining limiting distributions, be replaced by

$$M^{1/2}[\hat{\Lambda}_{0i}(t) - \Lambda_{0i}(t)]/\Lambda_{0i}(t),\ i = 1,2.$$

Since these two processes as well as $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent, it follows from Andersen and Gill[16] and some covariance calculations that $M^{1/2}[\hat{\rho}(t|z) - \rho(t|z)]$ $= M^{1/2}\sum_{i=1}^{2}[\log\hat{\Lambda}_i(t|z) - \log\Lambda_i(t|z)]$ converges weakly to the indicated Gaussian stochastic process.

Let $\sigma^2(t|z)$ be the variance of $W(t|z)$ at time t. We estimate it by

$$\hat{\sigma}^2(t|z) = \frac{n_2}{n_1+n_2}\hat{C}_1(t|z)/\hat{\Lambda}_{01}^2(t) + \frac{n_1}{n_1+n_2}\hat{C}_2(t|z)/\hat{\Lambda}_{02}^2(t)$$

where $\hat{C}_i(t|z)$ is the sample counterpart of $C_i(t|z)$. More precisely, let $N_i(u) = \Sigma_{j=1}^{n_i} I[T_{ij} \le u, \delta_{ij} = 1]$, then $\hat{C}_i(t|z)$ is $C_i(t|z)$ with ξ_i, μ_i and Σ_i estimated by

$$\hat{\Sigma}_i = \int_0^{t_1} V_i(u,\hat{\beta}_i)\,dN_i(u),\quad \hat{\xi}_i(t) = \int_0^t E_i(u,\hat{\beta}_i)\,\hat{\Lambda}_{0i}(du),\quad \hat{\mu}_i(t) = \int_0^t S_i^{(0)}(u,\hat{\beta}_i)^{-1}\,\hat{\Lambda}_{0i}(du).$$

Note that Σ_i^{-1} is the asymptotic covariance matrix of $n_i^{1/2}(\hat{\beta}_i - \beta_{0i})$ and that $\sqrt{M}[\hat{\beta}_2 - \hat{\beta}_1 - (\beta_{02} - \beta_{01})]$ is asymptotically normal with mean 0 and covariance matrix

$$\Sigma_{12}^{-1} = \eta_2\Sigma_1^{-1} + \eta_1\Sigma_2^{-1}.$$

Our estimate of Σ_{12}^{-1} is $\hat{\Sigma}_{12}^{-1} = (n_2\hat{\Sigma}_1^{-1} + n_1\hat{\Sigma}_2^{-1})/(n_1+n_2)$.

Acknowledgments We thank Ørnulf Borgan, Niels Keiding and Stewart Kroll for helpful comments. We also thank the referees for their suggestions which led to an improvement of the original paper.

References

1. Thall, P.F. and Lachin, J.M. 'Assessment of stratum-covariate interactions in Cox's proportional hazard regression model.' *Statistics in Medicine*, **5**, 73-83 (1986).

2. Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. 'Statistical models based on counting processes.' Springer Verlag, New York (1992).

3. Kay, R. 'Proportional hazards regression models and the analysis of censored survival data.' *Appl. Statist.*, **26**, 227-37 (1977).

4. Kalbfleisch, J.D. and Prentice, R.S. 'The Statistical Analysis of Failure Time Data.' Wiley, New York (1980).

5. Therneau, T.M., Grambsch, P.M. and Fleming, T.R. 'Martingale-based residuals for survival models.' *Biometrika*, **77**, 147-60 (1990).

6. Kronborg, D. and Aaby, P. 'Piecewise comparison of survival functions in stratified proportional hazards models.' *Biometrics*, **46**, 375-380 (1990).

7. Nielsen, G.G., Andersen, P.K., Gill, R.D. and Sørensen, T.I.A. 'A counting process approach maximum likelihood estimation infrailty models.' Technical Report 91/2. Statistical Research, University of Copenhagen, Denmark (1991).

8. Doll, R. 'Practical steps towards the prevention of bronchial carcinoma.' *Scot. Med. Journal*, **15**, 433-47 (1970).

9. Freedman, D. and Navidi, W. 'Ex-smokers and the multistage model for lung cancer.' *Epidemiology*, **1**, 21-29 (1990).

10. Fleming, T.R. and Harrington, D.P. 'Counting processes and survival analysis.' Wiley, New York (1990).

11. Dabrowska, D.M., Doksum, K.A. and Song, J.K. 'Graphical comparison of cumulative hazards for two populations.' *Biometrika*, **76**, 763-73 (1989).

12. Bie, O., Borgan, Ø. and Liestøl, K. 'Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties.' *Scand. J. Statist.*, **14**, 221-33 (1987).

13. Cox, D.R. 'Regression models and life tables.' *J. Roy. Statist. Soc. Ser. B*, **34**, 187-202 (1972).

14. Cox, D.R. 'Partial likelihood.' *Biometrika*, **62**, 269-276 (1975).

15. Breslow, N.E. 'Covariance analysis of censored survival data.' *Biometrics*, **30**, 89-99 (1974).

16. Andersen, P.K. and Gill, R.D. 'Cox's regression model for counting processes: a large sample study.' *Ann. Statist.*, **10**, 1100-1120 (1982).

17. Tsiatis, A.A. 'A large sample study of Cox's regression model.' *Ann. Statist.*, **9**, 93-108 (1981).

18. Feduska, N.J., Melzer, J.S., Amend, W.J., Tomlanovich, S., Vincenti, F., Garovoy, M., Husing, R., Salvatierra, O. 'Cyclosporine provides better success rates for both higher and lower-risk cadaveric kidney transplant recipients.' *Transpl Proc*, **20**, 102 (1988).

19. Terasaki, P.I., Perdue, S., Ayoub, G., et al. 'Reduction of accelerated failure rates by transfusion.' *Transplant Proc.*, **14**, 251 (1982).

20. Feduska, N.J., Vincenti, F., Amend, W.J., Duca, R., Cochrum, K., Salvatierra, O. 'Do blood transfusions enhance the possibility of a compatible transplant?' *Transplantation*, **27**, 365-38 (1979).

21. Salvatierra, O., Vincenti, F., Amend, W.J.C., Garovoy, M.R., Potter, D. and Feduska, N.J. 'The role of blood transfusions in renal transplantation.' *Urologic Clinics of North America*, **10**, 243-252 (1983).

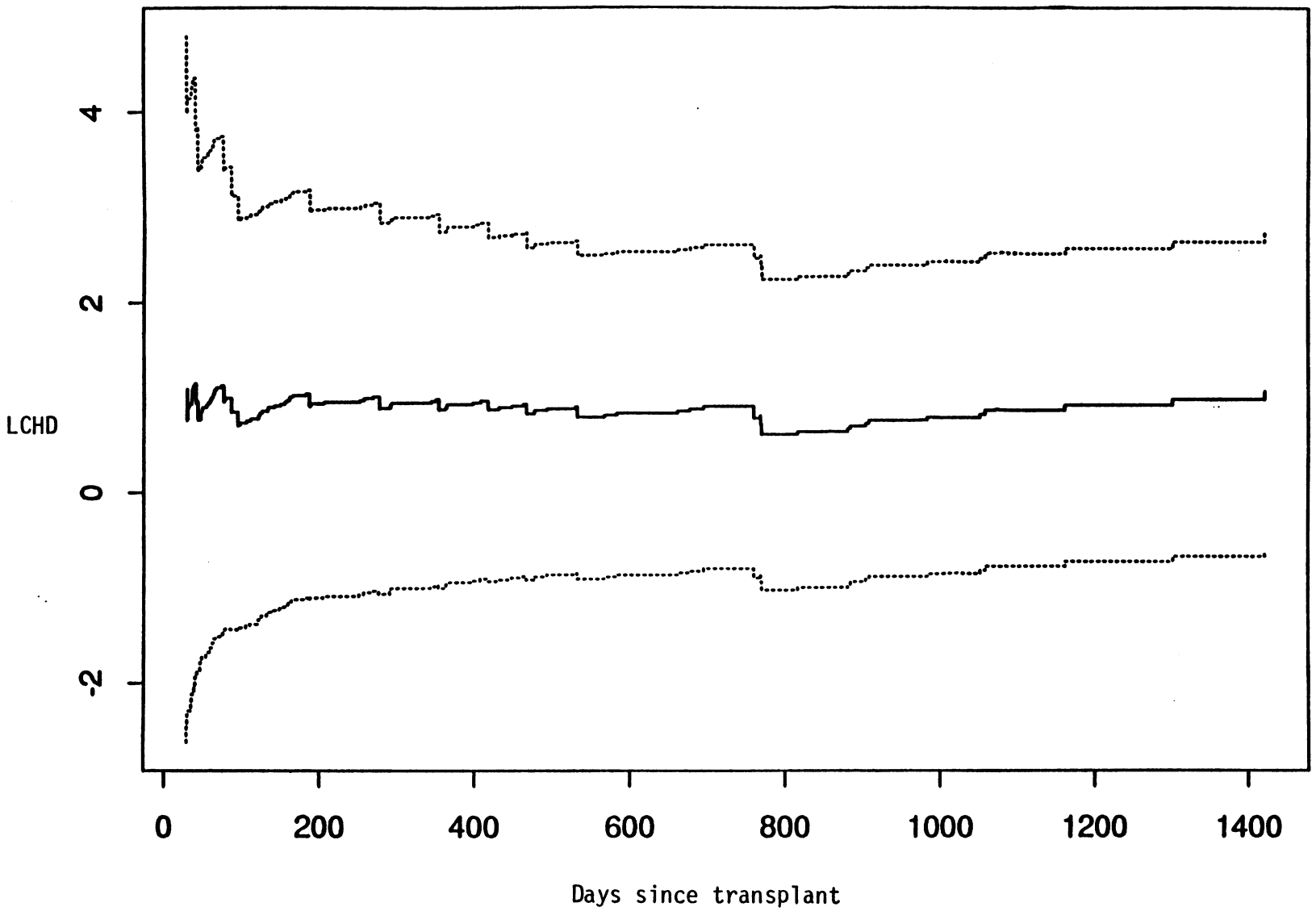
22. Doksum, K.A. and Yandell, B. Tests for exponentially. *Handbook of Statistics*, Vol. 4. Krishnaiah, P.R. and P.K. Sen, editors. Elsevier Science Publishers (1984).

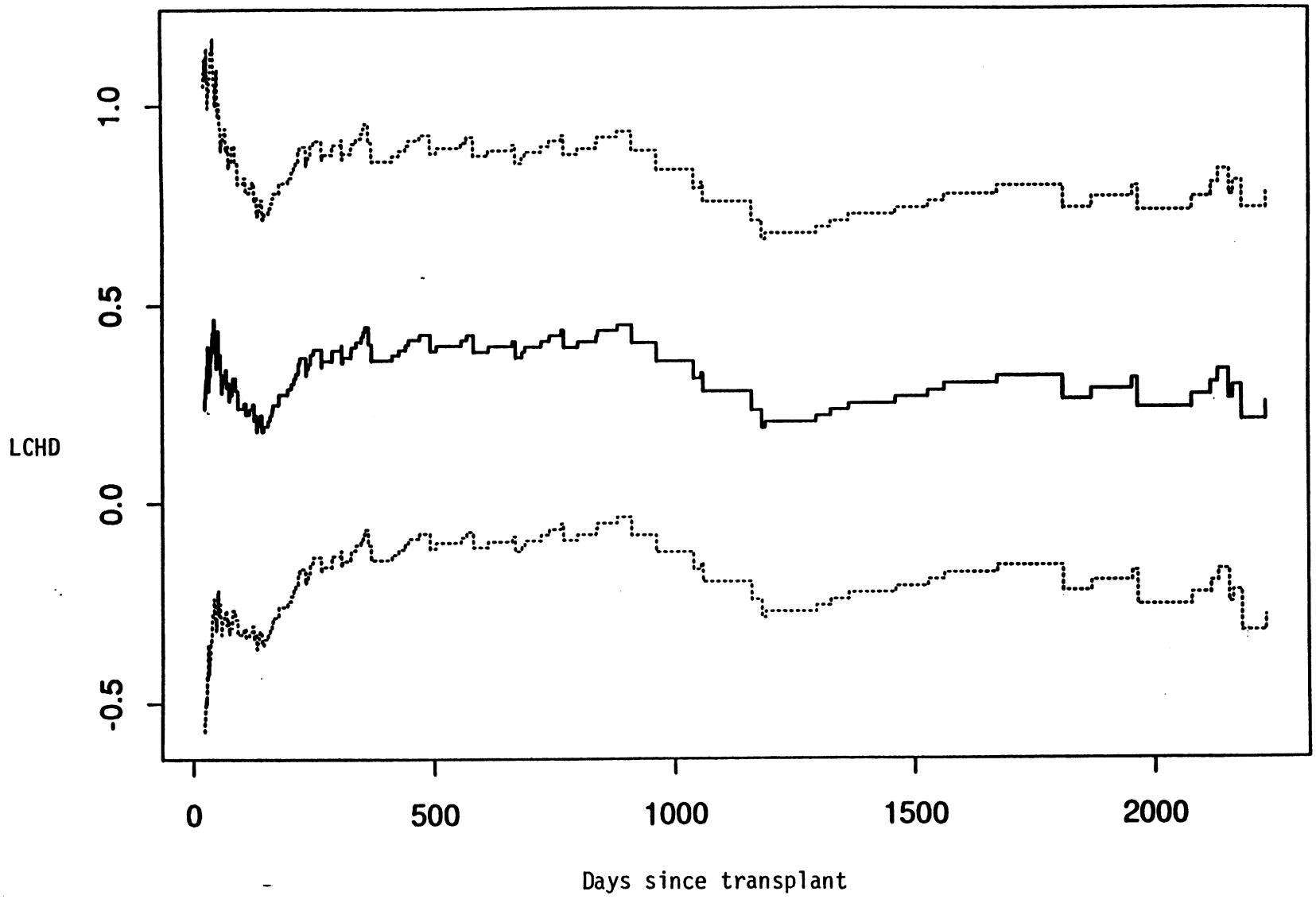
FIGURE CAPTIONS

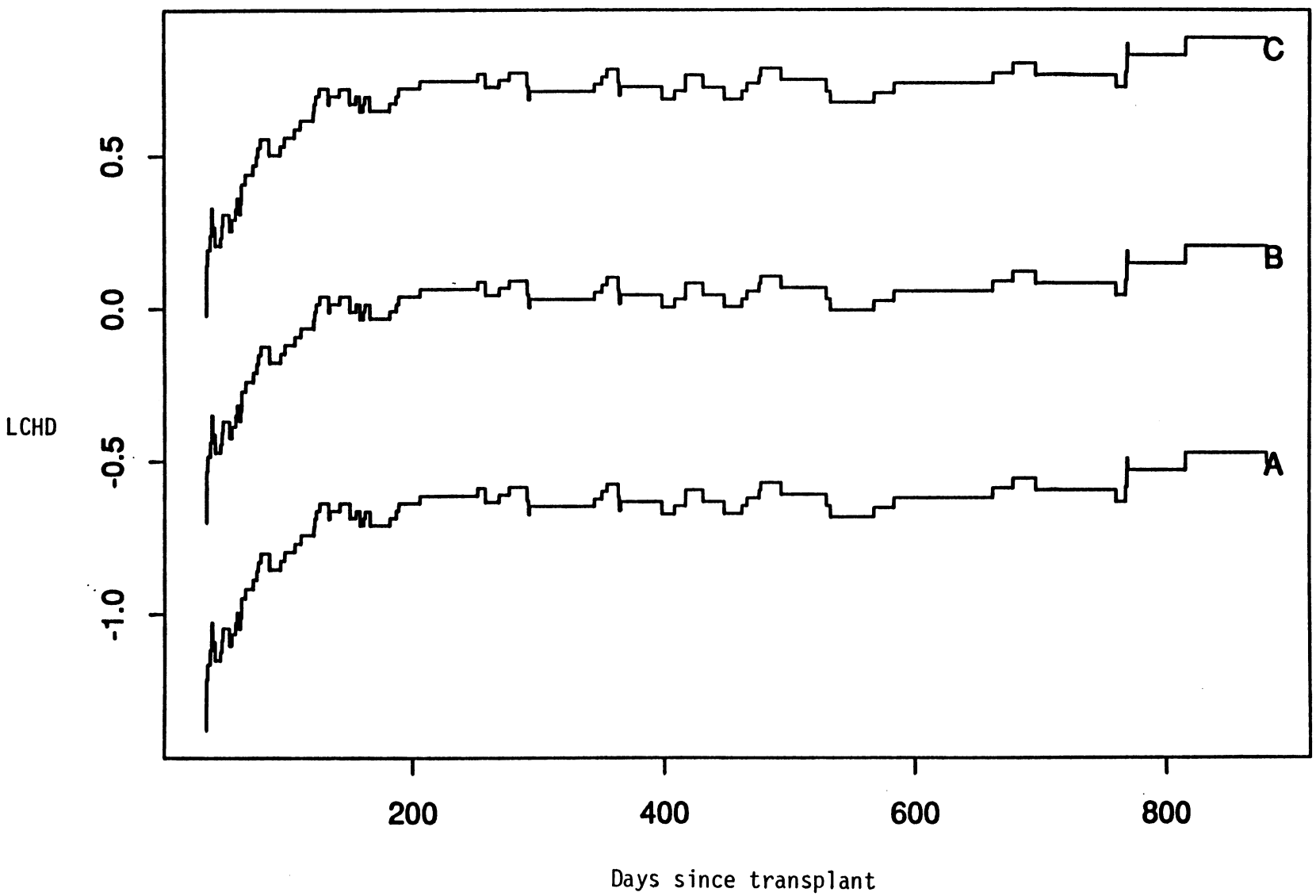
Figure 1. The log cumulative hazard difference (solid curve) comparing groups receiving more than 5 (group 1) and 5 or fewer (group 2) units of blood prior to transplant. The covariate age is set at $z = 40$. The dotted curves give a 90% simultaneous confidence band. Figure 1(a) is for the cyclosporine group ($n_1 = 142$, $n_2 = 481$) and Figure 1(b) is for the prednisone group ($n_1 = 98$, $n_2 = 256$).

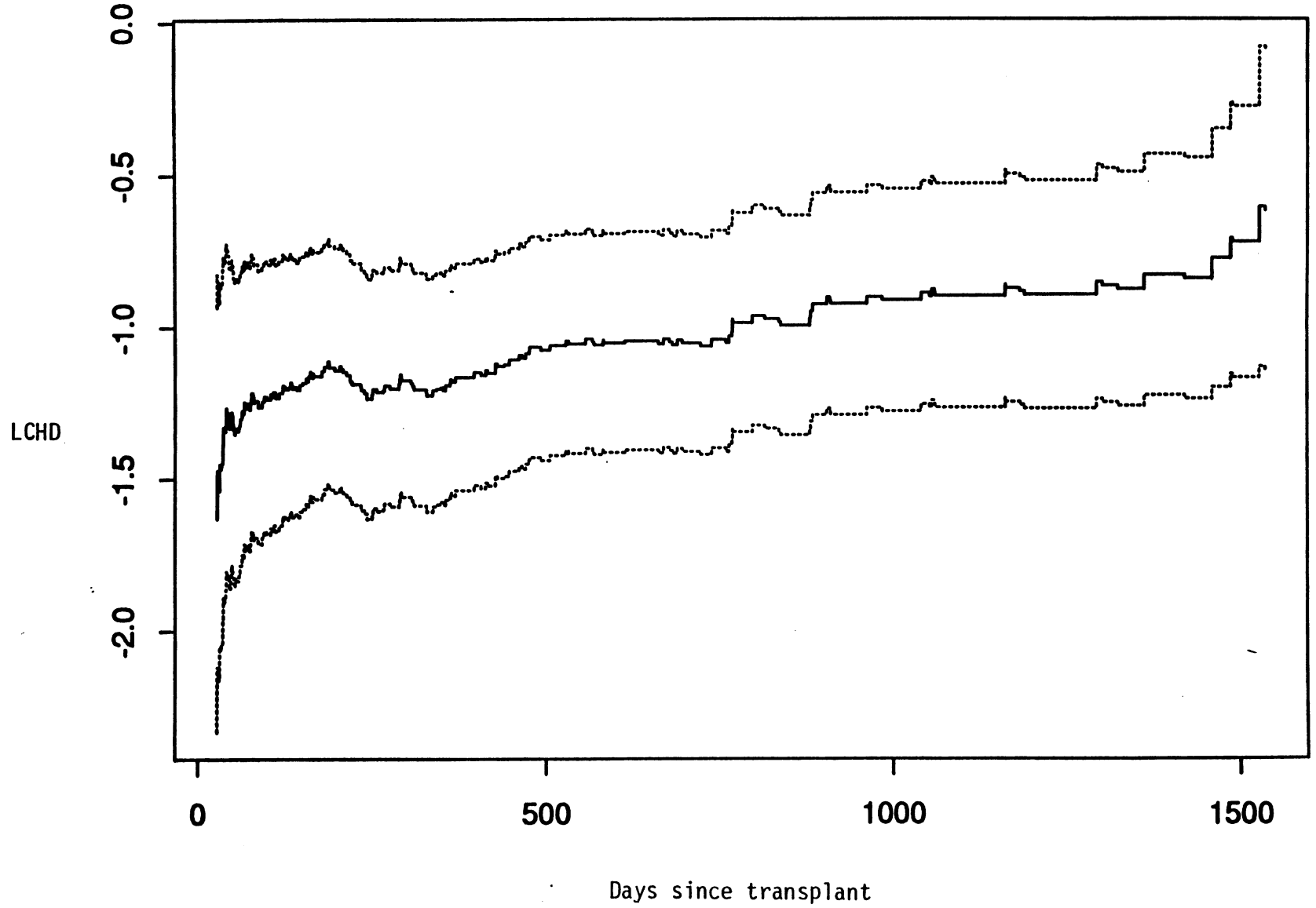
Figure 2. The log cumulative hazard difference comparing the nonsequential treatment (group 1) to the sequential treatment (group 2). All the transplant recipients were in the cyclosporine group. The covariate "no of units of transfused blood" is set at $z = 0,5$ and 10 and the corresponding relative risk curves are labelled A, B and C, respectively. The sample sizes are $n_1 = 157$ and $n_2 = 466$.

Figure 3. The log cumulative hazard difference (solid curve) comparing the treatments prednisone (group 1) and cyclosporine (group 2). The covariate "number of units of transfused blood" is set at $z = 5$. The dotted curves give a 90% simultaneous confidence band. The sample sizes are $n_1 = 354$ and $n_2 = 623$.


LCHD
4
2
0
-2
0
200
400
600
800
1000
1200
1400
Days since transplant


LCHD
1.0
0.5
0.0
-0.5
0
500
1000
1500
2000
Days since transplant


LCHD
0.5
0.0
-0.5
-1.0
C
B
A
200
400
600
800
Days since transplant


LCHD
0.0
-0.5
-1.0
-1.5
-2.0
0
500
1000
1500
Days since transplant