Normalised Local Hazard Plots

Nils Lid Hjort[a,b] and Thomas Lumley[b]

University of Oslo[a] and University of Oxford[b]

ABSTRACT. The purpose of this paper is to develop and illustrate certain classes of graphical plots that can be used for model verification in quite general survival data and life history data models. By suitably comparing nonparametric and parametric estimates of hazard rate functions over time a hazard comparison function can be constructed which under parametric model assumptions is approximately a zero-mean normal process. The test curves we propose are locally normalised versions of such hazard comparison functions. Under model conditions the test function is approximately a standard normal for each time point. This makes the normalised local hazard curves easy to interpret. We give explicit constructions for the most commonly used models of survival analysis, including the exponential, the Weibull, the Gompertz, the gamma, and for parametric Cox regression. Algorithms carrying this out have been developed in S-Plus. Various theoretical and practical issues are discussed, including detection power and extensions to time-discrete models. Illustrations are given on simulated and real data.

KEY WORDS: *counting processes, goodness of fit, hazard function models, normalised local hazard plots, parametric Cox model,* S-Plus*, time-discrete hazard models*

1. Introduction and summary. This paper discusses and illustrates certain graphical plots that can be used for model verification purposes in survival data situations and in fact also in much more general counting process models for life history data. Such models involve certain hazard rate functions. By suitably comparing nonparametric and parametric estimates of hazard rate functions over time we are able to produce a hazard comparison function which under the conditions of the parametric model is approximately a zero-mean normal process. And by normalising this local hazard function with a local estimate of the standard deviation we end up with a test function which is approximately a standard normal for each time point, provided the parametric model studied is valid.

We call these test curves *normalised local hazard plots.* Such a test curve is a concise summary of the discrepancy between the parametric and nonparametric model, and is easy to judge by eye, since the precision is the same throughout the time axis, and since everybody knows the standard normal. If the model is right then the test curve should stay inside the ± 1.96 horizontal band most of the time, and incorrect aspects of the model will lead to curves that wander outside this band.

There are also formal goodness of fit testing procedures, see Hjort (1990) and Andersen, Borgan, Gill & Keiding (1993, chapter VI), the latter henceforth referred to as ABGK. Such tests are valuable since they can guarantee e.g. a maximum 5% chance of incorrectly rejecting a model, of course, but the emphasis here is on graphical checking tools for various models. The plots will be useful when the statistician explores different modelling opportunities and can be more informative than the simple yes or no answer that a formal test provides, in that one is shown in what way or ways a given model does not hold. The formal tests can in this light be regarded as giving supplementary confirmatory information.

To construct the test curves properly one essentially needs (a) proofs of limiting normality of the processes, (b) expressions for and consistent estimates for the variances of the limit processes. Sufficient theory for this has in fact already been developed in Hjort (1985, 1990). The model test plot idea was also briefly mentioned in these papers, as an attractive consequence of the general results. Arulchelvam (1992) implemented one version of these plots for a couple of models, and she illustrated their use on simulated and real data. The intention of the present paper is

to systematically work out variance estimation and other necessary details for the most popular models, and to demonstrate the usefulness of the plots on real and simulated data.

For concreteness we choose to develop and discuss the plots mainly in the context of 'traditional survival analysis', involving possibly right-censored observations of life-times, with or without covariate information. The basic methods and results hold with suitable modifications for general counting process models for life history data, as briefly discussed in section 6. The traditional framework for homogeneous data is as follows: $T_1^0, \ldots, T_n^0$ are i.i.d. life-times from a distribution with hazard rate $h(s)$ and cumulative hazard rate $H(t) = \int_0^t h(s)\,\mathrm{d}s$. The data may be censored from the right, so what one observes is $t_j = \min(t_j^0, c_j)$ and $\delta_j = I\{t_j^0 \le c_j\}$, where c_j is the possibly interfering censoring time. The censoring mechanism is assumed noninformative (see ABGK for discussion of this). A parametric model is considered, say of the form $h(s) = h(s,\theta)$ for some unknown parameter $\theta = (\theta_1, \ldots, \theta_p)'$.

It is convenient to define and study both parametric and nonparametric estimation in terms of

$$N(t) = \sum_{j=1}^n I\{t_j \le t, \delta_j = 1\} \quad \text{and} \quad Y(t) = \sum_{j=1}^n I\{t_j \ge t\}. \tag{1.1}$$

Here $N(t)$ is the counting process, counting the number of observed failures, while the left-continuous $Y(t)$ is the at risk process, counting the number among the n original items that are still at risk just prior to time point t. Since $N(.)$ is flat between observed life-times $\int_0^t g(s)\,\mathrm{d}N(s)$ means simply $\sum_{j:t_j \le t} g(t_j)\delta_j$. Nonparametric estimation of H is carried out using the Nelson–Aalen estimator

$$\widehat{H}(t) = \int_0^t \mathrm{d}N(s)/Y(s) = \sum_{j:t_j \le t} \delta_j / Y(t_j). \tag{1.2}$$

As for parametric estimation, we work with the maximum likelihood (ML) estimator $\widehat{\theta}$. The log-likelihood can be written $\int_0^\tau \{\log h(s,\theta)\,\mathrm{d}N(s) - Y(s)h(s,\theta)\,\mathrm{d}s\}$, where $[0,\tau]$ is the finite or infinite time interval over which the processes are observed, see for example ABGK (chapter VI).

Hjort (1990) studied a large class of goodness of fit processes of the type

$$D_n(t) = \sqrt{n}\int_0^t K_n(s)\{\mathrm{d}\widehat{H}(s) - h(s,\widehat{\theta})\,\mathrm{d}s\} = \sqrt{n}\int_0^t \frac{K_n(s)}{Y(s)}\{\mathrm{d}N(s) - Y(s)h(s,\widehat{\theta})\,\mathrm{d}s\}. \tag{1.3}$$

This local hazard comparison function should take values around zero if the model holds. The weight function $K_n(s)$ can be chosen to suit different aims, as exemplified later, and is here meant to be scaled in a stable way, so that it has a well-defined limit in probability function $k(s)$ as n grows. Special cases of interest include the following, to be referred to later as Type A, Type B, Type C:

$$\begin{array}{lll} D_n(t) &= \sqrt{n}\{\widehat{H}(t) - H(t,\widehat{\theta})\}, & \text{for } K_n(s) = 1, \\ D_n(t) &= n^{-1/2}\{N(t) - \int_0^t Y(s)h(s,\widehat{\theta})\,\mathrm{d}s\}, & \text{for } K_n(s) = Y(s)/n, \\ D_n(t) &= n^{-1/2}\int_0^t G_n(s)\{\mathrm{d}N(s) - Y(s)h(s,\widehat{\theta})\,\mathrm{d}s\}, & \text{for } K_n(s) = \{Y(s)/n\}G_n(s). \end{array} \tag{1.4}$$

Type A uses $K_n(s) = 1$, and then $D_n(.)$ directly compares nonparametric versus parametric estimates of the cumulative hazard function. Type B employs $K_n(s) = Y(s)/n$, the proportion at risk at time s, in which case $D_n(.)$ compares the observed number of failures with the predicted number of failures under the parametric model. There are other ways of predicting the number of failures in the course of a time interval from the model, but this is a natural way based directly on the hazard rate interpretation; $\mathrm{d}N(s)$, the number of failures in $[s, s+\mathrm{d}s]$ among the $Y(s)$ that

are observed to be at risk just prior to time s, is simply a binomial with probability parameter $h(s,\theta)\,ds$. This leads to $\int_a^b Y(s)h(s,\widehat{\theta})\,ds$ as the 'expected' version of $N[a,b]$. Finally Type C, with $K_n(s) = \{Y(s)/n\}\,G_n(s)$, is as general as (1.3), of course, but some interesting procedures that are tailor-made to have optimal detection power against certain departures from the model are of this form, sometimes with a deterministic weight function $G_n(s) = g(s)$, as exemplified in 3.2.

Under a mild set of regularity conditions $D_n(.)$ tends in distribution to a certain zero-mean Gaußian process $D(.)$, see section 2. The basic normalised local hazard plot idea, as explained in the introductory paragraphs, is to draw the curve

$$\text{NLH}(t) = D_n(t)/\widehat{\kappa}(t) \quad \text{versus time } t \in (0,\tau), \tag{1.5}$$

where the denominator is an estimate of the local standard deviation. This curve has the property of being approximately a standard normal *for each* t, provided the model is correct.

The rest of the paper is organised as follows: Section 2 discusses some options for $\kappa(t)$ estimation and gives detailed general descriptions for three types of NLH-plots. This is next applied in section 3 to a selection of the most popular parametric models of survival analysis: The exponential model with a constant hazard rate, the Weibull, the gamma, et cetera. In section 4 similar goodness of fit plots are developed and illustrated for the case of a parametric Cox regression model. Behaviour of the plots outside model conditions, and the plots' detection power against various kinds of alternatives, is discussed in section 5. Section 6 shows how the previous methods and results extend to more general models for life history data, like time-inhomogeneous Markov chains. Although our emphasis is on time-continuous models and methods we take time out in section 7 to present the essential formulae for time-discrete hazard rate models. A number of supplementary remarks, including comparisons with other graphical goodness of fit methods that have been proposed in the literature, are placed in section 8.

The paper concludes in visual mode, providing a number of illustrations on real and simulated data in section 9. User-friendly well-commented algorithms producing the different plots, as well as the necessary parameter estimates and standard deviations, have been implemented in `S-Plus`. These are available from the authors upon polite request and are briefly described in the Appendix.

2. The three basic types of NLH-plots. We need the limit distribution of $D_n(t)$ of (1.3), and we need consistent estimates of the limit variance function.

2.1. LIMIT DISTRIBUTION OF THE GOODNESS OF FIT PROCESS. Some further notation and results are needed at this stage. A basic ingredient in the large-sample analysis of the behaviour of both parametric and nonparametric estimators is the limiting process $V(t)$ of the martingale $n^{-1/2}\{N(t) - \int_0^t Y(s)h(s)\,ds\}$. This is a zero-mean Gaußian process with independent increments of size $\text{Var}\{dV(s)\} = y(s)h(s)\,ds$, where $y(s)$ is the limit in probability function of $Y(s)/n$, i.e. the limiting proportion of items at risk at time s. One can show that $\sqrt{n}\{\widehat{H}(t) - H(t)\}$ tends in distribution to $\int_0^t y(s)^{-1}\,dV(s)$, see ABGK (chapter IV). Let next $\psi(s,\theta) = \frac{\partial}{\partial\theta}\log h(s,\theta)$, a $(p\times 1)$-vector. The $p\times p$-matrix

$$\Sigma = \int_0^\tau \psi(s,\theta)\psi(s,\theta)'y(s)\,h(s,\theta)\,ds \tag{2.1}$$

will enter several of our calculations. The ML estimator solves $\int_0^\tau \psi(s,\theta)\{dN(s) - Y(s)h(s,\theta)\,ds\} = 0$. It is consistent and satisfies

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d \Sigma^{-1}\int_0^\tau \psi(s,\theta)\,dV(s) \sim \mathcal{N}_p\{0,\Sigma^{-1}\}. \tag{2.2}$$

The limit of $D_n(t)$ can now be described: Under natural and mild regularity conditions, which include convergence in probability of the weight function $K_n(s)$ to an appropriate $k(s)$, it converges in distribution to

$$D(t) = \int_0^t \{k(s)/y(s)\}\,\mathrm{d}V(s) - \Big(\int_0^t k(s)h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big)'\Sigma^{-1}\int_0^\tau \psi(s,\theta)\,\mathrm{d}V(s).$$

This is proved in Hjort (1990, section 2). The limit is a zero-mean Gaußian process, and some calculations show that the variance is

$$\kappa(t)^2 = \int_0^t \frac{k(s)^2}{y(s)}h(s,\theta)\,\mathrm{d}s - \Big(\int_0^t k(s)h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big)'\Sigma^{-1}\Big(\int_0^t k(s)h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big). \qquad (2.3)$$

Of course the covariance needs to be given too in order to fully describe the $D(.)$ process, but our primary interest in this paper lies with the NLH-plots (1.5). Therefore the variance $\kappa(t)^2$ is the quantity we need to estimate, at least for $k(.)$ functions corresponding to Type A, Type B, Type C encountered in (1.4).

2.2. ESTIMATING THE VARIANCE. There are several choices, each with its own merits. A parametric plug-in option is to use $h(s,\widehat\theta)$ and $\psi(s,\widehat\theta)$, together with $\widehat y(s) = Y(s)/n$ for $y(s)$ and $K_n(s)$ for $k(s)$. A nonparametric plug-in option is to substitute $\mathrm{d}\widehat H(s) = \mathrm{d}N(s)/Y(s)$ for $h(s,\theta)\,\mathrm{d}s$ everywhere in (2.3), and again using $Y(s)/n$ and $K_n(s)$. It is shown in Hjort (1990, section 3) that each of these choices, or indeed combination of these choices, lead to consistent estimators of $\kappa(t)^2$.

The natural parametric estimator of Σ is

$$\widehat\Sigma_{\mathrm{pm}} = \int_0^\tau \psi(s,\widehat\theta)\psi(s,\widehat\theta)'\{Y(s)/n\}h(s,\widehat\theta)\,\mathrm{d}s = n^{-1}\sum_{j=1}^n \int_0^{t_j} \psi(s,\widehat\theta)\psi(s,\widehat\theta)'\,h(s,\widehat\theta)\,\mathrm{d}s, \qquad (2.4)$$

and is easy to calculate in cases where the integrals can be done explicitly. The natural nonparametric alternative is

$$\widehat\Sigma_{\mathrm{np}} = \int_0^\tau \psi(s,\widehat\theta)\psi(s,\widehat\theta)'\frac{Y(s)}{n}\frac{\mathrm{d}N(s)}{Y(s)} = n^{-1}\sum_{j:\delta_j=1}\psi(t_j,\widehat\theta)\psi(t_j,\widehat\theta_j)' = n^{-1}\sum_{j=1}^n \psi(t_j,\widehat\theta)\psi(t_j,\widehat\theta)'\delta_j,$$

which is easier to calculate where the integrals do not allow explicit expressions. These are both consistent estimates, under model conditions.

2.3. NLH-PLOTS, TYPE A. Type A of (1.4) uses $K_n(s) = 1$, and has

$$\kappa_A(t)^2 = \int_0^t \frac{h(s,\theta)}{y(s)}\,\mathrm{d}s - \Big(\int_0^t h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big)'\Sigma^{-1}\Big(\int_0^t h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big).$$

The parametric plug-in option is

$$\widehat\kappa_A(t)^2 = \int_0^t \frac{n}{Y(s)}h(s,\widehat\theta)\,\mathrm{d}s - H^*(t,\widehat\theta)'\widehat\Sigma_{\mathrm{pm}}^{-1}H^*(t,\widehat\theta), \qquad (2.5)$$

where we write $H^*(t,\theta)$ for $\frac{\partial}{\partial\theta}H(t,\theta) = \int_0^t h(s,\theta)\psi(s,\theta)\,\mathrm{d}s$, for notational convenience. The first integral is easiest to calculate by adding over the ordered $(t_{j-1},t_j]$ intervals to the left of time point t, since $Y(s)$ is constant over each of these. One then gets a sum of type $\sum_{j:t_j\le t}\{n/Y(t_{j-1})\}\{H(t_j,\widehat\theta)$

$- H(t_{j-1},\widehat\theta)\}$ plus a similar term for the remaining $(t_l,t]$ interval, where t_l is the last t_j before t. The nonparametric option is

$$\widehat\kappa_A(t)^2 = \int_0^t \frac{n\,dN(s)}{Y(s)^2} - \Big(\int_0^t \psi(s,\widehat\theta)\frac{dN(s)}{Y(s)}\Big)'\widehat\Sigma_{\rm np}^{-1}\Big(\int_0^t \psi(s,\widehat\theta)\frac{dN(s)}{Y(s)}\Big).$$

Note that the first term is n times the usual variance estimator for the Nelson–Aalen estimator, see ABGK (chapter IV). The second integral is just the finite sum $\sum_{t_j\le t}\psi(t_j,\widehat\theta)\delta_j/Y(t_j)$.

2.4. NLH-PLOTS, TYPE B. Next consider Type B of (1.4), which uses $K_n(s)=Y(s)/n$, and has

$$\kappa_B(t)^2 = \int_0^t y(s)h(s,\theta)\,ds - \Big(\int_0^t y(s)h(s,\theta)\psi(s,\theta)\,ds\Big)'\Sigma^{-1}\Big(\int_0^t y(s)h(s,\theta)\psi(s,\theta)\,ds\Big).$$

The parametric choice becomes

$$\widehat\kappa_B(t)^2 = n^{-1}\sum_{j=1}^n H(t_j\wedge t,\widehat\theta) - \Big(n^{-1}\sum_{j=1}^n H^*(t_j\wedge t,\widehat\theta)\Big)'\widehat\Sigma_{\rm pm}^{-1}\Big(n^{-1}\sum_{j=1}^n H^*(t_j\wedge t,\widehat\theta)\Big), \qquad (2.6)$$

while the nonparametric version is

$$\widehat\kappa_B(t)^2 = N(t)/n - \Big(n^{-1}\sum_{t_j\le t}\psi(t_j,\widehat\theta)\delta_j\Big)'\widehat\Sigma_{\rm np}^{-1}\Big(n^{-1}\sum_{t_j\le t}\psi(t_j,\widehat\theta)\delta_j\Big).$$

2.5. NLH-PLOTS, TYPE C. Finally consider the third choice in (1.4), which employs a weight function of the $K_n(s)=\{Y(s)/n\}G_n(s)$. Suppose for example that an omnibus goodness of fit test for the $h(s)=h(s,\theta)$ model is sought, that at the same time has good detection power for neighbouring alternatives of the form $h(s,\theta,\gamma)$, where $h(s,\theta)$ is the special case $h(s,\theta,\gamma_0)$. Then an optimal choice for weight function in (1.3) can be shown to be of the form

$$K_n(s) = n^{-1}Y(s)\Big\{\phi(s,\widehat\theta) - \psi(s,\widehat\theta)'\widehat\Sigma_{\rm pm}^{-1}\int_0^\tau n^{-1}Y(u)\phi(u,\widehat\theta)\psi(u,\widehat\theta)h(u,\widehat\theta)\,du\Big\}, \qquad (2.7)$$

where $\phi(s,\theta)=\frac{\partial}{\partial\gamma}\log h(s,\theta,\gamma_0)$. There is also a nonparametric alternative. See Hjort (1990, section 5) and Koning (1991, section 4) for details, and 3.2 and 5.2 below for illustrations.

The actual implementation of this case is reasonably similar to that of Type B above. One uses

$$\mathrm{NLH}_C(t) = n^{-1/2}\Big\{\int_0^t G_n(s)\,dN(s) - \sum_{j=1}^n\int_0^{t_j\wedge t} G_n(s)h(s,\widehat\theta)\,ds\Big\}/\widehat\kappa_C(t),$$

where

$$\widehat\kappa_C(t)^2 = n^{-1}\sum_{j=1}^n\int_0^{t_j\wedge t} G_n(s)^2h(s,\widehat\theta)\,ds - \Big(n^{-1}\sum_{j=1}^n\int_0^{t_j\wedge t} G_n(s)\psi(s,\widehat\theta)h(s,\widehat\theta)\,ds\Big)'\widehat\Sigma_{\rm pm}^{-1}\Big(n^{-1}\sum_{j=1}^n\int_0^{t_j\wedge t} G_n(s)\psi(s,\widehat\theta)h(s,\widehat\theta)\,ds\Big).$$

There is also a nonparametric alternative for $\kappa(t)$ estimation, for cases where calculating the above becomes too complicated.

2.6. DISCUSSION. The parametric plug-in versions are more statistically precise estimators than the nonparametric ones, and are chosen whenever they are not too cumbersome to compute. This means choosing Type A plots with (2.5) and Type B plots with (2.6). For some models, like the gamma and the log-normal, these can only be computed with numerical integration, however, in which case the nonparametric options are easier, involving only finite sums of explicit terms. We also note that positivity of the various $\kappa(.)$-estimate functions as given here is guaranteed, unlike for other choices that may seem natural and are as permissible from the asymptotic statistics point of view. See Remark 8E.

3. Special models.

3.1. A COMPLETELY SPECIFIED HAZARD FUNCTION. Sometimes one wants to compare the life-time distribution for a group of individuals with some established norm, say one with hazard rate $h_0(.)$. In this case the limit process of $D_n(.)$ is simpler than in the case with estimated parameters. It is $D(t) = \int_0^t \{k(s)/y(s)\}\,\mathrm{d}V(s)$, which is simply time-transformed Brownian motion, $W(\kappa(t)^2)$, where $\kappa(t)^2 = \int_0^t \{k(s)^2/y(s)\}h_0(s)\,\mathrm{d}s$. The test curve becomes

$$\mathrm{NLH}(t) = \sqrt{n}\int_0^t K_n(s)\{\mathrm{d}\widehat{H}(s) - h_0(s)\,\mathrm{d}s\}\Big/\Big\{\int_0^t K_n(s)^2\{n/Y(s)\}h_0(s)\,\mathrm{d}s\Big\}^{1/2}, \tag{3.1}$$

with appropriate simplifications for Type A and Type B. Its limit is a time-transformed normalised Brownian motion, see Remark 8A.

3.2. TESTING THE CONSTANT HAZARD RATE MODEL. Let the model be a constant rate $h(s,\theta) = \theta$. Then $H(t,\theta) = \theta t$ and $\psi(s,\theta) = 1/\theta$. The ML estimator solves $\int_0^\tau \{\mathrm{d}N(s) - Y(s)\theta\,\mathrm{d}s\} = 0$, that is,

$$\widehat{\theta} = \frac{N(\tau)}{\int_0^\tau Y(s)\,\mathrm{d}s} = \frac{\sum_{j=1}^n \delta_j}{\sum_{j=1}^n t_j}.$$

In this case both the parametric and nonparametric estimates of $\Sigma = \sigma^2 = \int_0^\tau y(s)\{1/\theta^2\}\theta\,\mathrm{d}s$ become equal to $\widehat{\sigma}^2 = n^{-1}(\sum_{j=1}^n t_j)^2/\sum_{j=1}^n \delta_j$. Note also that $\widehat{\sigma}^2 = 1/\widehat{\theta}^2$ in the case of non-censored data.

The NLH-plot of Type A for exponentiality is

$$\sqrt{n}\{\widehat{H}(t) - \widehat{\theta}t\}/\widehat{\kappa}_A(t), \tag{3.2}$$

where the parametric plug-in estimator for the variance is

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n}{Y(s)}\widehat{\theta}\,\mathrm{d}s - t^2/\widehat{\sigma}^2 = \widehat{\theta}\Big\{\int_0^t \frac{n}{Y(s)}\,\mathrm{d}s - \frac{n}{N(\tau)}\widehat{\theta}t^2\Big\}.$$

The nonparametric alternative is

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n\,\mathrm{d}N(s)}{Y(s)^2} - \Big\{\int_0^t \frac{1}{\widehat{\theta}}\frac{\mathrm{d}N(s)}{Y(s)}\Big\}^2/\widehat{\sigma}^2 = \int_0^t \frac{n\,\mathrm{d}N(s)}{Y(s)^2} - \{\widehat{H}(t)/\widehat{\theta}\}^2/\widehat{\sigma}^2.$$

Similarly there is a NLH-plot of Type B for exponentiality:

$$n^{-1/2}\Big\{N(t) - \int_0^t Y(s)\widehat{\theta}\,\mathrm{d}s\Big\}/\widehat{\kappa}_B(t) = n^{-1/2}\Big\{N(t) - \sum_{j=1}^n \widehat{\theta}(t_j \wedge t)\Big\}/\widehat{\kappa}_B(t). \tag{3.3}$$

The parametric choice for variance estimation is

$$\widehat{\kappa}_B(t)^2 = n^{-1}\sum_{j=1}^n (t_j \wedge t)\widehat{\theta} - \Big(n^{-1}\sum_{j=1}^n (t_j \wedge t)\Big)^2/\widehat{\sigma}^2,$$

while the nonparametric version is $n^{-1}N(t) - (n^{-1}N(t)/\widehat{\theta})^2/\widehat{\sigma}^2$.

Let us finally include a version of the Type C plots. If an overall test for exponentiality is sought that at the same time is good at detecting departures in the direction of Weibullness, then the general recipe described in 2.5 above leads to the tailor-made weight function $K_n(s) = \{Y(s)/n\}G_n(s)$, with

$$G_n(s) = \log s - \widehat{\phi} = \log s - \frac{\int_0^\tau Y(s)\log s\,\mathrm{d}s}{\int_0^\tau Y(s)\,\mathrm{d}s} = \log s - \frac{\sum_{j=1}^n (t_j \log t_j - t_j)}{\sum_{j=1}^n t_j}.$$

The result is

$$\mathrm{NLH}_C(t) = n^{-1/2}\Big\{\int_0^t (\log s - \widehat{\phi})\,\mathrm{d}N(s) - \sum_{j=1}^n \int_0^{t_j\wedge t} (\log s - \widehat{\phi})\widehat{\theta}\,\mathrm{d}s\Big\}/\widehat{\kappa}_C(t),$$

where

$$\widehat{\kappa}_C(t)^2 = n^{-1}\sum_{j=1}^n \int_0^{t_j\wedge t} (\log s - \widehat{\phi})^2\widehat{\theta}\,\mathrm{d}s - \Big\{n^{-1}\sum_{j=1}^n \int_0^{t_j\wedge t} (\log s - \widehat{\phi})\,\mathrm{d}s\Big\}^2/\widehat{\sigma}^2.$$

And of course the integrals here can be explicitly evaluated.

These test curves are illustrated in examples 9.1–9.3.

Note that the case of a model $h(s,\theta) = \theta h_0(s)$ specifying proportionality to a specified $h_0(s)$ can be treated in the same way, with small adjustments.

3.3. A CLASS OF TWO-PARAMETER MODELS. Suppose $h(s,\theta,\beta) = \theta h_0(s,\beta)$ for a specified h_0 function in terms of a single extra β parameter. Then $H(t,\theta,\beta) = \theta\int_0^t h_0(s,\beta)\,\mathrm{d}s = \theta H_0(t,\beta)$, and $\log h$ has derivatives $\psi(s,\theta,\beta) = (\theta^{-1}, \psi_0(s,\beta))$. The ML estimates maximise

$$L(\theta,\beta) = \int_0^\tau \big[\{\log\theta + \log h_0(s,\beta)\}\,\mathrm{d}N(s) - Y(s)\theta h_0(s,\beta)\,\mathrm{d}s\big]. \tag{3.4}$$

We see that $\widehat{\theta}(\beta) = N(\tau)/\int_0^\tau Y(s)h_0(s,\beta)\,\mathrm{d}s$, and $\widehat{\beta}$ maximises the resulting profile log-likelihood, so this can be made into a one-dimensional problem. See the computational notes of the Appendix. Next we need to estimate the matrix

$$\Sigma = \int_0^\tau \begin{pmatrix} 1/\theta \\ \psi_0(s,\beta)\end{pmatrix}\begin{pmatrix} 1/\theta \\ \psi_0(s,\beta)\end{pmatrix}' y(s)\theta h_0(s,\beta)\,\mathrm{d}s.$$

The parametric estimate is

$$\widehat{\Sigma}_{\mathrm{pm}} = n^{-1}\sum_{j=1}^n \int_0^{t_j} \begin{pmatrix} 1/\widehat{\theta}^2 & \psi_0(s,\widehat{\beta})/\widehat{\theta} \\ \psi_0(s,\widehat{\beta})/\widehat{\theta} & \psi_0(s,\widehat{\theta})^2\end{pmatrix}\widehat{\theta}h_0(s,\widehat{\beta})\,\mathrm{d}s, \tag{3.5}$$

and in many cases of interest the integrals can be evaluated explicitly. The nonparametric version is the finite sum

$$\widehat{\Sigma}_{\mathrm{np}} = n^{-1}\sum_{j=1}^n \begin{pmatrix} 1/\widehat{\theta}^2 & \psi_0(t_j,\widehat{\beta})/\widehat{\theta} \\ \psi_0(t_j,\widehat{\beta})/\widehat{\theta} & \psi_0(t_j,\widehat{\beta})^2\end{pmatrix}\delta_j.$$

NLH-plots of Type A plots are now of the form $\sqrt{n}\{\widehat{H}(t) - \widehat{\theta}H_0(t,\widehat{\beta})\}/\widehat{\kappa}_A(t)$ with

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n}{Y(s)}\widehat{\theta}h_0(s,\widehat{\beta})\,\mathrm{d}s - \begin{pmatrix} H_0(t,\widehat{\beta}) \\ \widehat{\theta}H_0^*(t,\widehat{\beta}) \end{pmatrix}' \widehat{\Sigma}_{\mathrm{pm}}^{-1} \begin{pmatrix} H_0(t,\widehat{\beta}) \\ \widehat{\theta}H_0^*(t,\widehat{\beta}) \end{pmatrix}, \tag{3.6}$$

in which $H_0^*(t,\beta) = \frac{\partial}{\partial\beta}H_0(t,\beta)$. Similarly NLH-plots of Type B are of the form $n^{-1/2}\{N(t) - \int_0^t Y(s)\widehat{\theta}h_0(s,\widehat{\beta})\,\mathrm{d}s\}/\widehat{\kappa}_B(t)$, with

$$\widehat{\kappa}_B(t)^2 = n^{-1}\sum_{j=1}^n \widehat{\theta}H_0(t_j\wedge t,\widehat{\beta}) - \begin{pmatrix} n^{-1}\sum_{j=1}^n H_0(t_j\wedge t,\widehat{\beta}) \\ n^{-1}\sum_{j=1}^n \widehat{\theta}H_0^*(t_j\wedge t,\widehat{\beta}) \end{pmatrix}' \widehat{\Sigma}_{\mathrm{pm}}^{-1} \begin{pmatrix} n^{-1}\sum_{j=1}^n H_0(t_j\wedge t,\widehat{\beta}) \\ n^{-1}\sum_{j=1}^n \widehat{\theta}H_0^*(t_j\wedge t,\widehat{\beta}) \end{pmatrix}. \tag{3.7}$$

The nonparametric options for variance estimation are also available and sometimes simpler to compute.

3.4. WEIBULL. Here $h(s,\theta,\beta) = \theta\beta s^{\beta-1}$ with cumulative hazard $H(t,\theta,\beta) = \theta t^\beta$. The ML estimates maximise $\sum_{j=1}^n[\{\log\theta + \log\beta + (\beta-1)\log t_j\}\delta_j - \theta t_j^\beta]$. The construction above leads after some calculations to

$$\widehat{\Sigma}_{\mathrm{pm}} = \begin{pmatrix} n^{-1}\widehat{\theta}^{-1}\sum_{j=1}^n t_j^{\widehat{\beta}} & n^{-1}\widehat{\beta}^{-1}\sum_{j=1}^n t_j^{\widehat{\beta}}\log t_j^{\widehat{\beta}} \\ n^{-1}\widehat{\beta}^{-1}\sum_{j=1}^n t_j^{\widehat{\beta}}\log t_j^{\widehat{\beta}} & n^{-1}\widehat{\theta}\widehat{\beta}^{-2}\sum_{j=1}^n t_j^{\widehat{\beta}}\{1+(\log t_j^{\widehat{\beta}})^2\} \end{pmatrix}.$$

We are now in a position to properly define NLH-plots for Weibullness. *Type A* uses

$$\mathrm{NLH}_A(t) = \sqrt{n}\{\widehat{H}(t) - \widehat{\theta}t^{\widehat{\beta}}\}/\widehat{\kappa}_A(t), \tag{3.8}$$

where

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n}{Y(s)}\widehat{\theta}\widehat{\beta}s^{\widehat{\beta}-1}\,\mathrm{d}s - \begin{pmatrix} t^{\widehat{\beta}} \\ \widehat{\theta}\widehat{\beta}^{-1}t^{\widehat{\beta}}\log t^{\widehat{\beta}} \end{pmatrix}' \widehat{\Sigma}_{\mathrm{pm}}^{-1} \begin{pmatrix} t^{\widehat{\beta}} \\ \widehat{\theta}\widehat{\beta}^{-1}t^{\widehat{\beta}}\log t^{\widehat{\beta}} \end{pmatrix}.$$

And the NLH-plot of *Type B* is

$$n^{-1/2}\Big\{N(t) - \int_0^t Y(s)\widehat{\theta}\widehat{\beta}s^{\widehat{\beta}-1}\,\mathrm{d}s\Big\}/\widehat{\kappa}_B(t) = n^{-1/2}\Big\{N(t) - \sum_{j=1}^n \widehat{\theta}(t_j\wedge t)^{\widehat{\beta}}\Big\}/\widehat{\kappa}_B(t), \tag{3.9}$$

where the denominator is the square root of

$$n^{-1}\sum_{j=1}^n \widehat{\theta}(t_j\wedge t)^{\widehat{\beta}} - \begin{pmatrix} n^{-1}\sum_{j=1}^n (t_j\wedge t)^{\widehat{\beta}} \\ n^{-1}\sum_{j=1}^n \widehat{\theta}(t_j\wedge t)^{\widehat{\beta}}\log(t_j\wedge t) \end{pmatrix}' \widehat{\Sigma}_{\mathrm{pm}}^{-1} \begin{pmatrix} n^{-1}\sum_{j=1}^n (t_j\wedge t)^{\widehat{\beta}} \\ n^{-1}\sum_{j=1}^n \widehat{\theta}(t_j\wedge t)^{\widehat{\beta}}\log(t_j\wedge t) \end{pmatrix}.$$

There are also nonparametric alternatives to these variance estimators.

An illustration is given in example 9.2.

3.5. FRAILTY MODELS. Suppose each individual has his own constant hazard rate, and that these are distributed in the population under study as a gamma distribution with mean θ and variance $\theta\beta$. Then the hazard rate for observed individuals is of the form $h(s,\theta,\beta) = \theta/(1+\beta s)$, nonincreasing a priori. The cumulative increases logarithmically. We call this the simple frailty model.

The ML estimates are found by maximising $\sum_{j=1}^n [\{\log\theta - \log(1+\beta t_j)\}\delta_j - \theta\beta^{-1}\log(1+\beta t_j)]$. With some efforts the entries of $\widehat{\Sigma}_{\rm pm}$ are found:

$$\widehat{\Sigma}_{1,1} = n^{-1}\sum_{j=1}^n \widehat{\theta}^{-1}\widehat{\beta}^{-1}\log(1+\widehat{\beta}t_j),$$

$$\widehat{\Sigma}_{1,2} = -n^{-1}\sum_{j=1}^n \widehat{\beta}^{-2}\{\log(1+\widehat{\beta}t_j) + (1+\widehat{\beta}t_j)^{-1} - 1\},$$

$$\widehat{\Sigma}_{2,2} = n^{-1}\sum_{j=1}^n \widehat{\theta}\widehat{\beta}^{-3}\{\log(1+\widehat{\beta}t_j) + 2(1+\widehat{\beta}t_j)^{-1} - \tfrac{1}{2}(1+\widehat{\beta}t_j)^{-2} - \tfrac{3}{2}\}.$$

NLH-plots to check the fit of the model can now be properly defined, parallelling (3.8) and (3.9). The Type A plot is

$$\text{NLH}_A(t) = \sqrt{n}\{\widehat{H}(t) - \widehat{\theta}\widehat{\beta}^{-1}\log(1+\widehat{\beta}t)\}/\widehat{\kappa}_A(t), \tag{3.10}$$

where

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n}{Y(s)}\frac{\widehat{\theta}}{1+\widehat{\beta}s}\,ds - \widehat{C}(t)'\widehat{\Sigma}_{\rm pm}^{-1}\widehat{C}(t),$$

in which

$$\widehat{C}(t) = \begin{pmatrix} \widehat{\beta}^{-1}\log(1+\widehat{\beta}t) \\ \widehat{\theta}\widehat{\beta}^{-2}\{\frac{\widehat{\beta}t}{1+\widehat{\beta}t} - \log(1+\widehat{\beta}t)\} \end{pmatrix}.$$

And the Type B plot uses

$$n^{-1/2}\Big\{N(t) - \int_0^t Y(s)\frac{\widehat{\theta}}{1+\widehat{\beta}s}\,ds\Big\}/\widehat{\kappa}_B(t) = n^{-1/2}\Big\{N(t) - \sum_{j=1}^n \widehat{\theta}\widehat{\beta}^{-1}\log(1+\widehat{\beta}(t_j\wedge t))\Big\}/\widehat{\kappa}_B(t), \tag{3.11}$$

where

$$\widehat{\kappa}_B(t)^2 = n^{-1}\sum_{j=1}^n \widehat{\theta}\widehat{\beta}^{-1}\log(1+\widehat{\beta}(t_j\wedge t)) - \Big(n^{-1}\sum_{j=1}^n \widehat{C}(t_j\wedge t)\Big)'\widehat{\Sigma}_{\rm pm}^{-1}\Big(n^{-1}\sum_{j=1}^n \widehat{C}(t_j\wedge t)\Big).$$

An illustration is given in example 9.3.

This simple frailty model should by its motivation and construction have $\beta \geq 0$. If the likelihood is maximised only in this region then the ML estimator will be equal to zero with positive probability, and one encounters the more involved 'corner asymptotics' problems associated with zero not being an inner point of the parameter space. In particular the plots would not be approximately standard normal any more. We avoid these difficulties by allowing an expanded parameter space $\beta > -\varepsilon$, where $1/\varepsilon$ is twice the largest observed t_j. And of course if the Nelson–Aalen plot for data suggests an increasing hazard rate one should immediately abandon the frailty model.

There are more general models based on frailty distributions and that have found uses in survival analysis and demography, see Aalen (1992). One important class of models is as follows. Let $\lambda(t)$ be a hazard rate function with cumulative $\Lambda(t) = \int_0^t \lambda(s)\,ds$. The individuals in the population under study have all hazard rates of the type $Z\lambda(t)$, but with Z, the unobservable frailty factor, varying from individual to individual. When Z has the particular compound Poisson

distribution considered by Aalen, with certain parameters α and δ, then the life-time of a randomly chosen individual has

$$\begin{aligned} \text{hazard rate } h(t) &= \frac{\lambda(t)}{\{1+(\delta/\alpha)\Lambda(t)\}^{\alpha}}, \\ \text{with cumulative } H(t) &= \frac{\alpha}{\alpha-1}\frac{1}{\delta}\Big[1-\Big(1+\frac{\delta}{\alpha}\Lambda(t)\Big)^{1-\alpha}\Big]. \end{aligned} \tag{3.12}$$

The last formula is valid for $\alpha \neq 1$. For $\alpha = 1$ one has $H(t) = \delta^{-1}\log\{1+\delta\Lambda(t)\}$, and the special case above is of this form. If the underlying $\lambda(.)$ is parametrised with two parameters then the scheme above leads to a four-parameter hazard model, for example. And the validity of these can all be tested with the NLH plot machinery. See example 9.6 for an illustration.

3.6. GOMPERTZ. Here $h(s) = \theta\exp(\beta s)$ on $[0,\tau]$. The ML estimators maximise $\sum_{j=1}^{n}[(\log\theta+\beta t_j)\delta_j - \theta\beta^{-1}\{\exp(\beta t_j)-1\}]$, and this function is actually concave in $(\log\theta,\beta)$. The entries of $\widehat{\Sigma}_{\rm pm}$ are

$$\begin{aligned} \widehat{\Sigma}_{1,1} &= n^{-1}\widehat{\theta}^{-1}\sum_{j=1}^{n}\widehat{\beta}^{-1}\{\exp(\widehat{\beta}t_j)-1\}, \\ \widehat{\Sigma}_{1,2} &= n^{-1}\sum_{j=1}^{n}\widehat{\beta}^{-2}\{(\widehat{\beta}t_j-1)\exp(\widehat{\beta}t_j)+1\}, \\ \widehat{\Sigma}_{2,2} &= n^{-1}\widehat{\theta}\sum_{j=1}^{n}\widehat{\beta}^{-3}[\{(\widehat{\beta}t_j)^2-2\widehat{\beta}t_j+2\}\exp(\widehat{\beta}t_j)-2]. \end{aligned}$$

And quite similarly to the cases above the A-plot and the B-plot take the forms

$$\sqrt{n}[\widehat{H}(t)-\widehat{\theta}\widehat{\beta}^{-1}\{\exp(\widehat{\beta}t)-1\}]/\widehat{\kappa}_A(t) \quad\text{and}\quad n^{-1/2}\Big\{N(t)-\int_0^t Y(s)\widehat{\theta}\exp(\widehat{\beta}s)\,ds\Big\}/\widehat{\kappa}_B(t),$$

in which

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{n}{Y(s)}\widehat{\theta}\exp(\widehat{\beta}s)\,ds - \widehat{C}(t)'\widehat{\Sigma}_{\rm pm}^{-1}\widehat{C}(t)$$

and

$$\widehat{\kappa}_B(t)^2 = n^{-1}\sum_{j=1}^{n}\widehat{\theta}\widehat{\beta}^{-1}\{\exp(\widehat{\beta}(t_j\wedge t))-1\} - \Big(n^{-1}\sum_{j=1}^{n}\widehat{C}(t_j\wedge t)\Big)'\widehat{\Sigma}_{\rm pm}^{-1}\Big(n^{-1}\sum_{j=1}^{n}\widehat{C}(t_j\wedge t)\Big).$$

This time

$$\widehat{C}(t) = \begin{pmatrix} \widehat{\beta}^{-1}\{\exp(\widehat{\beta}t)-1\} \\ \widehat{\theta}\widehat{\beta}^{-2}\{(\widehat{\beta}t-1)\exp(\widehat{\beta}t)+1\} \end{pmatrix}.$$

3.7. GAMMA. The gamma distribution has density $f(t,\alpha,\theta) = \{\theta^{\alpha}/\Gamma(\alpha)\}t^{\alpha-1}\exp(-\theta t)$, and is a useful class to work with in connection with life history data. In some situations a priori reasons could suggest a known integer value for the shape parameter α, in which case one also uses the name Erlang distribution. The constant hazard model corresponds to $\alpha = 1$, and with $\alpha = 3$, for example, the hazard function $h(t,\alpha,\theta) = f(t,\alpha,\theta)/\{1-F(t,\alpha,\theta)\}$ becomes $\theta\,\frac{1}{2}\theta^2t^2/\{1+\theta t+\frac{1}{2}\theta^2t^2\}$, smoothly increasing from 0 to its asymptote θ. It is not very difficult to construct NLH-plots for testing the fit of this one-parameter hazard model, or the other models with known integer α.

The general two-parameter model also fits into our general framework, but its analysis is more cumbersome than for the previous examples in that its hazard function and its derivates

w.r.t. model parameters involve special mathematical functions like the incomplete gamma integral. The cumulative distribution is $F(t,\alpha,\theta) = F_0(\theta t,\alpha)$, where $F_0(t,\alpha) = \int_0^t \Gamma(\alpha)^{-1} u^{\alpha-1} e^{-u}\,du$, for example. This function is available in S-Plus and other computer package libraries. Let us present the necessary formulae. First, the ML estimators maximise

$$\sum_{j:\delta_j=1} \{\alpha \log(\theta t_j) - \log\Gamma(\alpha) - \theta t_j\} + \sum_{j:\delta_j=0} \log\{1 - F_0(\theta t_j,\alpha)\}.$$

In this case the parametric options for estimating Σ and the $\kappa(.)$ functions are complicated, although they can be managed via numerical integrations, and the nonparametric options are easier. The two plots to test gamma-ness become

$$\sqrt{n}[\widehat{H}(t) + \log\{1 - F_0(\widehat{\theta}t,\widehat{\alpha})\}]/\widehat{\kappa}_A(t) \quad \text{and} \quad n^{-1/2}\Big[N(t) + \sum_{j=1}^{n} \log\{1 - F_0(\widehat{\theta}(t_j \wedge t),\widehat{\alpha})\}\Big]/\widehat{\kappa}_B(t),$$

where the denominators involve $\widehat{\Sigma}_{\rm np}$ and are the nonparametric options as given generally in 2.3 and 2.4. These are finite sums involving the model's $\psi(t,\alpha,\theta)$ function, with components

$$\begin{aligned}
\psi_\alpha(t,\alpha,\theta) &= \log(\theta t) - \psi(\alpha) + \{1 - F_0(\theta t,\alpha)\}^{-1} \int_0^t \{\log(\theta s) - \psi(\alpha)\} f(s,\alpha,\theta)\,ds \\
&= \log(\theta t) - \psi(\alpha) + \{1 - F_0(\theta t,\alpha)\}^{-1}\{F_0^*(\theta t,\alpha) - \psi(\alpha) F_0(\theta t,\alpha)\}, \\
\psi_\theta(t,\alpha,\theta) &= \theta^{-1}(\alpha - \theta t) + \{1 - F_0(\theta t,\alpha)\}^{-1} \int_0^t \theta^{-1}(\alpha - \theta s) f(s,\alpha,\theta)\,ds \\
&= \theta^{-1}(\alpha - \theta t) + \{1 - F_0(\theta t,\alpha)\}^{-1}\theta^{-1}(\theta t)^\alpha \exp(-\theta t)/\Gamma(\alpha).
\end{aligned}$$

Here $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$, and in addition to $F_0(t,\alpha)$ one needs its relative $F_0^*(t,\alpha) = \int_0^t \Gamma(\alpha)^{-1} \log u\, u^{\alpha-1} e^{-u}\,du$.

3.8. OTHER MODELS. Constructing NLH curves for other parametric models for survival data should not be difficult given the general machinery and the examples above. Some useful models include the Gompertz–Makeham one with $h(s) = a + b\exp(cs)$, the lognormal, the log-logistic, the inverse normal, and the three-parameter $h(s) = \exp(\beta_0 + \beta_1 s + \beta_2 s^2)$, capable of representing a wide variety of hazard curves on $[0,\tau]$.

4. NLH-plots for the parametric Cox regression model. Suppose covariate information in the form of features or measurements $z_j = (z_{j,1},\ldots,z_{j,p})'$ are available for individual j, in addition to the possibly right-censored life-time information (t_j,δ_j). The widely used semiparametric Cox regression model postulates that individual j has hazard rate of the form $h_j(s) = h_0(s)\exp(\beta' z_j)$, but without further specifying $h_0(.)$, the hazard rate for individuals with covariate zero. NLH-plots can be constructed to check the validity of a *parametric* Cox model, where

$$h_j(s) = h(s,\theta)\exp(\beta_1 z_{j,1} + \cdots + \beta_p z_{j,p}) \quad \text{for } j = 1,\ldots,n. \tag{4.1}$$

If such a parametric form for the baseline hazard can be validated it makes for a better understanding of the survival mechanisms under study, and also makes it possible to estimate the β parameters with increased precision.

We shall be content to illustrate the use of our plots for the important special case where the baseline hazard is a constant θ. This is sometimes a quite effective model. We also assume that the

covariates are constant in time. Generalisations to time-dependent covariates and to other $h(s,\theta)$ models are reasonably straightforward, in view of the theory developed in Hjort (1990, section 6).

The ML estimates are computed from the log-likelihood

$$\sum_{j=1}^n \int_0^\tau \{(\log\theta+\beta' z_j)\,\mathrm{d}N_j(s) - Y_j(s)\theta\exp(\beta' z_j)\,\mathrm{d}s\} = \sum_{j=1}^n \{(\log\theta+\beta' z_j)\delta_j - \theta\exp(\beta' z_j)\,t_j\}.$$

Here $Y_j(s) = I\{t_j \geq s\}$ is the at risk indicator and $N_j(t) = I\{t_j \leq t, \delta_j = 1\}$ the 0–1 counting process for no. j, summing to the $N = \sum_{j=1}^n N_j$ counting process. Define

$$\widehat{\Sigma}_{1,1} = n^{-1}\widehat{\theta}^{-1}\sum_{j=1}^n \exp(\widehat{\beta}' z_j)\,t_j,\ \widehat{\Sigma}_{2,1} = n^{-1}\sum_{j=1}^n z_j\exp(\widehat{\beta}' z_j)\,t_j,\ \widehat{\Sigma}_{2,2} = n^{-1}\widehat{\theta}\sum_{j=1}^n z_j z_j'\exp(\widehat{\beta}' z_j)\,t_j.$$

These are the blocks of a consistent estimator $\widehat{\Sigma}$ for the inverse variance matrix for the limit distribution of $\sqrt{n}(\widehat{\theta}-\theta, \widehat{\beta}-\beta)$, under the (4.1) model, as can be shown from more general results of Hjort (1990, section 6).

A NLH-plot of Type A takes as its starting point

$$D_n(t) = \sqrt{n}\{\widehat{H}(t) - \widehat{\theta}t\} = \sqrt{n}\Big\{\int_0^t \frac{\mathrm{d}N(s)}{\sum_{j=1}^n Y_j(s)\exp(\widehat{\beta}' z_j)} - \widehat{\theta}t\Big\}. \tag{4.2}$$

One might also argue in favour of plugging in the maximum partial likelihood estimator $\widetilde{\beta}$ here, but the following formulae relate to the ML estimate from the parametric regression. Several natural consistent estimates for its limit variance can be constructed along the general lines of Hjort (1990, section 6). One version is as follows. Let $R(s,\beta) = n^{-1}\sum_{j=1}^n Y_j(s)\exp(\beta' z_j)$, $R_{(1)}(s,\beta) = n^{-1}\sum_{j=1}^n Y_j(s) z_j\exp(\beta' z_j)$, and finally $E(s,\beta) = R_{(1)}(s,\beta)/R(s,\beta)$. These are next used to define

$$\widehat{C}_A(t) = \begin{pmatrix} t \\ \int_0^t E(s,\widehat{\beta})\widehat{\theta}\,\mathrm{d}s \end{pmatrix},\quad \widehat{C}_B(t) = \begin{pmatrix} \int_0^t R(s,\widehat{\beta})\,\mathrm{d}s \\ \int_0^t R_{(1)}(s,\widehat{\beta})\,\widehat{\theta}\,\mathrm{d}s \end{pmatrix}.$$

In the end define

$$\widehat{\kappa}_A(t)^2 = \int_0^t \frac{\widehat{\theta}\,\mathrm{d}s}{R(s,\widehat{\beta})} - \widehat{C}_A(t)'\widehat{\Sigma}^{-1}\widehat{C}_A(t),$$

and use $\mathrm{NLH}_A(t) = \sqrt{n}\{\widehat{H}(t) - \widehat{\theta}t\}/\widehat{\kappa}_A(t)$.

Similarly a NLH-plot of Type B would use

$$D_n(t) = n^{-1/2}\Big\{N(t) - \sum_{j=1}^n \int_0^t Y_j(s)\widehat{\theta}\exp(\widehat{\beta}' z_j)\,\mathrm{d}s\Big\} = n^{-1/2}\Big\{N(t) - \sum_{j=1}^n \widehat{\theta}\exp(\widehat{\beta}' z_j)(t_j\wedge t)\Big\}.$$

In this case one may use $\widehat{\kappa}_B(t)^2 = \int_0^t R(s,\widehat{\beta})\widehat{\theta}\,\mathrm{d}s - \widehat{C}_B(t)'\widehat{\Sigma}^{-1}\widehat{C}_B(t)$, with $\widehat{C}_B(t)$ as above. The simplest way to compute the various integrals appearing here is by noting that the integrands are constant over the ordered intervals $(t_{j-1}, t_j]$, turning the integrals into finite sums.

We also note that the log-likelihood is concave once reparametrised with $\theta = \exp(\beta_0)$, making computation relatively easy, and, incidentally, making rigorous proofs of the necessary large-sample distribution statements easier. That the required $\mathrm{NLH}_A(t) \to_d \mathcal{N}\{0,1\}$ and $\mathrm{NLH}_B(t) \to_d \mathcal{N}\{0,1\}$ hold, under standard regularity conditions, follows indeed from the general results of Hjort (1990,

section 6), but can be proven relatively easily under almost minimal regularity assumptions when one uses methods of Hjort and Pollard (1993, section 7).

See example 9.5 for an illustration of these techniques.

5. Detection power. Some analysis makes it possible to predict the behaviour of NLH-plots outside model conditions. Information in this section should also help in pinpointing exactly which aspects of a proposed model are wrong, in cases where the test curve determinedly wanders outside the ± 1.96 band.

5.1. A FIXED ALTERNATIVE. Suppose the true hazard is $h(.)$ rather than belonging to the parametric class $h(.,\theta)$. Then one can establish

$$D_n(t)/\sqrt{n} = \int_0^t K_n(s)\{\mathrm{d}\widehat{H}(s) - h(s,\widehat{\theta})\,\mathrm{d}s\} \to_p \pi(t) = \int_0^t k(s)\{h(s) - h(s,\theta_0)\}\,ds, \qquad (5.1)$$

and there is uniform convergence in probability under suitable assumptions, see Hjort (1990, section 5). For Type A plots this is simply $H(t) - H(t,\theta_0)$. The θ_0 parameter value here is not 'true' but rather 'least false' in the sense of making $h(s,\theta_0)$ the best parametric approximant to $h(s)$, see Hjort (1992) for details necessary to prove this and for the precise distance measure which is being minimised.

This result shows that one can expect the NLH plot to decrease in regions where $h(t) < h(t,\theta_0)$ and increase in regions where $h(t) > h(t,\theta_0)$ (this assumes $k(t)$ positive, as with Type A and Type B). It also indicates that $\mathrm{NLH}(t)/\sqrt{n}$ is estimating a well-defined discrepancy function $\pi(t)/\kappa_0(t)$, and which is zero only if the model being tested is correct. Yet another pleasing theoretical consequence is that every departure from the model will be detected by the NLH plots with probability 1 as n grows; a non-zero $\pi(t)$ will send $|\mathrm{NLH}(t)|$ towards infinity with speed $\sqrt{n}|\pi(t)|$.

If the model being tested is that of a constant hazard θ, then $\theta_0 = \int_0^\tau y(s)h(s)\,\mathrm{d}s / \int_0^\tau y(s)\,\mathrm{d}s$. Each of the NLH plots of 3.2 will then tend to decrease in regions where the true hazard is less than θ_0 and tend to increase when it is greater than θ_0. If the true state of affairs is a Weibull with $\beta > 1$, for example, then the expected plot decreases up to certain t_0 and increases afterwards. The opposite happens if the true hazard is a Weibull with $\beta < 1$.

5.2. LOCAL ALTERNATIVES. Suppose $h_n(s) \doteq h(s,\theta)\{1 + \phi(s,\theta)\delta/\sqrt{n}\}$. If one for example considers model alternatives of the form $h(s,\theta,\beta)$, where $\beta = \beta_0$ gives back $h(s,\theta)$, then $\phi(s,\theta) = \frac{\partial}{\partial\beta}\log h(s,\theta,\beta_0)$. Methods of Hjort (1990, section 5) can be used to prove

$$\mathrm{NLH}(t) \to_d D(t)/\kappa(t) + \delta a(t)/\kappa(t) \sim \mathcal{N}\{\delta a(t)/\kappa(t), 1\}, \qquad (5.2)$$

where

$$a(t) = \int_0^t k(s)h(s,\theta)\phi(s,\theta)\,\mathrm{d}s - \Big(\int_0^t k(s)h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big)'\Sigma^{-1}\Big(\int_0^\tau y(s)\phi(s,\theta)h(s,\theta)\psi(s,\theta)\,\mathrm{d}s\Big).$$

This indicates the local detection power of the NLH plots. The optimal choice for weight function is as in (2.7), leading to a test curve of Type C.

For illustration take the constant θ model again, and suppose the true hazard is $\theta\{1 + \phi(s)\delta/\sqrt{n}\}$ for a suitable ϕ function. Then calculations show that $\mathrm{NLH}(t)$ is approximately a normal with variance one and mean parameter

$$\delta\frac{\int_0^t k(s)\{\phi(s) - \bar{\phi}\}\,\mathrm{d}s}{\big[\int_0^t\{k(s)^2/y(s)\}\,\mathrm{d}s - \{\int_0^t k(s)\,\mathrm{d}s\}^2/\int_0^\tau y(s)\,\mathrm{d}s\big]^{1/2}},$$

in which $\bar{\phi} = \int_0^\tau y(s)\phi(s)\,ds / \int_0^\tau y(s)\,ds$. The optimal choice corresponds to $k(s) = y(s)\{\phi(s) - \bar{\phi}\}$.

For another illustration, consider the heterogeneity situation where there is a 'hidden covariate'. If the true hazard for individual i is of the proportional hazard form $h_i(s) = h(s|z_i) = h_0(s)\exp(\beta z_i)$, and the frailty factor $\exp(\beta z_i)$ has a gamma distribution with mean 1 and variance σ^2 (which amounts to gamma parameters equal to $1/\sigma^2$ and $1/\sigma^2$), then calculations show that the hazard rate for a randomly selected individual is of the form $h_0(s)/\{1 + \sigma^2 H_0(s)\}$, where $H_0(.)$ is the cumulative hazard for $h_0(.)$. Suppose in particular that $h_i(s) = \theta\exp(\beta z_i)$ for an unobserved covariate of this form, with population variance $\sigma^2 = \delta/\sqrt{n}$. Then the framework above is appropriate with $h_n(s) \doteq \theta(1 - \theta s\delta/\sqrt{n})$. If we further assume that there is no censoring and the observation window is the full halfline then calculations with $\phi(s,\theta) = -\theta s$, for the B plot, give

$$a(t) = \theta t\exp(-\theta t) \quad \text{and} \quad \kappa(t)^2 = \{1 - \exp(-\theta t)\}\exp(-\theta t).$$

In the end this leads to

$$\text{NLH}_B(t) \approx \mathcal{N}\Big\{\sqrt{n}\sigma^2 \frac{\theta t\exp(-\frac{1}{2}\theta t)}{\{1-\exp(-\theta t)\}^{1/2}}, 1\Big\}. \tag{5.3}$$

This detection power approximation is valid for small σ^2, but illustrates well the plot's ability to detect the presence of a missing covariate. Note also that the whole B plot will tend to lie above the time axis, according to these calculations, if there is such a frailty departure from the constant hazard model. The A plot, on the other hand, will tend to lie above the time axis as long as $\theta t < 2$ and below the time axis when $\theta t > 2$, as borne out by similar calculations. The optimal C plot uses $G_n(s) = 1 - \widehat{\theta} s$, and gives most weight to the smallest and then the largest time values.

Results similar to those given here can also be reached for the parametric Cox regression plots of section 4.

6. General counting process models. Our paper has so far been concerned with right-censored survival times, with or without covariate information. The plots and the results about them can be generalised with surprisingly few modifications to much more general counting process models for life history data, see Hjort (1990). The necessary nonparametric and parametric machinery work specifically for Aalen's multiplicative intensity model, as broadly surveyed in the ABGK book. Examples include models with left truncated entry times, competing risks, and time-inhomogeneous Markov chains, where individuals move between states with certain hazard or transition rates.

The practical consequence, as far as the normalised local hazard plots are concerned, is that the formulae developed earlier are still true as long as $Y(s)/n$ is used everywhere to estimate the asymptotic $y(s)$ function, where $Y(s)$ is the number of individuals at risk (for the particular transition in question) just prior to time s. Formulae based on the $\sum_{j=1}^n I\{t_j \geq s\}$ identity are not true in this more general situation, however. For example, the first expression for $\widehat{\Sigma}_{\text{pm}}$ of (2.4) is still valid in these more general models, but the second is not. We illustrate our plots in a competing risk situation in example 9.3 and in a left-truncated entry study in example 9.4.

An important fact about situations with several transition rates is that the collection of Nelson–Aalen estimates are statistically independent asymptotically. This is a fruit of the martingale theory developed by Aalen. It also follows that the different Type B plots, say, for testing different parametric models for the transition rates under study, become approximately independent. This makes it easier to judge separate models for separate transition mechanisms in a many-state study.

7. Time-discrete models. There are quite similar models and methods for time-discrete survival analysis. Although our emphasis has been clearly on the time-continuous case we take the necessary time out to provide the most important results for the time-discrete case. These are of interest in their own right, since data even on time-continuous phenomena often are recorded on a time-discrete basis, e.g. in demography, the social sciences, national health statistics, and for many censuses of central bureau of statistics type. The results can also be used to construct correction factors to the time-continuous formulae in case where data are grouped in time.

7.1. GENERAL RESULTS. Suppose that life-times observations are recorded only at time points $a_0 < a_1 < \cdots < a_k = \tau$, where τ could be infinity. If a distribution has point mass f_i at a_i then the hazard rate at that point is $h_i = f_i / \sum_{j \geq i} f_j$. Suppose a p-parametric model is considered of the form $h_i = h_i(\theta)$, and let there be ΔN_i observed failures at a_i among the Y_i at risk at that point. Then the log-likelihood can be written $\sum\{\Delta N_i \log h_i + (Y_i - \Delta N_i)\log(1-h_i)\}$, see e.g. Cox and Oakes (1984, chapter 3). The nonparametric estimate is $\widehat{h}_i = \Delta N_i / Y_i$. The parametric ML estimator $\widehat{\theta}$ solves

$$U_n(\theta) = \sum_{a_i \leq \tau} \Big\{ \frac{\Delta N_i}{h_i(\theta)} - \frac{Y_i - \Delta N_i}{1 - h_i(\theta)} \Big\} h_i^*(\theta) = \sum_{a_i \leq \tau} \frac{\Delta N_i - Y_i h_i(\theta)}{h_i(\theta)(1-h_i(\theta))} h_i^*(\theta) = 0, \tag{7.1}$$

where $h_i^*(\theta) = \frac{\partial}{\partial \theta} h_i(\theta)$, a p-vector. One can prove that $\sqrt{n}(\widehat{\theta} - \theta)$ tends to a $\mathcal{N}_p\{0, \Sigma^{-1}\}$, where

$$\Sigma = \sum_{a_i \leq \tau} \frac{r_i}{h_i(\theta)(1-h_i(\theta))} h_i^*(\theta) h_i(\theta)^* = \sum_{a_i \leq \tau} r_i \psi_i(\theta) \psi_i(\theta)' h_i(\theta)/(1-h_i(\theta)), \tag{7.2}$$

writing $\psi_i(\theta) = \frac{\partial}{\partial \theta} \log h_i(\theta) = h_i^*(\theta)/h_i(\theta)$. Here r_i is the limit in probability of Y_i/n, the limiting proportion of individuals present just before time a_i. We omit the proof here, but it proceeds along lines similar to those for the time-continuous case, as presented in Hjort (1992) and ABGK (chapter VI), with the crucial modification that the time-discrete martingale $M(t) = \sum_{a_i \leq t} (\Delta N_i - Y_i h_i)$ has variance process $\langle M, M \rangle(t) = \sum_{a_i \leq t} Y_i h_i (1 - h_i)$ whereas the time-continuous martingale $M(t) = N(t) - \int_0^t Y(s) h(s)\, ds$ has variance process $\langle M, M \rangle(t) = \int_0^t Y(s) h(s)\, ds$. Note the similarity but also the slight correction to the time-continuous case, cf. (2.1)–(2.2).

The weighted hazard difference function takes the form

$$D_n(t) = \sqrt{n} \sum_{a_i \leq t} K_i \{\Delta N_i / Y_i - h_i(\widehat{\theta})\} = \sqrt{n} \sum_{a_i \leq t} (K_i / Y_i) \{\Delta N_i - Y_i h_i(\widehat{\theta})\}, \tag{7.3}$$

for suitable weight function K_i converging in probability to some k_i. Limit theorems for time-discrete martingales can be used to prove

$$D_n(t) \to_d D(t) = \sum_{a_i \leq t} (k_i / r_i) V_i - \Big(\sum_{a_i \leq t} k_i h_i^*(\theta) \Big)' \Sigma^{-1} \sum_{a_i \leq \tau} h_i^*(\theta) V_i / \{h_i(\theta)(1-h_i(\theta))\},$$

where the V_is are independent and $V_i \sim \mathcal{N}\{0, r_i h_i (1-h_i)\}$. The variance is

$$\kappa(t)^2 = \sum_{a_i \leq t} (k_i^2 / r_i) h_i(\theta)(1-h_i(\theta)) - \Big(\sum_{a_i \leq t} k_i h_i^*(\theta) \Big)' \Sigma^{-1} \Big(\sum_{a_i \leq t} k_i h_i^*(\theta) \Big), \tag{7.4}$$

which can be estimated consistently. The end product is therefore a NLH plot of the form $D_n(t)/\widehat{\kappa}(t)$, plotted for each $t = a_i$.

As an illustration, consider the constant hazard rate model $h_i = \theta$. Note that $\theta \in (0,1)$ in the present framework. The estimate is $\widehat{\theta} = \sum_{a_i \le \tau} \Delta N_i / \sum_{a_i \le \tau} Y_i$, and

$$\mathrm{NLH}_B(t) = \frac{n^{-1/2} \sum_{a_i \le t} (\Delta N_i - Y_i \widehat{\theta})}{[\widehat{\theta}(1-\widehat{\theta}) v(\tau)^2 \{\widehat{p}(t) - \widehat{p}(t)^2\}]^{1/2}} \to_d \frac{W^0(p(t))}{\{p(t) - p(t)^2\}^{1/2}} \sim \mathcal{N}\{0,1\}$$

for each $t = a_i$. Here $v(t)^2 = \sum_{a_i \le t} r_i$ and $p(t) = v(t)^2 / v(\tau)^2$.

7.2. OBSERVING TIME-CONTINUOUS PHENOMENA ON A TIME-DISCRETE SCALE. Suppose life-times have a continuous distribution and that a hazard rate model is considered of the form $h(s) = h(s,\theta)$. Assume however that data only are collected on a time-discrete basis, will cells $I_i = [i, i+1)$ for $i = 0, 1, \ldots$. This means a time-discrete framework with hazards

$$h_i(\theta) = F_\theta[i, i+1) / F_\theta[i, \infty) = 1 - \exp[-\{H(i+1,\theta) - H(i,\theta)\}],$$

and the framework above applies. Note that $h_i^*(\theta) = (1 - h_i(\theta)) \frac{\partial}{\partial \theta} \{H(i+1,\theta) - H(i,\theta)\}$. See examples 9.4 and 9.6 for illustrations.

Let us finally mention that the above methods also lead to

$$\Delta D_n(a_i) = \sqrt{n} K_i \{\Delta N_i / Y_i - h_i(\widehat{\theta})\} \to_d \Delta D(a_i) \sim \mathcal{N}\{0, w_i^2\}, \tag{7.5}$$

where $w_i^2 = k_i^2 \{r_i^{-1} h_i(\theta)(1 - h_i(\theta)) - h_i(\theta)^2 \psi_i(\theta)' \Sigma^{-1} \psi_i(\theta)\}$. Plotting $\Delta D_n(t)$ is nugatory in situations with small time cells and correspondingly small values of $\Delta N_i h_i(\widehat{\theta})$, but in many large-scale studies in demography and social sciences one would have substantial Y_is and not too small ΔN_is. Aalen's (1992) two examples are of this sort, for example, and one of these are analysed as example 9.6 below. In such situations plotting of $\sqrt{n}\{\Delta N_i / Y_i - h_i(\widehat{\theta})\} / \widehat{w}_i$ (taking $K_i = 1$) is quite informative regarding the validity of the model.

8. Concluding remarks.

8A. MAXIMAL VALUES OF TEST CURVES. Our test curves will at each given time point by construction stay inside ± 1.96 with probability approximately 0.95. This is a point-wise statement, and the maximal absolute value of the random curve is substantially larger in distribution than the absolute value of a standard normal. In the notation of section 2 the maximal absolute value over $[a_1, a_2]$ tends in distribution to $\max_{a_1 \le t \le a_2} |D(t)| / \kappa(t)$, as a consequence of process convergence. This limit distribution is quite intricate in general, but can be handled in special cases.

Consider a general one-parametric model $h(s,\theta)$, and use NLH plots of Type C with $K_n(s) = n^{-1} Y(s) \psi(s, \widehat{\theta})$. Notice that the $D_n(.)$ process of (1.4) then starts and ends at zero, by definition of the ML estimator. By covariance calculations and Gaußianeity one can show that $D_n(t) \to_d v(\tau) W^0(p(t))$, a scaled and time-transformed Brownian bridge, where $v(t)^2 = \int_0^t y(s) \psi(s,\theta)^2 h(s,\theta)\, ds$ and $p(t) = v(t)^2 / v(\tau)^2$. One can also see from (2.3) that $\kappa(t)^2 = v(\tau)^2 p(t)(1 - p(t))$. It follows that

$$\mathrm{NLH}_C(t) \to_d \frac{D(t)}{\kappa(t)} = \frac{W^0(p(t))}{\{p(t)(1-p(t))\}^{1/2}} \quad \text{over } (0,\tau),$$

using convergence of stochastic processes theory. Consequently

$$M_n = \max_{a_1 \le t \le t_2} |\mathrm{NLH}_C(t)| \to_d M = \max_{b_1 \le u \le b_2} |W^0(u)| / \{u(1-u)\}^{1/2},$$

where $b_1 = p(a_1)$ and $b_2 = p(a_2)$, involving the normalised Brownian bridge. An approximation to this distribution is

$$\Pr\{M \geq m\} \doteq 4\phi(m)/m + \phi(m)(m - m^{-1})\log(c_2/c_1),$$

where $c_1 = b_1/(1-b_1)$ and $c_2 = b_2/(1-b_2)$; see Miller and Siegmund (1982).

If the model considered is $h(s) = \theta h_0(s)$, then $\psi(s,\theta) = 1/\theta$ is constant and factors out, which means that the Type C plot is in fact the same as the Type B plot, and the above applies. Study a Type B plot for constant hazard rate, for example, and let a_1 and a_2 be chosen as the empirical versions of $p(a_1) = 0.10$ and $p(a_2) = 0.90$, where $p(t) = \int_0^t y(s)\,\mathrm{d}s / \int_0^\tau y(s)\,ds$. Then the above shows that the maximal absolute value of the plot, over the $[a_1, a_2]$ interval, exceeds the pointwise 1.96 limit with probability about 0.49, and that an upper 5% limit for this maximum is about 3.05.

Similar and in fact somewhat simpler calculations can be carried out for the fully specified case of 3.1, where the test curve is asymptotically distributed as $W(\kappa(t)^2)/\kappa(t)$, a time-transformed normalised Brownian motion process. Here $\kappa(t)^2 = \int_0^t \{k(s)^2/y(s)\}h_0(s)\,\mathrm{d}s$.

8B. OTHER EMPIRICAL TEST CURVES. Our methods and results rely on the weak convergence results described in section 2. Similar results can be and have been reached for $\sqrt{n}\{\widehat{F}(t) - F(t,\widehat{\theta})\}$, for example, where $\widehat{F}$ is the Kaplan–Meier estimate. Test curves and test statistics can be constructed based on this. We have found it most useful to work with weighted versions of hazard differences instead, however, partly since the hazard rate quantity is more central and more easily generalisable to other survival data models, and partly since results tend to be simpler. Our results belong to the tradition originating with the work by Durbin (1973) for empirical processes with estimated parameters. We find satisfaction in seeing practical and even visual uses of theoretical results in which "interest ... died down when the intractable limit processes asserted themselves", as Pollard (1984, p. 118) remarked.

8C. TESTING VALIDITY OVER A SUBINTERVAL. Suppose that one wishes to test a parametric model only over the subinterval $[a,b]$, perhaps because it is obvious that the model cannot hold to the left of time a. One can then use a C plot with G_n equal to 1 on the interval and zero outside, but this is not quite satisfactory since the plot uses $\widehat{\theta}$, the ML estimate calculated from the full time interval $[0,\tau]$. The natural remedy is to use test curves with parameter estimates $\widetilde{\theta}$ that only use $[a,b]$-information.

Consider in general terms the estimator $\widetilde{\theta}$ that solves $\int_0^\tau w(s)\{\log h(s,\theta)\,\mathrm{d}N(s) - Y(s)h(s,\theta)\,\mathrm{d}s\} = 0$. This is an M-estimator, or a maximum weighted likelihood estimator, and its large-sample properties are known, see Hjort (1985, 1992) and ABGK (chapter VI). Some work, involving the combination of large-sample arguments of Hjort (1990) with such of Hjort (1992), leads to the following generalisation of the basic result of section 2.1:

$$\begin{aligned} D_n(t) &= \sqrt{n}\int_0^t K_n(s)\{\mathrm{d}\widehat{H}(s) - h(s,\widetilde{\theta})\,\mathrm{d}s\} \\ &\to_d D(t) = \int_0^t \{k(s)/y(s)\}\,\mathrm{d}V(s) - \Big(\int_0^t k(s)\psi(s,\theta)h(s,\theta)\,\mathrm{d}s\Big)' J_w^{-1} \int_0^\tau w(s)\psi(s,\theta)\,\mathrm{d}V(s), \end{aligned}$$

and this zero-mean Gaußian limit process has variance function

$$\kappa(t)^2 = \int_0^t (k^2/y)h\,\mathrm{d}s - 2\Big(\int_0^t k\psi h\,\mathrm{d}s\Big)' J_w^{-1}\Big(\int_0^t kw\psi h\,\mathrm{d}s\Big) + \Big(\int_0^t k\psi h\,\mathrm{d}s\Big)' J_w^{-1} K_w J_w^{-1}\Big(\int_0^t k\psi h\,\mathrm{d}s\Big).$$

Here $J_w = \int_0^\tau wy\psi\psi' h\,\mathrm{d}s$ and $K_w = \int_0^\tau w^2 y\psi\psi' h\,\mathrm{d}s$. The results previously used in this paper correspond to the special case $w(s) = 1$ on the whole interval.

Let now $w = 1$ on $[a,b]$ and zero outside, affecting the parameter estimate, and let also the test weight function K_n be zero outside the interval, affecting the test curve. Then the natural NLH curve is

$$\mathrm{NLH}(t) = \sqrt{n}\int_a^t K_n(s)\{\mathrm{d}\widehat{H}(s) - h(s,\widetilde{\theta})\,\mathrm{d}s\}/\widetilde{\kappa}(t) \quad \text{on } [a,b],$$

where the denominator is the square root of a suitable estimator of

$$\int_a^t (k^2/y)h\,\mathrm{d}s - \Big(\int_a^t k\psi h\,\mathrm{d}s\Big)'\Sigma_{[a,b]}^{-1}\Big(\int_a^t k\psi h\,\mathrm{d}s\Big),$$

and where $\Sigma_{[a,b]} = \int_a^b y\psi\psi' h\,\mathrm{d}s$. This gives A, B and C plots for the validity of the model on $[a,b]$. An A plot for constant hazard rate on $[a,b]$ is for example $\sqrt{n}\{\widehat{H}(t) - \widehat{H}(a) - \widetilde{\theta}(t-a)\}/\widetilde{\kappa}(t)$, where $\widetilde{\theta} = N[a,b]/\int_a^b Y(s)\,\mathrm{d}s$ and $\widetilde{\kappa}(t)^2 = \int_a^b\{\widetilde{\theta}/\widehat{y}(s)\}\,\mathrm{d}s - \widetilde{\theta}(t-a)^2/\int_a^t \widehat{y}(s)\,\mathrm{d}s$. See also Hjort (1993a) for use of such interval tests to dynamic likelihood hazard rate estimation.

The results reported on here are also relevant to the question of making plots with robustly estimated parameter values.

8D. THE FIRST FEW AND THE LAST FEW VALUES. The basic property of the plots is that $\mathrm{NLH}(t)$ is approximately distributed as a standard normal if the model in question is correct. This is really a large-sample statement for each fixed t, as n grows, and we cannot necessarily trust the ± 1.96 limits to correspond precisely to 95% coverage probability for the smallest values of t. Consider the (3.2) and (3.3) plots for the exponential model, for example. Assume that there is no censoring, and let the observed life times be ordered as $t_1 < t_2 < \cdots$. A well known transformation is to new random variables nt_1, $(n-1)(t_2 - t_1)$, $(n-2)(t_3 - t_2)$ and so on; these are independent and exponentially distributed with parameter θ. Using this fact one can prove, in the framework with fixed t_k and increasing n, that

$$\text{both } \mathrm{NLH}_A(t_k) \text{ and } \mathrm{NLH}_B(t_k) \to_d \sqrt{k}(1-\bar{V}_k)/\bar{V}_k^{1/2},$$

where $\bar{V}_k$ is the average of independent unit exponentials $V_1,\ldots,V_k$. Using $\bar{V}_k = \chi^2_{2k}/2k$ one can therefore compute approximate probabilities for exceeding 1.96 in absolute value at t_k, valid at least for k small and n large. The first few probabilities are 0.165, 0.111, 0.092, 0.081, 0.075, and there is convergence to 0.05. This shows that somewhat higher values than under the normal can be expected for the first five or so t_ks, and that not too much emphasis should be placed on the behaviour of the test curve here. (If the nonparametric $\kappa(.)$ estimators are used instead of the parametric ones, in (3.2) and (3.3), then the limit here is $\sqrt{k}(1-\bar{V}_k)$ instead.)

The last few values have a similar behaviour, at least for the Type B plots. With somewhat more work than for the first few one can prove, for the uncensored exponential case, that $\mathrm{NLH}_B(t_{n-l}) \to_d \sqrt{l}(\bar{W}_l - 1)/\bar{W}_l^{1/2}$, where $\bar{W}_l$ is the average of i.i.d. unit exponentials $W_1,\ldots,W_l$.

Some remedies could be thought of in connection with these calculations, including a slight down-weighing at the start and end of the plot and also skewness-reducing transformations, but this is not pursued here.

8E. POSITIVITY OF VARIANCE ESTIMATES. Some natural-looking estimates of the $\kappa^2(.)$ function of (2.3) will turn out negative in places, even if they are perfect from a large-sample point of view. The estimators used in this paper are however safe from such embarrassments. To see this, note first that the following holds, for any measure $\nu(\mathrm{d}s)$, provided only that $\int_0^t vv'\,\mathrm{d}\nu$ has an inverse:

$$\int_0^t u^2\,\mathrm{d}\nu \ge \Big(\int_0^t uv\,\mathrm{d}\nu\Big)'\Big(\int_0^t vv'\,\mathrm{d}\nu\Big)^{-1}\Big(\int_0^t uv\,\mathrm{d}\nu\Big),$$

for any functions $u\colon [0,s] \to I\!R$ and $v\colon [0,1] \to I\!R^p$. This is actually a generalisation of the Cauchy–Schwartz inequality, and is true since the matrix

$$\Gamma = \begin{pmatrix} \int_0^t u^2 \, d\nu & \int_0^t uv' \, d\nu \\ \int_0^t uv \, d\nu & \int_0^t vv' \, d\nu \end{pmatrix}$$

is non-negative definite, and consequently $\Gamma_{1,1} - \Gamma_{1,2}\Gamma_{2,2}^{-1}\Gamma_{2,1}$, in usual block notation, is non-negative definite too. And this implies

$$\int_0^t (k^2/y) \, d\nu \geq \Big(\int_0^t k\psi \, d\nu\Big)' \Big(\int_0^\tau y\psi\psi' \, d\nu\Big)^{-1} \Big(\int_0^t k\psi \, d\nu\Big), \quad \text{for all } t \leq \tau,$$

and again for any measure $\nu(ds)$. In our setting $\nu(ds)$ would be an estimate of $h(s,\theta)\,ds$ and $y(s)$ would be replaced with $Y(s)/n$.

8F. OTHER GRAPHICAL TEST PLOTS. One of the independent origins of the Nelson–Aalen estimator is the 1972 paper by Nelson, advocating the sensible idea of plotting both the nonparametric $\widehat{H}(t)$ and a parametric $H(t,\widehat{\theta})$ (and sometimes with ad hoc estimates for the parameters) in the same diagram. The difficulty of judging such pairs of estimated cumulative hazards is that the variability differs both between the two curves and over time. One may view our Type A plots as more worked out and sophisticated versions of the same idea, incorporating the correct local stabilising scaling of the difference.

Another sensible idea is to plot a nonparametric estimate $\widehat{h}(s)$ of the hazard rate itself with the parametric $h(s,\widehat{\theta})$. Versions of $\widehat{h}(.)$ are discussed in ABGK (chapter IV) and in Hjort (1991, 1993a). Some of the latter ones are inspired by the idea of making nonparametric corrections to parametric estimates, and will directly or indirectly give indications of the fit of the parametric model. Again a direct comparison can be difficult since the precision of the curves are quite different. Still other graphical model plots are discussed in ABGK (chapter VI). They include in particular a non-normalised version of our Type B plot, and plotted against $N(t)$ instead of direct time t.

8G. THE SEMIPARAMETRIC COX MODEL. We gave methods for checking the parametric Cox model in section 4. The possibly overused semiparametric version postulates only $h_j(s) = h_0(s)\exp(\beta' z_j)$ with no structure imposed on the baseline hazard $h_0(.)$. With notation as in section 4 we have $dN_j(s) = Y_j(s)\exp(\beta' z_j)h_0(s)\,ds$ + noise. The cumulative baseline hazard $H_0(.)$ is usually estimated using

$$\widehat{H}_0(t) = \int_0^t \frac{d\sum_{j=1}^n N_j(s)}{\sum_{j=1}^n Y_j(s)\exp(\widetilde{\beta}' z_j)\,ds},$$

where $\widetilde{\beta}$ is the partial likelihood Cox estimator. Now consider the test function

$$D_n(t) = n^{-1/2} \sum_{j=1}^n \int_0^t k_j(s)\{dN_j(s) - Y_j(s)\exp(\widetilde{\beta}' z_j) d\widehat{H}_0(s)\},$$

for suitable weight functions $k_j(.)$. The choice $k_j(s) = 1$ gives simply zero, but other choices like $k_j(s) = \exp(-\widetilde{\beta}' z_j)$ will give curves of interest and that if the Cox model is correct should lie around zero. We can work out the necessary asymptotic theory to find the limiting variance $\kappa(t)^2$ of $D_n(t)$, after which we would have a NLH curve $D_n(t)/\widehat{\kappa}(t)$ to test the validity of the Cox model, but this is not pursued here.

An increasingly popular alternative to the multiplicative Cox model is Aalen's linear hazard nonparametric regression model, see Aalen (1989) and ABGK (chapter VII). Hjort (1993b) discusses also parametric versions of this model, and the validity of such can again be assessed using appropriate test curve constructions of the NLH variety. We have worked out the necessary theory but this will be presented elsewhere.

9. Examples and illustrations.

EXAMPLE 9.1. TESTING A CONSTANT HAZARD. Figure 9.1a displays NLH plots of Type A and Type B for 100 simulated unit exponentials (i.e. having constant hazard rate 1). Figure 9.1b presents A and B curves in a situation with 100 simulated unit exponentials but censored with another 100 simulated unit exponentials. The plots behave as expected, cf. also Remark 8A.

FIGURE 9.1. *NLH plots of Type A (line) and Type B (dashed) are shown in 9.1a for 100 simulated unit exponentials. 9.1b shows A and B curves for 100 unit exponentials in a situation with about 50% censoring (the censoring variables are another set of 100 unit exponentials).*

EXAMPLE 9.2. TESTING A WEIBULL. 100 variables simulated from the Weibull (10,1.3) distribution, i.e. with hazard rate $13s^{0.3}$, gave rise to the plots of 9.2a and 9.2b. Type A plots for the exponential and the Weibull are shown in 9.2a and Type B plots in 9.2b. The curves for the exponential model clearly wanders away from the acceptable band while the Weibull curves stay within. Note also that the plots indicate a Weibull shape parameter greater than 1, in view of remarks made in section 5. A Weibull (10,0.7) distribution would for example give exponential model plots that first increased and then decreased.

FIGURE 9.2. *NLH plots of Type A are shown for the exponential model (line) and Weibull model (dashed) are shown in 9.2a for 100 simulated Weibull (10,1.3) variables. NLH plots of Type B for the same data are shown in 9.2b.*

EXAMPLE 9.3. IUD EXPULSION AND REMOVAL DATA. These data are from Peterson (1975) and are the experience of a sample of 100 women using an experimental intrauterine contraceptive device (IUD). There are several competing risks in this experiment, and following Aalen's 1982 analysis we have focused on two: unplanned removal and expulsion of the IUD. Most of the removals were planned, leading to heavy censoring. A plot of the Nelson–Aalen estimator of the two cumulative hazards are given in figure 9.3a. The cumulative hazard for unplanned removal appears to be linear, and that for expulsion appears to increase logarithmically. This suggests an exponential model for removals and a simple frailty model for expulsions. The NLH plots for unplanned removal are given in figure 9.3b and show good agreement with the model. Figure 9.3c, comparing the IUD expulsions with the simple frailty model, shows close agreement. The dramatic step in the model at a time of one year is due to the planned removal of IUDs (censoring) in a large proportion of the remaining women at this time. Our conclusions agree with more informal analysis by Aalen (1982).

FIGURE 9.3. *Nelson–Aalen estimators for the cumulative hazard rates for expulsion (dotted line) and unplanned removal (line) are shown in 9.3a. NLH plots for constant hazard rate are given in 9.3b for unplanned removal and for the simple frailty model of 3.5 in 9.3c for expulsion. As in figures 9.1 and 9.2 the A plots are with a line and the B plots are dashed. In both cases there is close agreement with the model.*

EXAMPLE 9.4. GOMPERTZ MORTALITY RATES FOR FYN DIABETICS. ABGK (example I.3.2 and later on) discuss several aspects of data that have been obtained by Green et al. (1981) on the mortality for 716 women and 783 men suffering from insulin-dependent diabetes mellitus in the Danish county of Fyn. We have got the full data set from Andersen and Borgan (personal communication). The life-times are partly right-censored since the study was finished 1 January 1982, and also partly left-truncated, since only persons alive by 1 July 1973 could enter the study. Although the traditional framework of sections 2 and 3 is too narrow our methods still work, as explained in section 6, with $Y(s)$ of the form $\sum_{j=1}^{n} I\{t_j \geq s > v_j\}$, where v_j denotes entry age and t_j is exit age (age of death if death occurred before 01.01.1982 and age at this date otherwise).

ABGK consider the women group of this example in some depth in their chapter VI, and we shall complement their analysis by studying the fit of the time-continuous Gompertz model $h(s) = \theta \exp(\beta s)$ for the mortality rate. The data are only given in whole years, and there are ties, so we shall in fact use the time-discrete machinery outlined in section 7. There are Y_i women alive and diagnosed with diabetes when entering year interval $[i, i+1)$, and ΔN_i of these die during this year, where $i = 1, 2, \ldots, 98$. This framework is slightly more precise, presumably, given the grouped data, although the time-continuous apparatus also is acceptable here as an approximation, since the time intervals are relatively short (ABGK use time-continuous techniques). The hazard for the ith year interval is

$$h_i(\theta, \beta) = F[i, i+1)/F[i, \infty) = 1 - \exp\{-H[i, i+1)\} = 1 - \exp\big[-(\theta/\beta)(e^{(i+1)\beta} - e^{i\beta})\big]. \quad (9.1)$$

ML estimates $0.957/10^3$ ($0.317/10^3$) and $61.110/10^3$ ($4.668/10^3$) for θ and β were found by a Newton–Raphson algorithm (estimated standard deviations in parentheses). ABGK use an alternative parametrisation bc^{s-50} and finds (p. 413) 0.0199 and 1.066 for b and c using a likelihood appropriate to time-continuous data; our values give instead 0.0203 and 1.063. Using our ML estimates we can display nonparametric versus parametric estimates of cumulative hazards, and thereby produce a close relative of figure IV.3.5 in ABGK, but perhaps with even better fit for age ≥ 80 years.

FIGURE 9.4. *Time-discrete NLH plots of Type A (connected dots) and Type B (connected plusses) for the Gompertz model are shown for the group of 716 women in 9.4a and for the 783 men in 9.4b. In both cases there is close agreement with the model. NLH plots to compare the mortality rate for men with diabetes with the estimated Gompertz mortality rate for women with diabetes are given in 9.4c, and shows that the mortality rate for men is higher.*

The A and B plots to assess the validity of the Gompertz model, or strictly speaking rather the inherited model (9.1) for yearly hazards, are shown in figure 9.4a, and are a convincing show of support for the model. We note that ABGK were able to detect a certain unexplained departure from the Gompertz model using a Khmaladze test (see their example VI.3.9). This difference in opinion might partly stem from their use of time-continuous machinery and perhaps partly from their use of a τ value at the very end of the time scale (note their discussion of this "very delicate matter" on p. 466). To indicate the time-discreteness the A plot is shown with yearly dots. A similar analysis was carried out for the men group (not similarly analysed in ABGK). ML estimates (again using the time-discrete likelihood) are $1.097/10^3$ ($0.314/10^3$) and $64.809/10^3$ ($4.390/10^3$) for θ and β. This translates to 0.028 and 1.067 for b and c. The A and B plots for this group are given in 9.4b. We conclude that both men and women of Fyn look perfectly Gompertzian to us. To show that men and women are different we have also included a Type A plot to assess the hypothesis that the men have the hazard rates $h_i = h_{i,0}$ as specified by (9.1) with parameter estimates from

the women stratum. In this fully specified case the variance expression (7.4) simplifies by losing its second term, cf. 3.1.

EXAMPLE 9.5. PARAMETRIC COX REGRESSION FOR DANISH MELANOMA SURVIVAL DATA. This data set, though with fewer variables, was analysed extensively in ABGK. The full set has been given us by Andersen and Borgan (personal communication). They relate to 205 patients who had a particular operation to remove malignant melanoma (a form of skin cancer) in Odense, Denmark in the period 1962–77. Various risk factors were recorded at the time of the operation. The patients were observed until 1977, and the time (in days) until they died of melanoma was recorded. Patients dead of other causes or still alive in 1977 are treated as censored observations.

One of the analyses performed by ABGK was a semiparametric Cox regression. The estimated baseline cumulative hazard appeared fairly nearly linear, leading us to select these data for a parametric analysis with constant baseline hazard. We used the 'Poisson error trick' of Aitkin and Clayton (1980) to fit the model as a generalised linear model with Poisson error and log link. ABGK used thickness of the tumour (in mm) and sex of the patient as covariates, and our data set has a further three: the layer of skin to which the tumour penetrated (coded 1 to 4), presence of ulceration on the tumour, and presence of a certain type of cell (epithelioide cells) in the tumour. Men had a higher hazard than women, and the actions of the other variables were all in the directions which would be expected. In the parlance of generalised linear models the deviance of the resulting model was 187.1 on 198 degrees of freedom (null deviance was 232.1 on 204 df), giving an acceptable fit. A measure of the effect of a variable z is the range of values of βz, which measures how much deviation from uniform hazard is attributable to this variable. The variables with the highest values of $\mathrm{Var}(\beta_i z_i)$, and so associated with the greatest variation in hazard, were 'presence of ulceration' and 'skin layer penetrated'.

The NLH plots of this exponential regression model show the expected relationship between the exponential and regression model plots. In the type B plot, figure 9.5b, the curve for the exponential model lies above that for the regression model, and in the Type A plot, figure 9.5a, the curves cross as expected, cf. remarks at the end of section 5. The plots do not, however, indicate a better fit from the regression model. This calls into question the assumption of constant baseline hazard over the whole time period. Looking at the data we find that are no deaths from melanoma after about 9 years, and fewer than would be expected in the first two years. The sparse nature of the data after 9 years means that the estimated variance is very high and so the NLH-plots only show the initial dip in hazard. The nonparametric variance estimates are not as sensitive to the spacing of points and should work better at the end of the time interval, and in fact figure 9.5c shows that they allow the exponential model to be rejected by a Type A plot.

Figure 9.5d displays an estimate of the baseline hazard for a semiparametric Cox model fitted to these data. The hazard is estimated by kernel smoothing $\int b^{-1}K(h^{-1}(s-t))\,\mathrm{d}\widehat{H}_0(t)$ of the estimated cumulative hazard given in 8G. The kernel function chosen is $K(z) = \frac{15}{8}(1-8z^2+16z^4)$ on $[-\frac{1}{2},\frac{1}{2}]$, since it is the simplest one obeying the natural requirements of being a symmetric unimodal probability density with existing derivatives zero at $\pm\frac{1}{2}$ (the hazard estimate will not have a continuous derivative without the last requirement, which is why we avoided the Yepanechnikov kernel). We have worked out some theory for cross validation and for adaptive smoothing and also experimented with orthogonal expansion estimators. A uniform bandwidth value of 4 years was found to be satisfactory. An estimate for times near zero was achieved by augmenting the observed points $t_1, t_2, \ldots$ with 'reflected' points $-t_1, -t_2, \ldots$, following a boundary technique for density estimation, see for example Scott (1992, section 6.2.3.5). This allows estimation of the baseline hazard down to $s = 0$ but constrains it to have zero derivative at $s = 0$.

The plots we have given show that the constant baseline hazard suggested by the cumulative hazard plot cannot be justified under closer examination. We also mention that ABGK (example VII.3.1) as well as Murphy (1993) have pointed to some problems with the proportional hazards assumptions for these data. It should be noted, however, that a semiparametric Cox model that was fitted in order to produce figure 9.5d gave almost identical coefficient estimates and overall significance level, suggesting that the conclusions in this case are not very sensitive to assumptions made in the analysis.

FIGURE 9.5. *NLH-plots of Type A and B are in respectively 9.5a and 9.5b for the parametric Cox regression model (dashed lines) and the homogeneous exponential model (solid line). In 9.5c an A plot for the homogeneous exponential model is given, suggesting that the true hazard is lower at the beginning and at the end of the time interval. Figure 9.5d gives a kernel smoothed estimate of the baseline hazard in a semiparametric Cox model for the melanoma data.*

EXAMPLE 9.6. TIME-DISCRETE FOUR-PARAMETER HETEROGENEITY MODEL FOR THE TIME TO NEXT BIRTH. Data have been extracted from the Norwegian Medical Birth Registry by O. Aalen and B. Sandstad to find the time to next birth for young women experiencing a stillbirth. The data set consists of all the 451 Norwegian women who had their first birth during 1967 to 1971, who were at the time of this birth below 25 years of age and married, and for whom the child was stillborn. We have got the data from Aalen (personal communication). Aalen (1992) discusses this data set and fits a four-parameter frailty model: woman j is thought of as having 'hazard' or intensity rate $Z_j\lambda(t)$, where $\lambda(t)$ is a Weibull starting after nine months, while the distribution of the Z_js among the women is thought to have a certain compound Poisson distribution. The result is as in (3.12), with

$$\lambda(t) = a(t-9/12)^k \quad \text{and} \quad \Lambda(t) = a(t-9/12)^{k+1}/(k+1), \quad \text{for } t \geq 9/12. \tag{9.2}$$

The data are only collected time-discretely, however, so we cannot produce NLH plots to test (9.2) directly. There are 24 time intervals of variable lengths, starting with $(9/12, 10/12)$, $(10/12, 11/12)$ and ending with $(12, 13)$, $(13, 15)$ (years). Letting the ith time interval be $(t_{i,l}, t_{i,r})$ the number Y_i of women having not yet given birth after the stillbirth when entering this time interval is known, as is the number ΔN_i of these that then give birth within this interval. The hazard rate for the ith interval is

$$h_i = h_i(a,k,\alpha,\delta) = 1 - \exp[-\{H(t_{i,r},a,k,\alpha,\delta) - H(t_{i,l},a,k,\alpha,\delta)\}], \tag{9.3}$$

with the $H(.)$ function as in (3.12), and the log-likelihood is $\sum_{i=1}^{24}\{\Delta N_i \log h_i + (Y_i - \Delta N_i)\log(1-h_i)\}$. We wrote a Newton–Raphson programme in S-Plus to find ML estimates 5.141 (1.600) for a, 1.152 (0.193) for k, 1.305 (0.114) for α, and 1.551 (0.315) for δ. This required lengthy partial derivatives calculations. The numbers in parentheses are the estimated standard errors, obtained from Σ^{-1}/n and (7.2). The a value is incorrectly given in Aalen (1992), but his plots showing a quite good agreement between model and data are correct. (He also estimated the standard errors differently.) In figure 9.6 the time-discrete normalised hazard difference $\sqrt{n}\{\Delta N_i/Y_i - h_i(\widehat{a},\widehat{k},\widehat{\alpha},\widehat{\delta})\}/\widehat{w}_i$ is plotted against the midpoints of the time intervals, see (7.5). The plot indicates a very good fit. Similarly the three-parameter model where $k = 1$, corresponding to a linearly increasing individual intensity rate, would also give a quite accurate fit.

Of course these three- and four-parameter classes are quite rich and one should not overstate this particular probabilistical explanation of *N*ature. It is nevertheless an attractive model with a

natural socio-biological interpretation and an impressive fit. Note that each woman has a steadily increasing intensity rate for giving birth a second time (up to 40 years of age). A feature of the model is also that a certain proportion of the women, namely $\exp(-\frac{\alpha}{\alpha-1}\delta)$ and here estimated to be 6.4%, will never have the second birth.

FIGURE 9.6. *Time-discrete normalised hazard difference plot assessing the compound Poisson heterogeneity model for the time to next birth following stillbirth for 451 young Norwegian women.*

Appendix: S-Plus procedures. A package of S-Plus procedures has been produced to produce the Type A and Type B plots and to perform the necessary parameter estimation for the exponential, Weibull, Gompertz, simple frailty and exponential (Cox) regression models. These procedures were used to produce the plots in this paper, and are available by electronic mail from either of the authors upon receipt of an ethnic postcard.

The package consists of a single function for the user to call that then sorts the data and calls separate procedures to compute the appropriate maximum likelihood parameter estimates and calculate and draw all the requested plots. The parameter estimates and approximate standard deviations are also reported.

These procedures calculate the normalised local hazards only at times when cases are observed, and interpolates linearly between these points. This does not change the asymptotic behaviour of the plots, but saves computation and produces smoother pictures for small samples. The parameter estimates use a variety of methods. There is an explicit formula for the exponential model. The exponential regression model can be rewritten as a generalised linear model using the Poisson error trick explained for example in Aitkin and Clayton (1980), and is then fitted using the `glm` procedure in S-Plus. The other models are fitted by direct numerical maximisation of the log-likelihood. The variance-covariance matrix of the parameter estimates is approximated by Σ^{-1}/n, with the same matrix Σ^{-1} as is used in calculating the plots. The calculations for the plots are carried out entirely within S-Plus, rather than using C or FORTRAN routines. While S-Plus procedures are easier to write, debug and extend than C or FORTRAN, they can be much slower; a plot for 500 cases can take a couple of minutes.

It should be a straightforward exercise to extend the procedures to compute different NLH plots, once the correct formulae are derived. The only significant complications arise from the structure of S-Plus. While there are efficient commands for manipulating matrices, vectors and lists as single objects, loops to manipulate components of these structures can run very slowly. For this reason, the integrals required in the plots are calculated by first working out the increments between successive time points and then taking cumulative sums of these increments. If other systems were used instead of S-Plus, it would probably be easiest to use the simpler finite-sum expressions.

Acknowledgements. We are grateful to Odd Aalen, Ørnulf Borgan and Per Kragh Andersen for giving us copies of various data sets with necessary explanations, for permission to use them, and for their interest. This work was carried out while the first author was visiting Oxford with a grant from the Royal Norwegian Research Council.

References

Aitkin, M. and Clayton, D.G. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics* **29**, 156–163.

Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Arulchelvam, M. (1992). Some topics pertaining to parametric survival data models. Cand. Scient. thesis, Department of Mathematics and Statistics, University of Oslo.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Durbin (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

Green, A., Hauge, M., Holm, N.V. and Rasch, L.L. (1981). Epidemiological studies of diabetes mellitus in Denmark. II. A prevalence study based on insulin prescriptions. *Diabetologia* **20**, 468–470.

Hjort, N.L. (1985). Contribution to the discussion of Andersen and Borgan's 'Counting process models for life history data: a review'. *Scandinavian Journal of Statistics* **12**, 141–150.

Hjort, N.L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Annals of Statistics* **18**, 1221–1258.

Hjort, N.L. (1991). Semiparametric estimation of parametric hazard rates. In *Survival Analysis: State of the Art*, Kluwer, Dordrecht, pp. 211–236. Proceedings of the *NATO Advanced Study Workshop on Survival Analysis and Related Topics*, Columbus, Ohio, eds. P.S. Goel and J.P. Klein.

Hjort, N.L. (1992). On inference in parametric survival data models. *International Statistical Review* **60**, 355–387.

Hjort, N.L. (1993a). Dynamic likelihood hazard rate estimation. Submitted for publication.

Hjort, N.L. (1993b). Efficiency of three estimators in Aalen's linear hazard rate regression model. Submitted for publication.

Hjort, N.L. and Pollard, D.B. (1993). Asymptotics for minimisers of convex processes. Submitted for publication.

Koning, A. (1991). Stochastic integrals and goodness-of-fit tests. Proefschrift ter verkrijging van de graad van doctor, Universiteit Twente.

Miller, R.G. and Siegmund, D. (1982). Maximally selected chi-square statistics. *Biometrics* **38**, 1011–1016.

Murphy, S.A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scandinavian Journal of Statistics* **20**, 35–50.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945–965.

Peterson, A. V. (1975). Nonparametric estimation in the competing risks problem. PhD thesis, Department of Statistics, Stanford University.

Pollard, D.B. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.

Aalen, O.O. (1982). Practical applications of the non-parametric statistical theory for counting processes. Statistical Research Report, Department of Mathematics and Statistics, University of Oslo.

Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8**, 907–925.

Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability* **2**, 951–972.

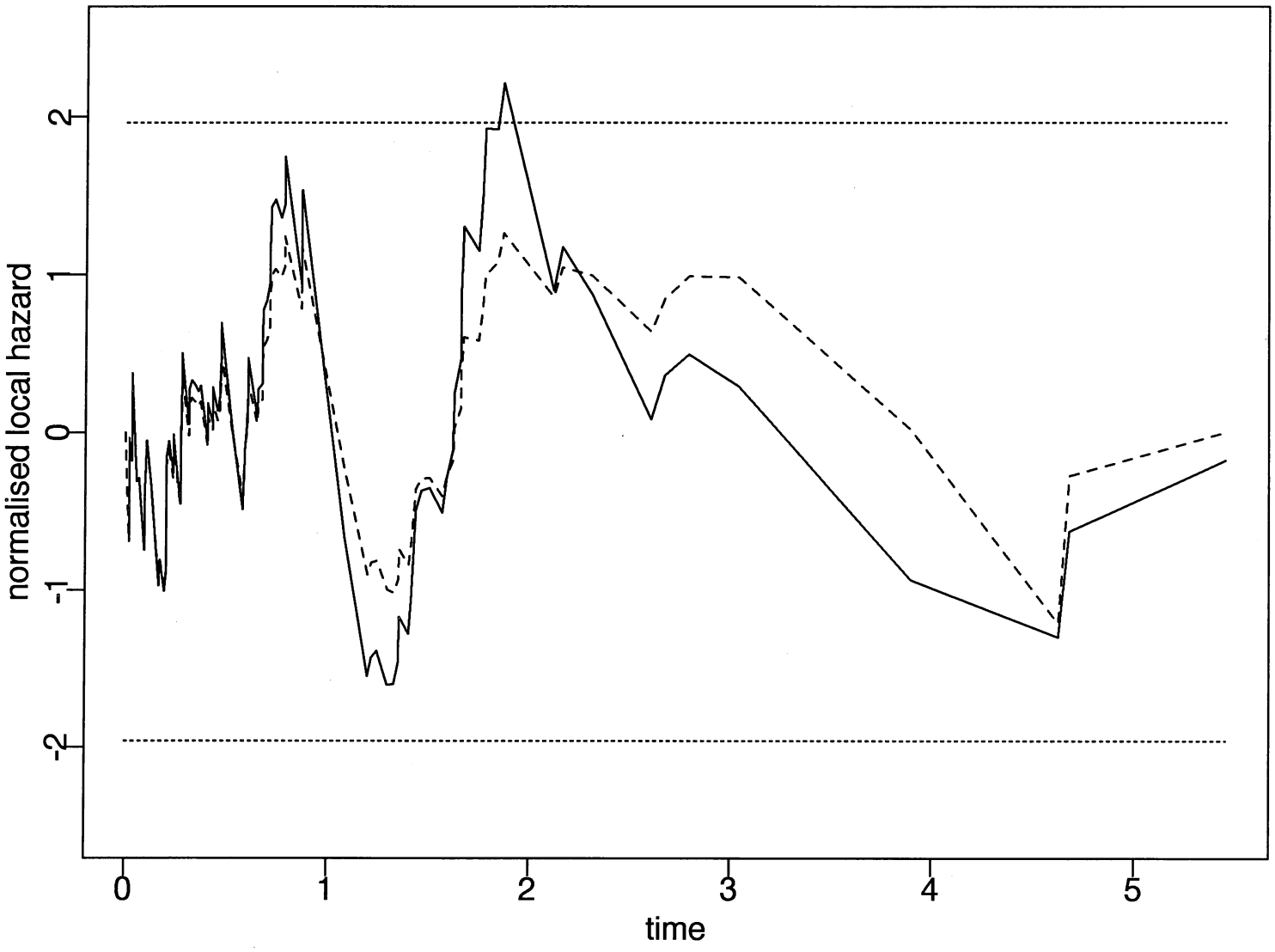

Figure 9.1.a.

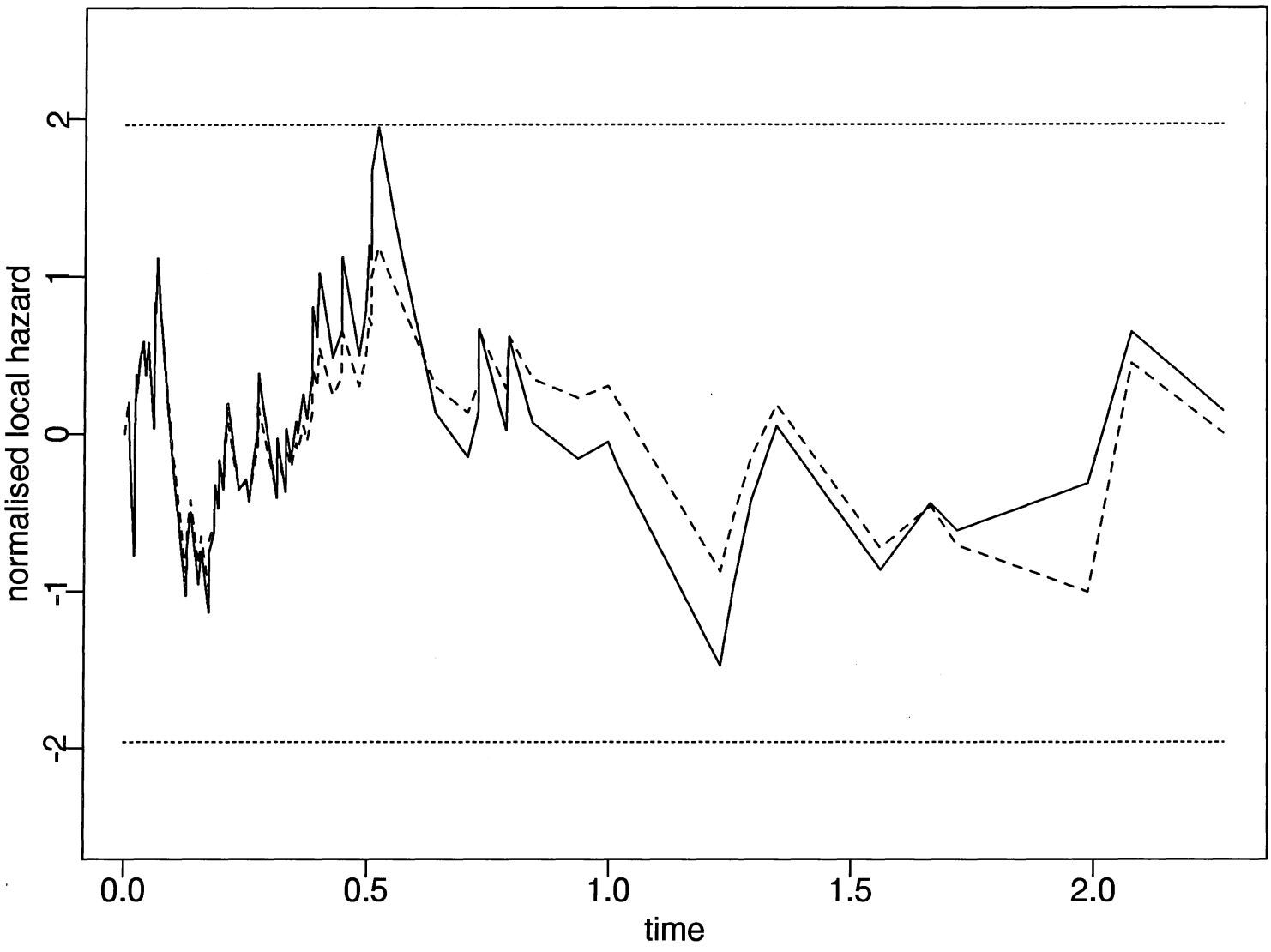

Figure 9.1.b.

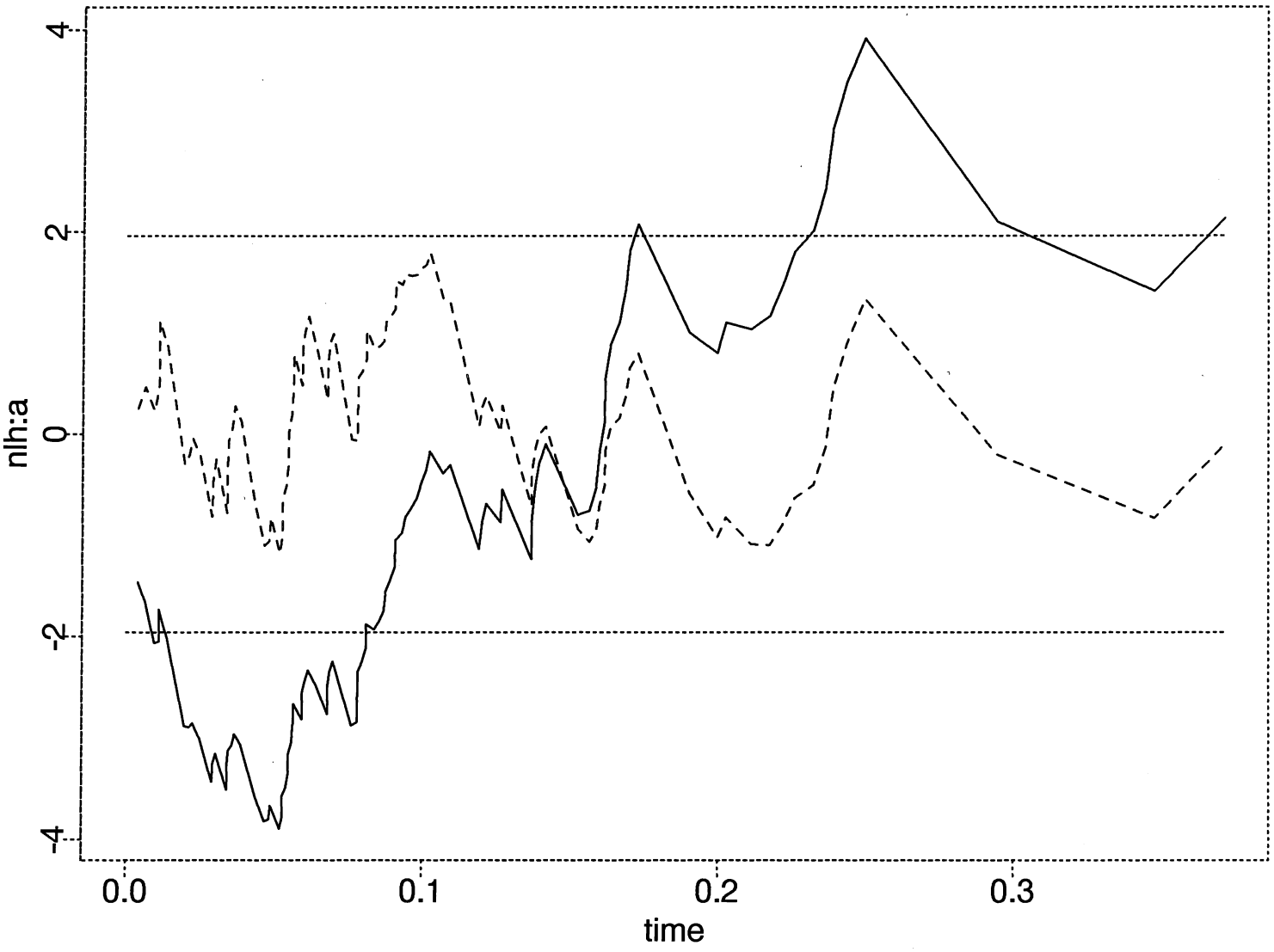

Figure 9.2.a.

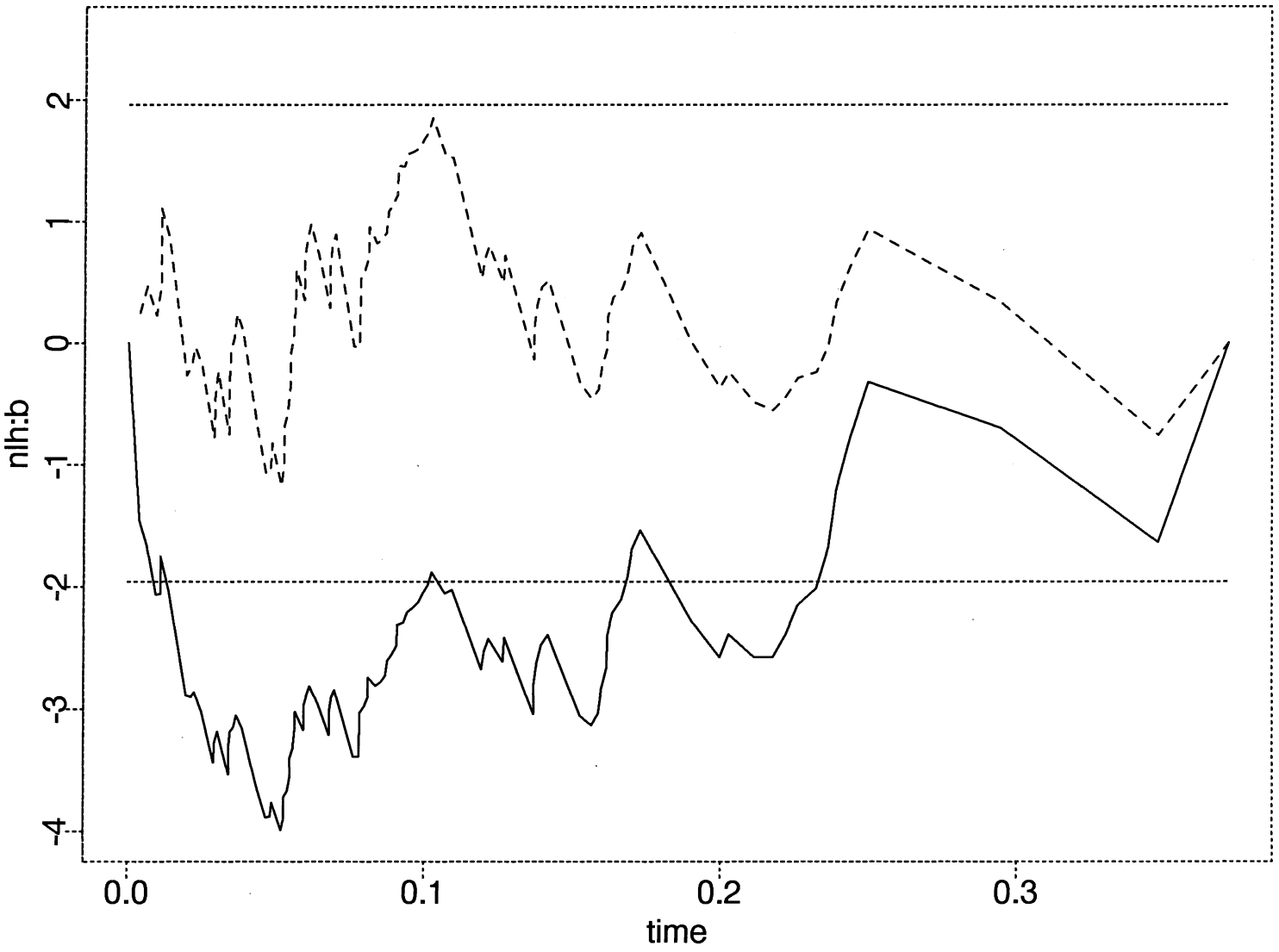

Figure 9.2.b.

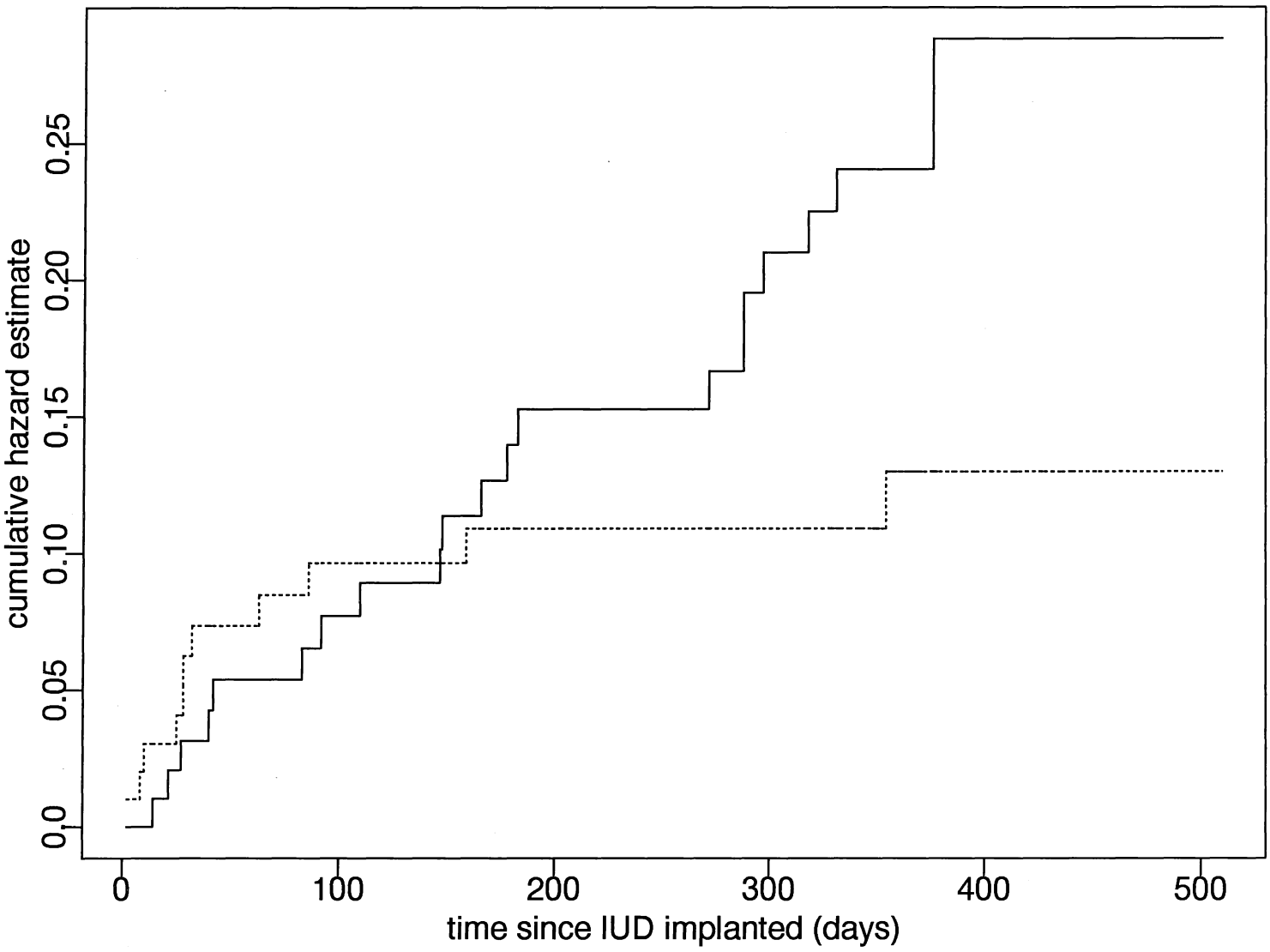
cumulative hazard estimate
0.0
0.05
0.10
0.15
0.20
0.25
0
100
200
300
400
500
time since IUD implanted (days)

Figure 9.3.a.

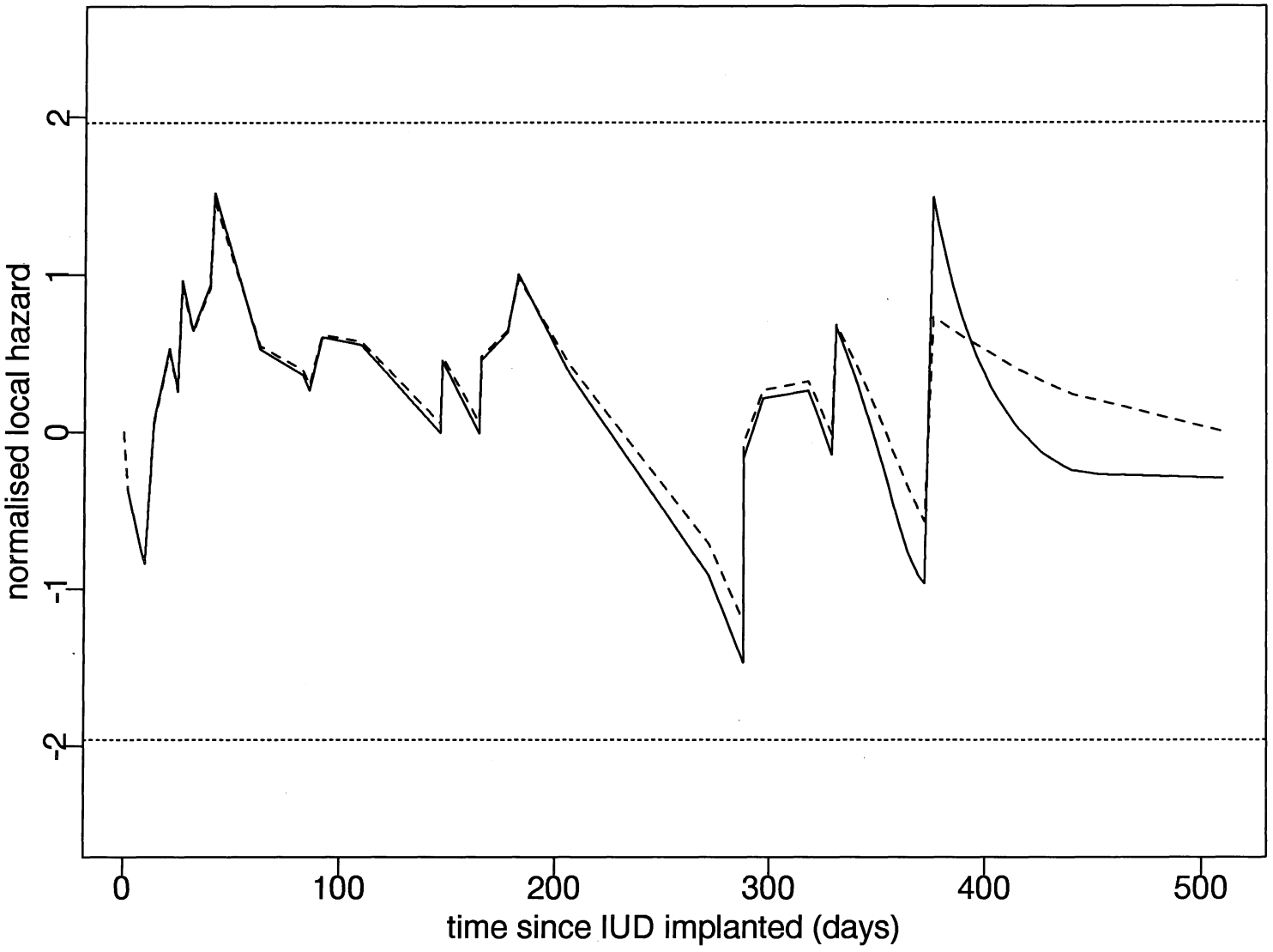
normalised local hazard
2
1
0
-1
-2
0
100
200
300
400
500
time since IUD implanted (days)

Figure 9.3.b.

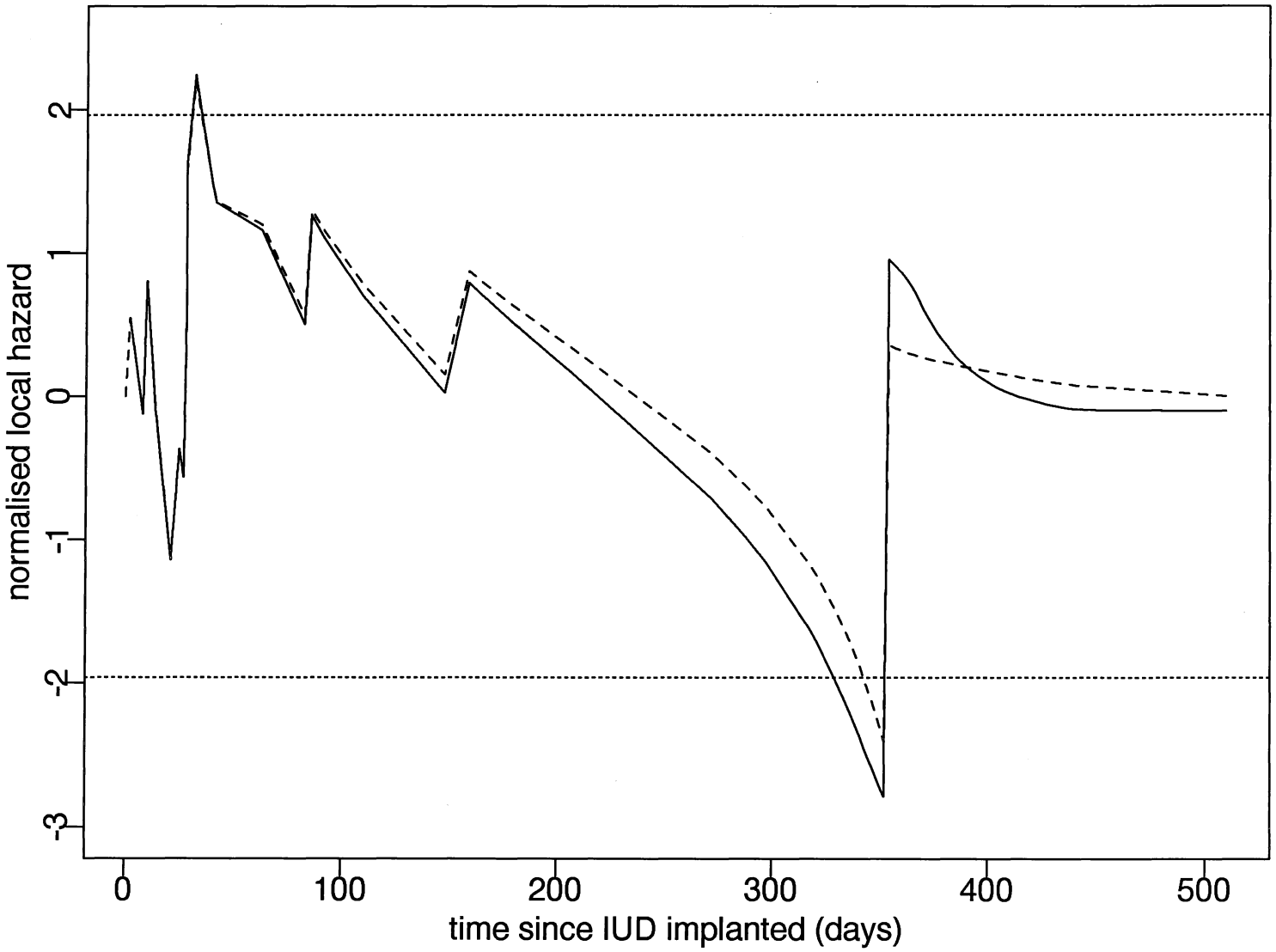
normalised local hazard
2
1
0
-1
-2
-3
0
100
200
300
400
500
time since IUD implanted (days)

Figure 9.3.c.

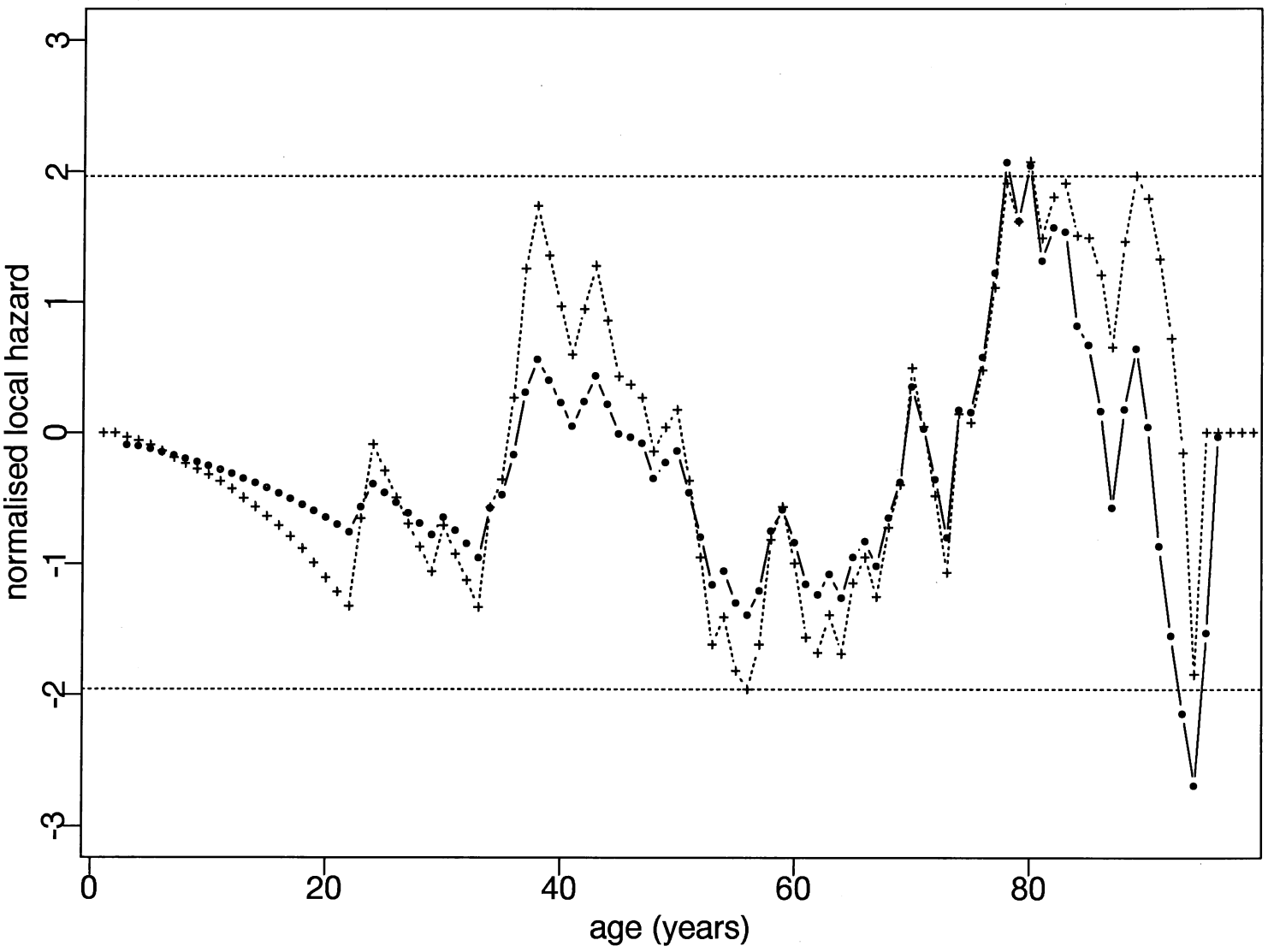

Figure 8.4.a.

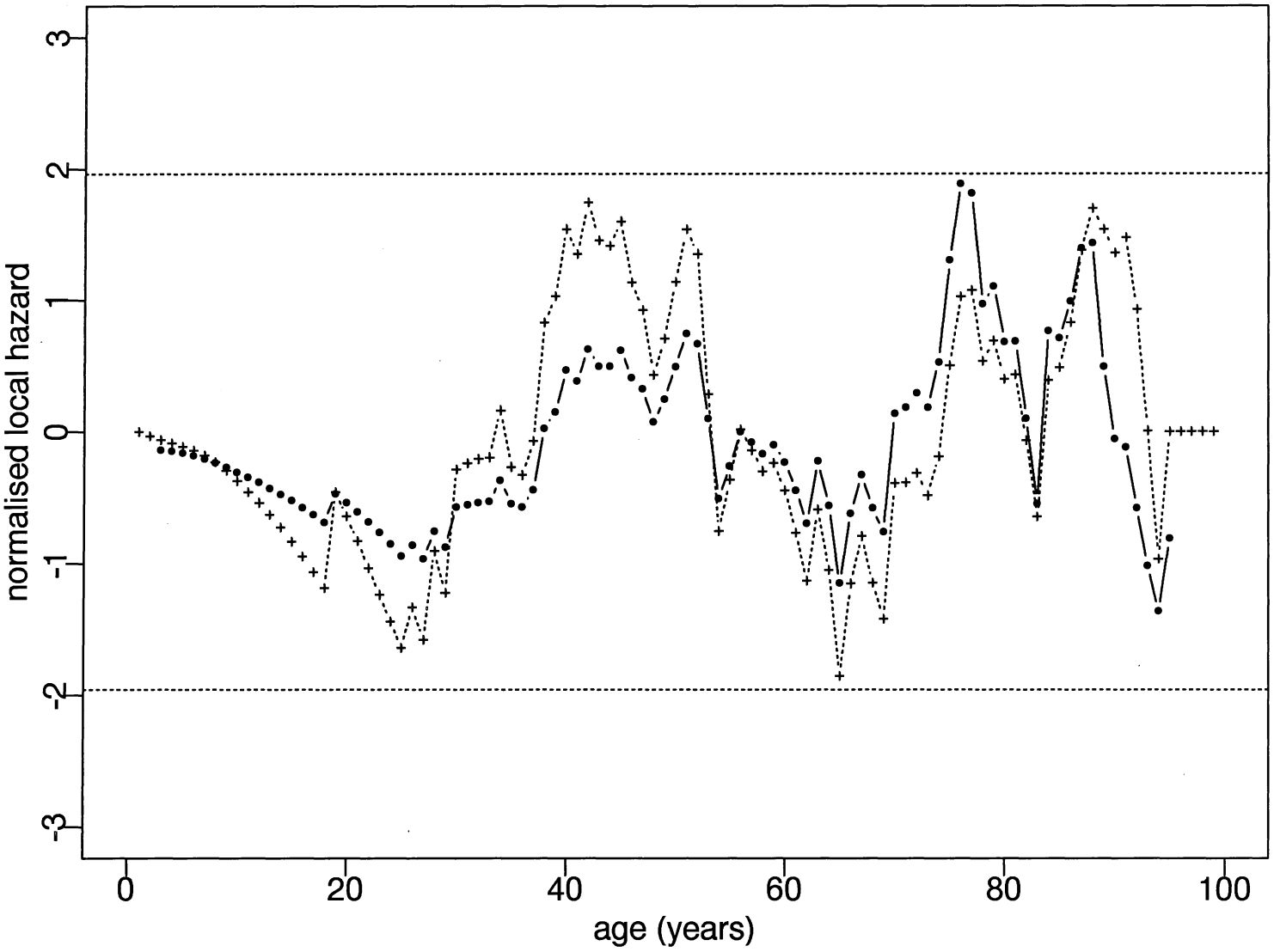

Figure 9.4.b.

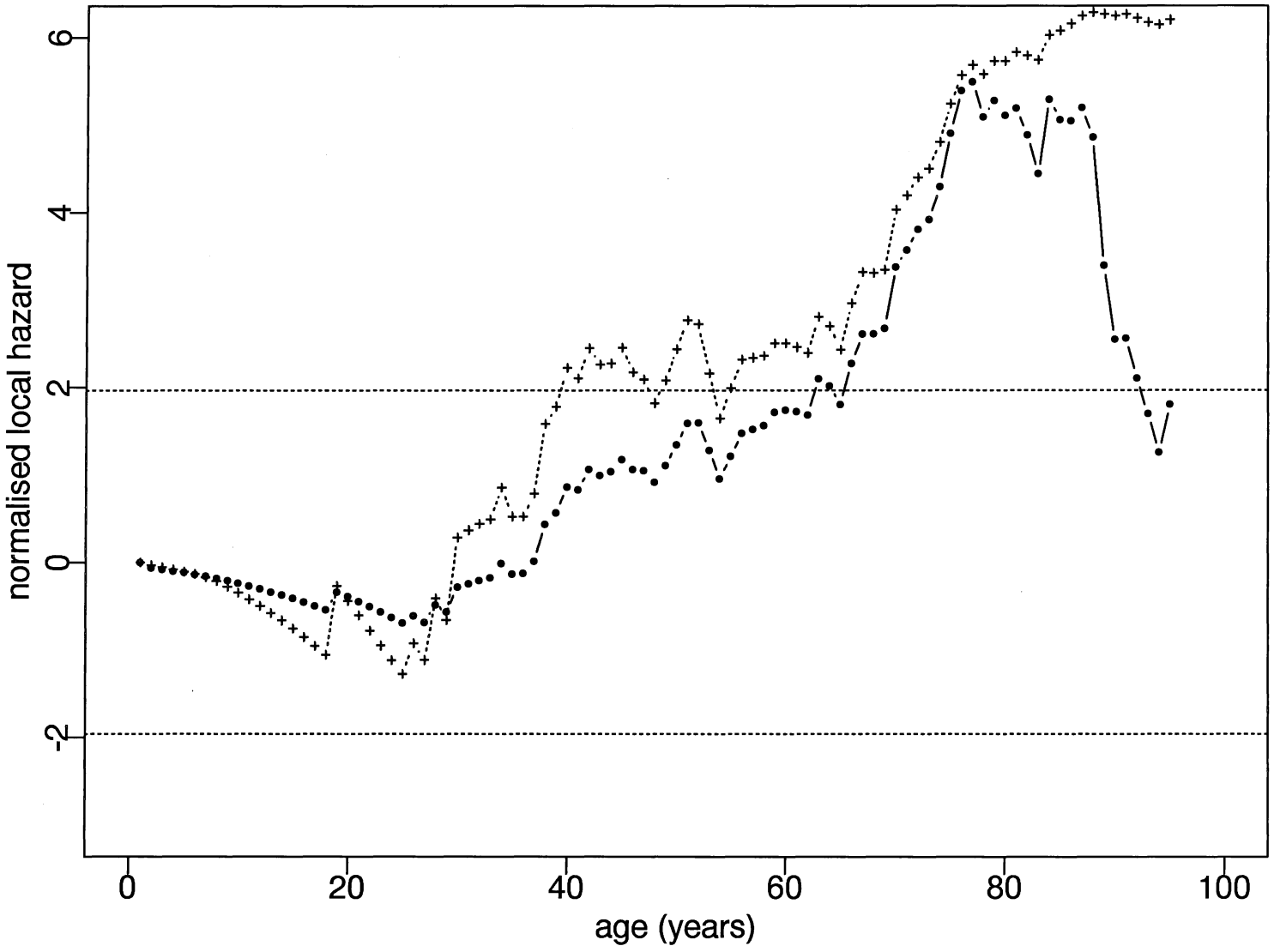

Figure 9.4.c.

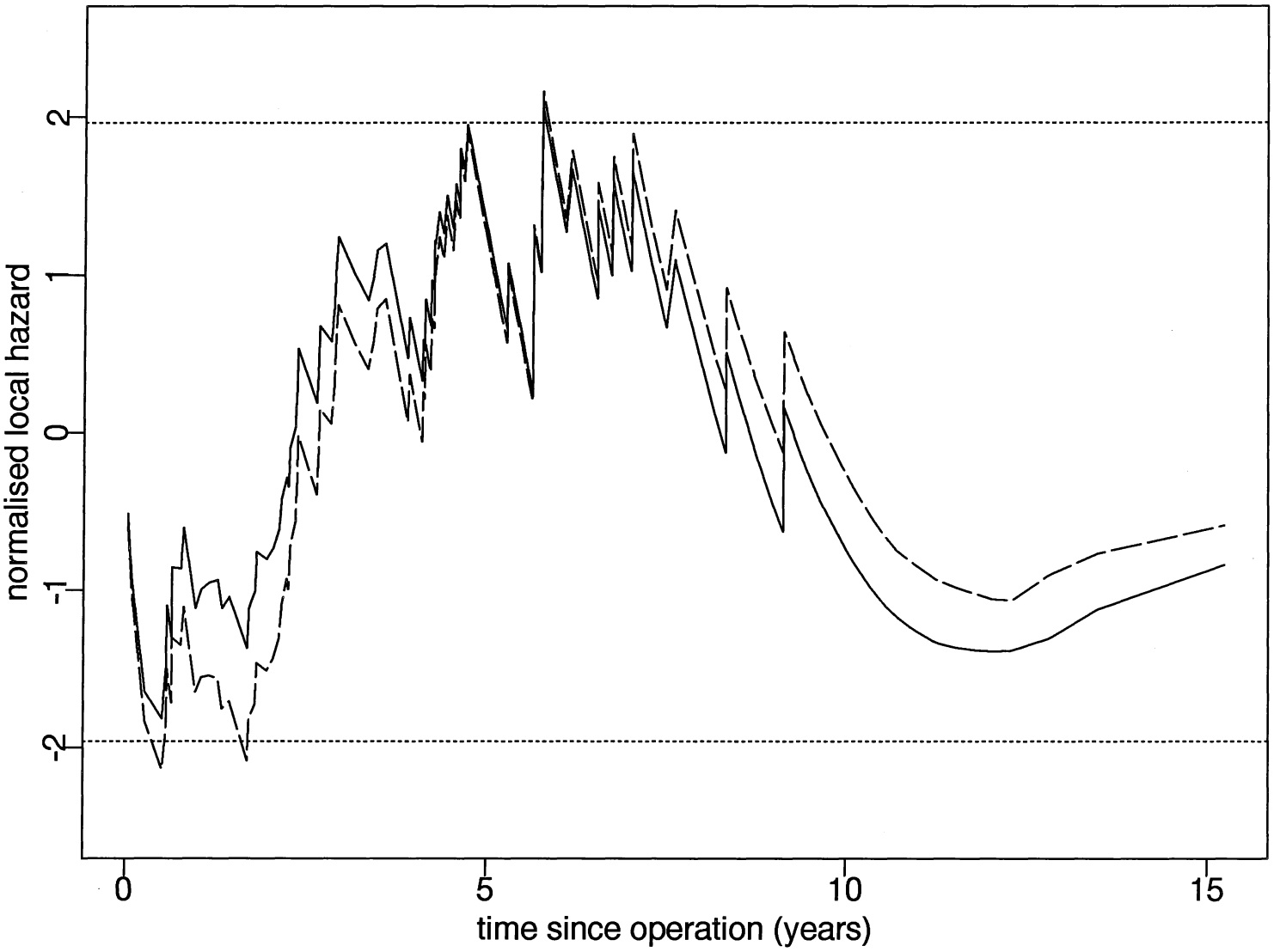
normalised local hazard
2
1
0
-1
-2
0
5
10
15
time since operation (years)

Figure 9.5.a.

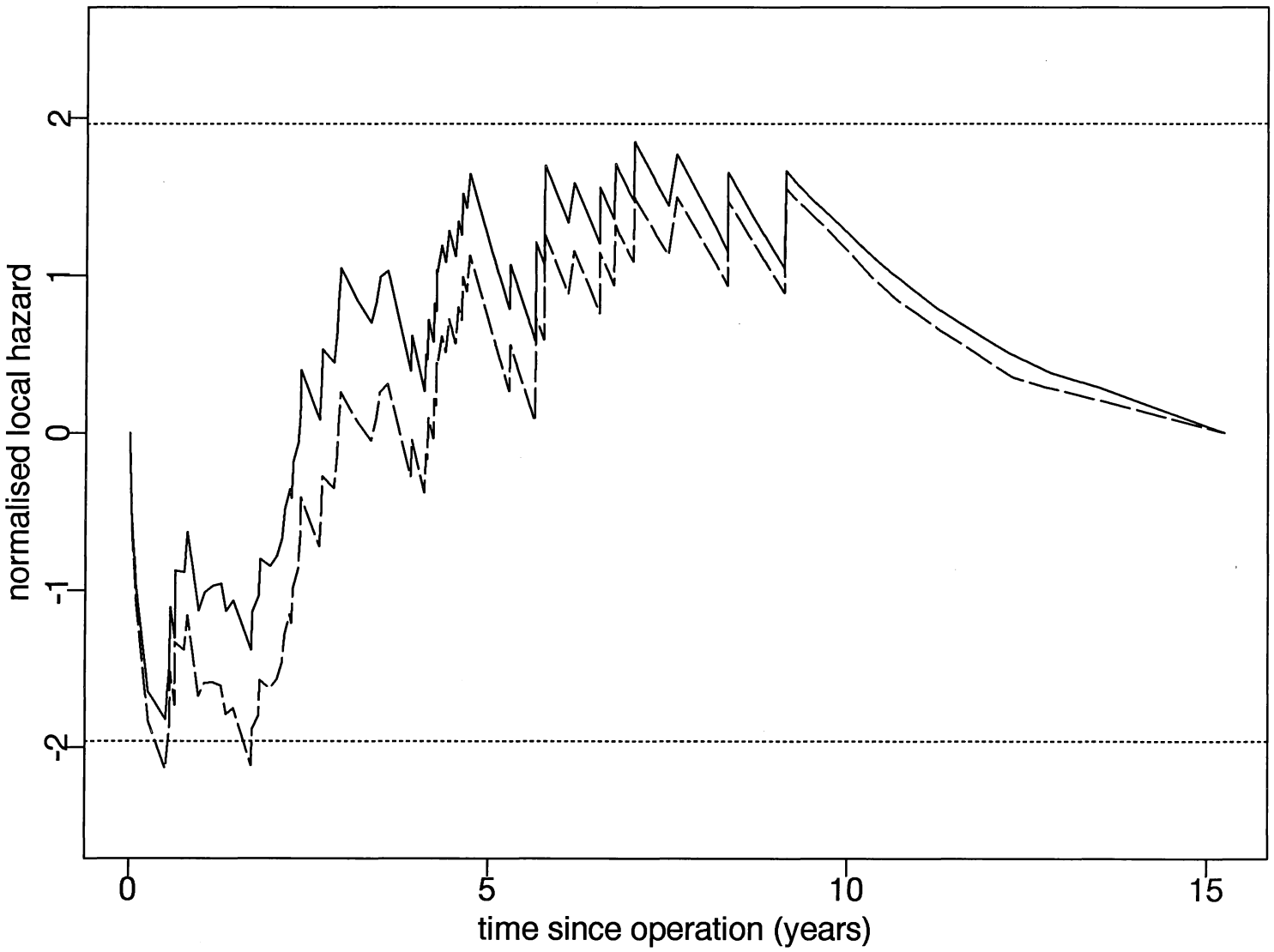
normalised local hazard
2
1
0
-1
-2
0
5
10
15
time since operation (years)

Figure 9.5.b.

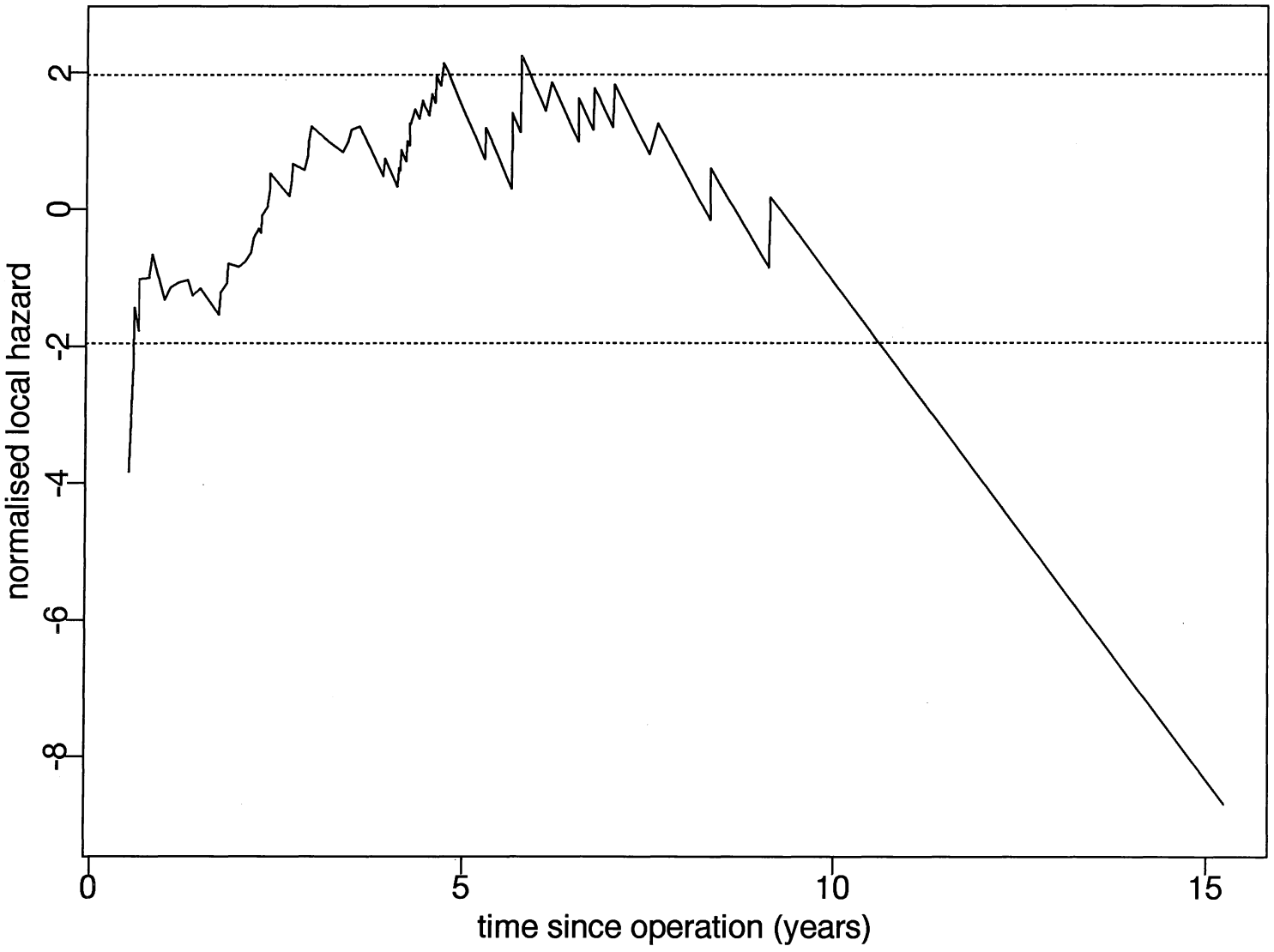
normalised local hazard
2
0
-2
-4
-6
-8
0
5
10
15
time since operation (years)

Figure 9.5.c.

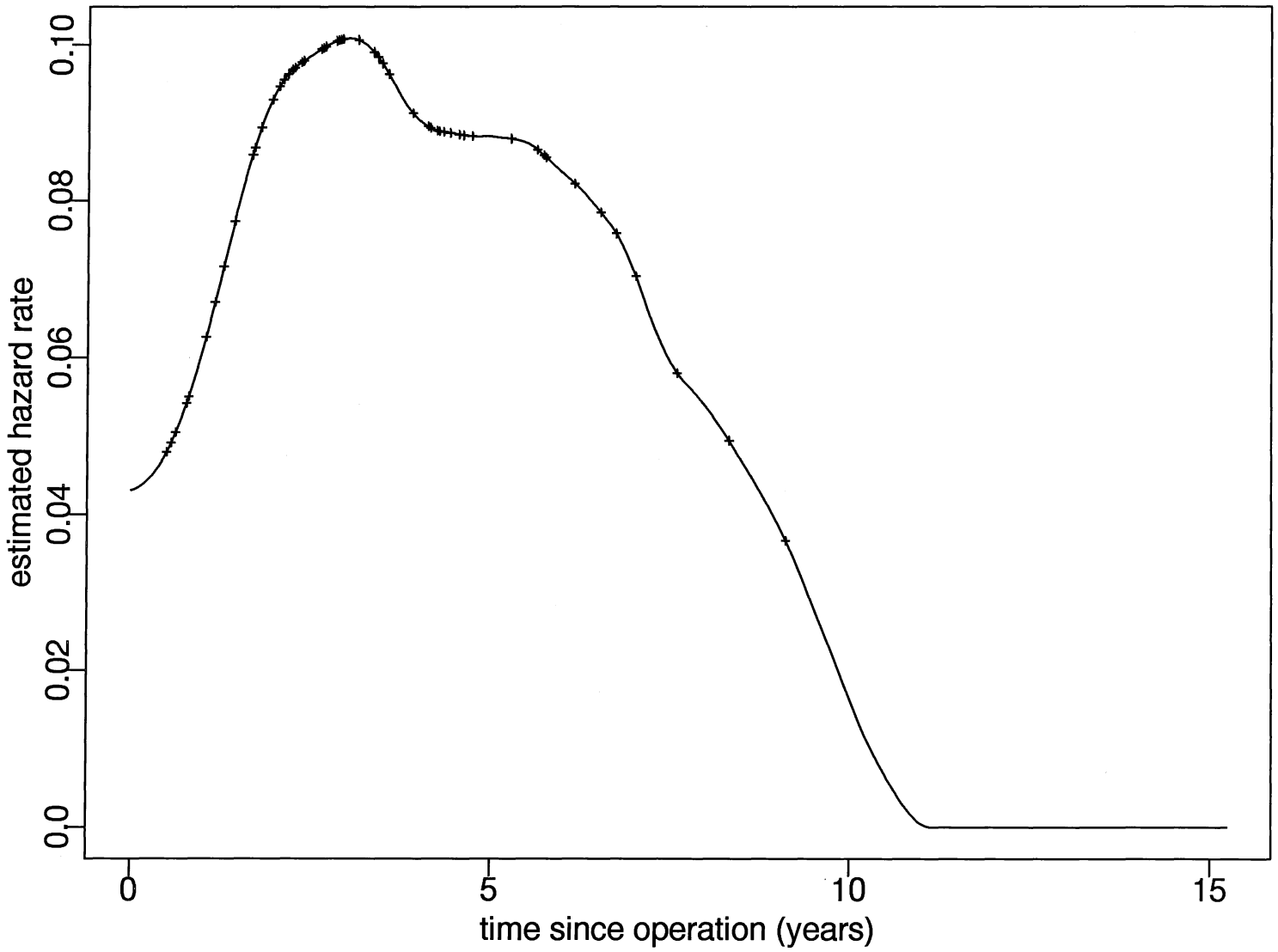
estimated hazard rate
0.0
0.02
0.04
0.06
0.08
0.10
0
5
10
15
time since operation (years)

Figure 9.5.d.

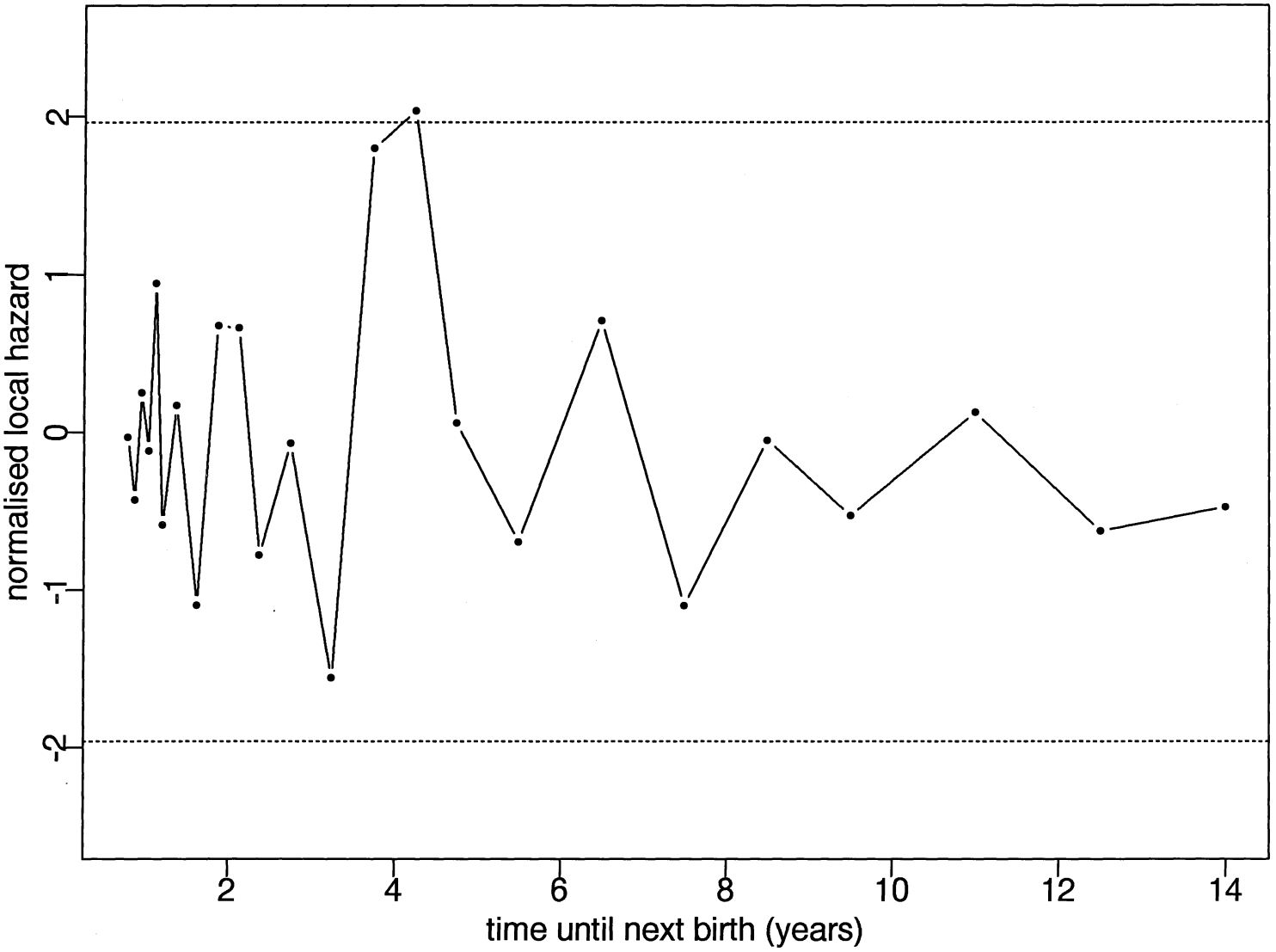
normalised local hazard
2
1
0
-1
-2
2
4
6
8
10
12
14
time until next birth (years)

Figure 9.6.