Biometrika (2018), **105**, 4, *pp.* 905–916
Printed in Great Britain
doi: 10.1093/biomet/asy035
Advance Access publication 21 August 2018

Transforming cumulative hazard estimates

By PÅL C. RYALEN, MATS J. STENSRUD and KJETIL RØYSLAND

Department of Biostatistics, University of Oslo, Domus Medica Gaustad, Sognsvannsveien 9, 0372 Oslo, Norway

p.c.ryalen@medisin.uio.no m.j.stensrud@medisin.uio.no kjetil.roysland@medisin.uio.no

Summary

Time-to-event outcomes are often evaluated on the hazard scale, but interpreting hazards may be difficult. Recently in the causal inference literature concerns have been raised that hazards actually have a built-in selection bias that prevents simple causal interpretations. This is a problem even in randomized controlled trials, where hazard ratios have become a standard measure of treatment effects. Modelling on the hazard scale is nevertheless convenient, for example to adjust for covariates; using hazards for intermediate calculations may therefore be desirable. In this paper we present a generic method for transforming hazard estimates consistently to other scales at which these built-in selection biases are avoided. The method is based on differential equations and generalizes a well-known relation between the Nelson–Aalen and Kaplan–Meier estimators. Using the martingale central limit theorem, we show that covariances can be estimated consistently for a large class of estimators, thus allowing for rapid calculation of confidence intervals. Hence, given cumulative hazard estimates based on, for example, Aalen's additive hazard model, we can obtain many other parameters without much more effort. We give several examples and the associated estimators. Coverage and convergence speed are explored via simulations, and the results suggest that reliable estimates can be obtained in real-life scenarios.

Some key words: Asymptotics; Delta method; Hazard; Marginal effect measure; Survival analysis.

1. Introduction

Applied researchers often have to deal with time-to-event outcomes. The temporal aspect makes causal inference for such outcomes challenging, since effects and biases could be time-varying. In particular, the proportional hazards model provides the standard approach to analysing survival data in medicine, but the assumption of proportional hazards cannot be justified in many real-life scenarios. Furthermore, since estimating hazards involves conditioning on recent survival, the presence of a seemingly innocent unobserved heterogeneity may mean that we will condition on a so-called collider, and therefore activate non-causal pathways from the exposure to the risk of an event at short term (Aalen et al., 2015). These commonly used effect measures, such as the hazard ratio, are hard to interpret causally as short-term risks (Robins & Greenland, 1989; Greenland, 1996; Aalen et al., 2008, 2015; Hernán, 2010; Stensrud et al., 2017).

However, to evaluate time-to-event outcomes, we do not need to assess hazards per se. Rather, we may be interested in effect measures on, say, the survival scale. Such measures can be easier to interpret, and may allow for causal evaluations (Hernán, 2010; Hernán & Robins, 2017). The survival function, for example, will not suffer from the collider bias, since no conditioning on survival is needed. Heuristically, the survival function at any time t is a function of the whole

study sample and is not derived only from a surviving subpopulation. Still, modelling the hazard scale could be useful as an intermediate step, for example to adjust for covariates.

Additive hazard models are becoming more popular, at least in the causal inference literature (Lange & Hansen, 2011; Martinussen et al., 2011; Tchetgen Tchetgen et al., 2015). In contrast to the proportional hazards model, the additive models are collapsible (Martinussen & Vansteelandt, 2013), and they easily allow for time-dependent effects. The additive models are also less prone to selection biases when unmeasured factors follow linear structural models (Martinussen & Vansteelandt, 2013; Strohmaier et al., 2015). On the other hand, additive models have been criticized because the effect estimates may be harder to understand. Often effect estimates are directly plotted as cumulative hazard curves, which may seem unsatisfactory to applied researchers (Bradburn et al., 2003). Indeed, the cumulative hazard curves may themselves not have an immediate causal interpretation.

With the above in mind, we aim to transform cumulative hazard estimates by using the theory of ordinary differential equations. We will estimate parameters that solve such equations, which are typically driven by cumulative hazards. The estimators we consider are solutions of naturally associated stochastic differential equations that are straightforward to solve numerically on a computer. Although we are mostly concerned with such equations that are driven by cumulative hazards, our method is not limited to this setting. We use a simple variant of the delta method to prove that applying our transformations to consistent cumulative hazard estimators also yields consistent estimators. Furthermore, we provide estimates of the asymptotic variance. In fact, our approach to finding the variance not only is simple but also allows us to go further than the functional delta method suggested by Gill et al. (1989) and Andersen et al. (1993). The functional delta method involves topologies that sometimes need to be carefully designed for each example and may be difficult to handle in practice; see, for example, van der Vaart (1998). In other words, the functional delta method would require us to derive results for each parameter separately, by linearizing an explicit expression of the parameter. In contrast, we consider parameters nonexplicitly in the form of differential equations that are already linearized. This allows us to use general stability results for stochastic differential equations to show consistency of the naturally associated estimators. Our approach is thus more generic than the functional delta method.

2. PARAMETERS

The parameters we consider are functions on a finite interval $[0, \mathcal{T}]$ that are solutions to an ordinary differential equation system of the form

$$X_t = X_0 + \int_0^t F(X_{s-})\,\mathrm{d}A_s. \tag{1}$$

Here A is a continuous k-dimensional function of bounded variation, and $F : \mathbb{R}^n \to M_{n,k}(\mathbb{R})$ is sufficiently smooth and satisfies a linear growth bound. Our differential equation formulation is motivated by a well-known relationship between cumulative hazards and survival functions; if T is a terminal endpoint and S_t and A_t are the accompanying survival and cumulative hazard, then the survival can be written as $S_t = 1 - \int_0^t S_{s-}\,\mathrm{d}A_s$.

There are several examples of parameters of this form in the survival analysis literature, but (1) has not received much attention. In the following we show some examples of parameters that can be expressed in this way. More examples can be found in the Supplementary Material.

Example 1 (Relative survival). When assessing effects of exposures, it is intriguing to consider the probability of survival directly, rather than estimating, say, hazard ratios. The relative survival

is routinely used in clinical medicine (Perme et al., 2012). In particular, five-year relative survival rates are standard measures of survival in cancer patients versus the general population, and such numbers are conventionally reported from cancer registries. Suppose that we are given two groups 1 and 0, and we want to identify the relative survival $\text{RS}_t = S_t^1/S_t^0$ of the positive group compared with the negative group at time t. If we let A^1 and A^0 denote the corresponding cumulative hazards, then a simple calculation of derivatives shows that

$$\text{RS}_t = \text{RS}_0 + \int_0^t \left(-\text{RS}_{s-},\ \text{RS}_{s-}\right) \mathrm{d}\begin{pmatrix} A_s^1 \\ A_s^0 \end{pmatrix}.$$

Example 2 (Mean and restricted mean survival). Instead of studying the relative survival as a function of time, we may be interested in the mean survival in different populations. Indeed, the mean survival is simply the area under the survival curve. In practice, however, the tail of the estimated survival function will strongly influence the integral of the estimated survival curve, and the mean survival estimates may be unreliable. As a remedy, it is possible to study the restricted mean survival $E\{TI(T \leqslant \tau)\}$, i.e., the area R_τ under the survival curve up to some restricted time τ. The restricted mean survival up to time τ has a useful interpretation: it is the mean event-free survival up to τ. In practice, τ may be a clinically important time-point that is prespecified. Advocates claim that the restricted mean survival to a prespecified τ should be reported in clinical trials (Royston & Parmar, 2011, 2013; Trinquart et al., 2016), particularly when the proportional hazard assumption is invalid. Indeed, R_τ is readily found by solving the system

$$\begin{pmatrix} R_\tau \\ S_\tau \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \int_0^\tau \begin{pmatrix} S_s & 0 \\ 0 & -S_{s-} \end{pmatrix} \mathrm{d}\begin{pmatrix} s \\ A_s \end{pmatrix},$$

where A is the cumulative hazard for death.

We may compare the restricted mean survival in different groups. In particular, the life expectancy difference and life expectancy ratio represent, respectively, the absolute and relative differences in restricted mean survival (Dehbi et al., 2017). Explicit expressions for these quantities are given in the Supplementary Material. In contrast to the hazard ratio, these parameters may be easier to interpret causally, because they do not involve conditioning on survival.

Example 3 (Cumulative incidence and competing risks). In practice we are often interested in the time to a particular event, but our study may be disrupted by other, competing events. For example, in medical contexts we may be interested in the time to onset of a disease D, but subjects may die before they develop D. If there are competing risks, it is incorrect to assume a one-to-one relation between cause-specific hazards and the cumulative incidence (Andersen et al., 2012); for example, if we treat death as censoring, we obtain estimates from a hypothetical population in which subjects cannot die without D.

Alternatively, we can include the hazards of the competing events in our model. Consider a situation with k competing causes of death, with cause-specific cumulative hazards $A^1, \ldots, A^k$. Following Andersen et al. (2002), we see that

$$\begin{pmatrix} S_t \\ C_t^1 \\ \vdots \\ C_t^k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \int_0^t \begin{pmatrix} -S_{s-} & -S_{s-} & \cdots & -S_{s-} \\ S_{s-} & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & S_{s-} \end{pmatrix} \mathrm{d}\begin{pmatrix} A_s^1 \\ \vdots \\ A_s^k \end{pmatrix}$$

describes all the cause-specific cumulative incidences $C_t^1, \ldots, C_t^k$, subject to the remaining competing risks. To intuitively understand our expression for cumulative incidences, we emphasize that a subject survives if she is unaffected by the k mutually exclusive causes of death, which motivates the top row. Experiencing a cause-specific death of type j by time t happens only if a subject survived other causes up to s. Summing over all s gives us the integrals above.

Example 4 (Cumulative sensitivity and specificity after screening). The effectiveness of disease screening is frequently studied in medicine. Often the screening is performed at $t = 0$ but the disease may advance or develop at times $t > 0$. Subjects may be followed over time to assess whether certain events occur, such as the onset of disease. In such scenarios, conventional methods of assessing sensitivity and specificity are inadequate, because they do not account for differences in follow-up times. We instead express sensitivity and specificity as functions of time, as suggested by Chen et al. (2012) and Nygård et al. (2014). Let A^1 be the cumulative hazard of the disease event for the group with positive screening results, A^0 the cumulative hazard of the negative results, β the prevalence of the disease, and Z the screening outcome. We let U_t denote the cumulative positive predictive value $\mathrm{pr}(N_t > 0 \mid Z = 1)$, let V_t denote the cumulative negative predictive value $\mathrm{pr}(N_t = 0 \mid Z = 0)$, let W_t denote the cumulative sensitivity $\mathrm{pr}(Z = 1 \mid N_t > 0)$, and let X_t denote the cumulative specificity $\mathrm{pr}(Z = 0 \mid N_t = 0)$. Applying Bayes' rule gives the equation

$$\begin{pmatrix} U_t \\ V_t \\ W_t \\ X_t \end{pmatrix} = \begin{pmatrix} U_0 \\ V_0 \\ W_0 \\ X_0 \end{pmatrix} + \int_0^t \begin{pmatrix} 1 - U_{s-} & 0 \\ 0 & -V_{s-} \\ \frac{W_{s-}^2 (1-V_{s-})(1-U_{s-})}{\gamma U_{s-}^2} & -\frac{W_{s-}^2 V_{s-} U_{s-}}{\gamma U_{s-}^2} \\ \frac{\gamma X_{s-}^2 (1-U_{s-})}{V_{s-}} & -\frac{\gamma X_{s-}^2 (1-U_{s-})}{V_{s-}} \end{pmatrix} \mathrm{d}\begin{pmatrix} A_s^1 \\ A_s^0 \end{pmatrix},$$

where we have introduced the odds $\gamma = \beta/(1-\beta)$ and where V_{s-} and U_{s-} cannot be too small.

3. PLUG-IN ESTIMATORS

The estimators we propose will build on a cumulative hazard estimator $\hat{A}^{(n)}$ that is given by counting process integrals, that is,

$$\hat{A}_t^{(n)} = \int_0^t G_{s-}^{(n)} \, \mathrm{d}N_s^{(n)}, \tag{2}$$

where $G^{(n)} = (g^{(n)1} \ldots g^{(n)l})$ is a predictable $(k \times l)$-dimensional matrix-valued process and $N^{(n)}$ is an adapted l-dimensional counting process. It should be noted that the results here are applicable to more general situations than (2). Suppose this estimator is consistent in the sense that

$$\lim_{n \to \infty} \mathrm{pr}\Big(\sup_{s \leqslant T} \big| \hat{A}_s^{(n)} - A_s \big| \geqslant \epsilon \Big) = 0$$

for every $\epsilon > 0$, or such that

$$W^{(n)} = n^{1/2} \big(\hat{A}^{(n)} - A \big)$$

converges in law to an independent-increments zero-mean Gaussian local martingale W, relative to the Skorohod metric. The latter is typically a consequence of the martingale central limit

theorem, as stated in Andersen et al. (1993, Theorem II.5.1) or Jacod & Shiryaev (2003, Theorem VIII.3.22). Two standard examples in event history analysis where this condition holds are the Nelson–Aalen cumulative hazard estimator and Aalen's additive hazard regression (Andersen et al., 1993, Theorems VII.4.1, IV.1.1 and IV.1.2).

In addition to consistency, another property of $\hat{A}^{(n)}$ that will be crucial in this setting is predictably uniform tightness; see Jacod & Shiryaev (2003, VI.6a) for an exact definition. However, as we will not need this property in full generality, we can use the following lemma to determine whether processes are predictably uniformly tight in most of the situations we might encounter.

LEMMA 1. *Let $\{Z^{(n)}\}_n$ be a sequence of semi-martingales on $[0, \mathcal{T}]$, and let $\{\rho^{(n)}\}_n$ be predictable processes such that every*

$$M_t^{(n)} = Z_t^{(n)} - \int_0^t \rho_s^{(n)} \, \mathrm{d}s$$

defines a square-integrable local martingale. Suppose that:

(i) $\lim_{J\to\infty} \sup_n \mathrm{pr}(\sup_{s\leqslant\mathcal{T}} |\rho_s^{(n)}|_1 \geqslant J) = 0$;
(ii) $\lim_{J\to\infty} \sup_n \mathrm{pr}(\mathrm{tr}\langle M^{(n)}\rangle_{\mathcal{T}} \geqslant J) = 0$.

Then $\{Z^{(n)}\}_n$ is predictably uniformly tight.

We are now ready to define the estimators for parameters defined through (1).

DEFINITION 1. *Let $\hat{X}_0^{(n)}$ be an estimator for X_0. The plug-in estimator $\hat{X}^{(n)}$ for X that satisfies (1) is the solution of the stochastic differential equation*

$$\hat{X}_t^{(n)} = \hat{X}_0^{(n)} + \int_0^t F(\hat{X}_{s-}^{(n)}) \, \mathrm{d}\hat{A}_s^{(n)}. \tag{3}$$

Plug-in estimators are relatively easy to implement on a computer due to their recursive form. If $\tau_1, \tau_2, \ldots$ are the jump times of $\hat{A}^{(n)}$, then

$$\hat{X}_t^{(n)} = \hat{X}_{\tau_{k-1}}^{(n)} + F(\hat{X}_{\tau_{k-1}}^{(n)}) \Delta\hat{A}_{\tau_k}^{(n)} \tag{4}$$

whenever $\tau_k \leqslant t < \tau_{k+1}$. These estimators are also consistent in many situations; we formulate this result as a theorem, which is proved in the Supplementary Material.

THEOREM 1. *Suppose that X satisfies the ordinary differential equation (1), where F is locally Lipschitz continuous, and also satisfies the linear growth condition on a domain that contains $\{X_t : t \in [0, \mathcal{T}]\}$. Assume furthermore that $\{\hat{A}^{(n)}\}_n$ and $\{\hat{X}_0^{(n)}\}_n$ are consistent, i.e.,*

$$\lim_{n\to\infty} \mathrm{pr}\Big(\sup_{s\leqslant\mathcal{T}} \big|\hat{A}_s^{(n)} - A_s\big| \geqslant \epsilon \Big) = 0, \quad \lim_{n\to\infty} \mathrm{pr}\Big(\big|\hat{X}_0^{(n)} - X_0\big| \geqslant \epsilon \Big) = 0$$

for every $\epsilon > 0$, and that $\hat{A}^{(n)}$ is predictably uniformly tight; see Proposition 1 for additive hazards. Then

$$\lim_{n\to\infty} \mathrm{pr}\Big(\sup_{s\leqslant\mathcal{T}} \big|\hat{X}_s^{(n)} - X_s\big| \geqslant \epsilon \Big) = 0$$

for every $\epsilon > 0$; that is, $\{\hat{X}^{(n)}\}_n$ defines a consistent estimator of X.

We can also handle the case where $A_t = t$; see the Supplementary Material for details.

If $n^{1/2}(\hat{A}_t^{(n)} - A_t)$ defines a local martingale that satisfies the martingale central limit theorem, then the sequence $\{n^{1/2}(\hat{X}_t^{(n)} - X_t)\}$ converges to a solution of a given stochastic differential equation that is driven by a Gaussian process, as described in the next theorem.

THEOREM 2. *Suppose that* $F = \big(F_1, \ldots, F_k\big)$ *has bounded and continuous first and second derivatives on a domain that contains* $\{X_t : t \in [0, \mathcal{T}]\}$. *Consider*

$$Z_0^{(n)} = n^{1/2}\big(\hat{X}_0^{(n)} - X_0\big), \qquad W^{(n)} = n^{1/2}\big(\hat{A}^{(n)} - A\big)$$

such that $Z_0^{(n)}$ *converges in law to a zero-mean Gaussian vector* Z_0 *as* $n \to \infty$, *and* $W^{(n)}$ *converges in law, relative to the Skorohod metric, to a zero-mean Gaussian martingale with independent increments and is predictably uniformly tight. Suppose also that* $\{[W^{(n)}]\}_n$ *is predictably uniformly tight (see Proposition 1 for additive hazards). Let* Z *denote the unique solution to the stochastic differential equation*

$$Z_t = Z_0 + \sum_{j=1}^{k} \int_0^t \nabla F_j(X_{s-}) Z_{s-} \, \mathrm{d}A_s^j + \int_0^t F(X_{s-}) \, \mathrm{d}W_s.$$

Then Z *is a zero-mean Gaussian process, and*

$$Z^{(n)} = n^{1/2}\big(\hat{X}^{(n)} - X\big)$$

converges in law, relative to the Skorohod metric, to Z *as* $n \to \infty$.

Moreover, let V *denote the covariance of* Z. *Suppose that* $\{\hat{V}_0^{(n)}\}$ *is a consistent estimator for* V_0, *and let* $\hat{V}^{(n)}$ *denote the solution of the stochastic differential equation*

$$\begin{aligned} \hat{V}_t^{(n)} = \hat{V}_0^{(n)} &+ \sum_{j=1}^{k} \int_0^t \hat{V}_{s-}^{(n)} \nabla F_j(\hat{X}_{s-}^{(n)})^{\mathrm{T}} + \nabla F_j(\hat{X}_{s-}^{(n)}) \hat{V}_{s-}^{(n)} \, \mathrm{d}\hat{A}_s^{(n)j} \\ &+ n \int_0^t F(\hat{X}_{s-}^{(n)}) \, \mathrm{d}[\hat{A}^{(n)}]_s F(\hat{X}_{s-}^{(n)})^{\mathrm{T}}. \end{aligned} \tag{5}$$

Then $\hat{V}^{(n)}$ *defines a consistent estimator of* V, *i.e.,*

$$\lim_{n \to \infty} \mathrm{pr}\left(\sup_{s \leqslant T} \big| \hat{V}_s^{(n)} - V_s \big| \geqslant \epsilon \right) = 0$$

for every $\epsilon > 0$, *and* V *solves the ordinary differential equation*

$$V_t = V_0 + \sum_{j=1}^{k} \int_0^t V_s \nabla F_j(X_s)^{\mathrm{T}} + \nabla F_j(X_s) V_s \, \mathrm{d}A_s^j + \int_0^t F(X_s) \, \mathrm{d}[W]_s F(X_s)^{\mathrm{T}}.$$

We can compute covariances and hence also pointwise confidence intervals for our estimates based on the differential equations from § 1, by substituting $\hat{X}^{(n)}$ for X and $\hat{A}^{(n)}$ for A. Analogously

to (3), it is straightforward to solve the stochastic differential equation (5), since it is driven by jump processes. If $\tau_1, \tau_2, \ldots$ denote the jump times of $\hat{A}^{(n)}$, then we have that

$$\begin{aligned}\hat{V}_t^{(n)} = \hat{V}_{\tau_{k-1}}^{(n)} + \sum_{j=1}^{k} \left\{ \hat{V}_{\tau_{k-1}}^{(n)} \nabla F_j(\hat{X}_{\tau_{k-1}}^{(n)})^{\mathrm{T}} + \nabla F_j(\hat{X}_{\tau_{k-1}}^{(n)}) \hat{V}_{\tau_{k-1}}^{(n)} \right\} \Delta \hat{A}_{\tau_k}^{(n)j} \\ + nF(\hat{X}_{\tau_{k-1}}^{(n)}) H_{\tau_k}^{(n)} F(\hat{X}_{\tau_{k-1}}^{(n)})^{\mathrm{T}},\end{aligned} \tag{6}$$

whenever $\tau_k \leqslant t < \tau_{k+1}$. Here $H_t^{(n)}$ is the $k \times k$ matrix $[\hat{A}^{(n)} - A]_t$. Following Remark 1 in the Supplementary materials we find that $\left(H_t^{(n)}\right)_{i,j} = 0$ if either $A^i = t$ or $A^j = t$. Otherwise, $\left(H_t^{(n)}\right)_{i,j} = \Delta \hat{A}_t^{(n),i} \Delta \hat{A}_t^{(n),j}$.

The assumptions of Theorems 1 and 2 are indeed satisfied for Aalen's additive regression under relatively weak assumptions.

PROPOSITION 1. *Assume that we have n independent and identically distributed individuals and that the intensity of each N^i at time t is of the form $\lambda_t^i = Y_t^i \alpha_t^{\mathrm{T}} U_{t-}^i$ where α is bounded and continuous. Write $U^{(n)}$ for the matrix in which the ith row is equal to U^i, and write $Y^{(n)}$ for the diagonal matrix where the ith diagonal element is equal to Y^i. Suppose that $E(\sup_{t \leqslant \mathcal{T}} |U_t^i|_3^3) < \infty$ for each i and that*

$$\lim_{J \to \infty} \inf_n \mathrm{pr}\left[\sup_{t \leqslant \mathcal{T}} \mathrm{tr}\left\{ \left(\frac{U_{t-}^{(n)\mathrm{T}} Y_t^{(n)} U_{t-}^{(n)}}{n} \right)^{-1} \right\} \geqslant J \right] = 0.$$

Consider

$$\begin{aligned}\hat{A}_t^{(n)} &= \int_0^t \left(U_{s-}^{(n)\mathrm{T}} Y_s^{(n)} U_{s-}^{(n)} \right)^{-1} U_{s-}^{(n)\mathrm{T}} Y_s^{(n)} \, \mathrm{d}N_s^{(n)}, \\ W^{(n)} &= n^{1/2}(\hat{A}^{(n)} - A).\end{aligned}$$

Then $\{\hat{A}^{(n)}\}_n$, $\{[W^{(n)}]\}_n$ and $\{W^{(n)}\}_n$ are predictably uniformly tight and the latter converges in law, relative to the Skorohod metric, towards a zero-mean Gaussian martingale with independent increments.

Using (4) and (6) we are thus able to express plug-in estimators for our class of parameters and their variances generically. A simple yet illustrative example is the plug-in estimator for the survival function; employing the notation used after Definition 1, we see that it takes the form $\hat{S}_t = \hat{S}_{\tau_{k-1}} - \hat{S}_{\tau_{k-1}} \Delta \hat{A}_{\tau_k}$. This is none other than the Kaplan–Meier estimator expressed as a difference equation. The plug-in variance estimator in this case is $\hat{V}_t = \hat{V}_{\tau_k - 1} - 2\hat{V}_{\tau_k - 1} \Delta \hat{A}_{\tau_k} + n\hat{S}_{\tau_{k-1}} (\Delta \hat{A}_{\tau_k})^2$, which is different from the commonly used Greenwood estimator. The plug-in estimators for several parameters and their variances are given in the Supplementary Material. Simulations of selected parameters are shown in Figures 1 and 2.

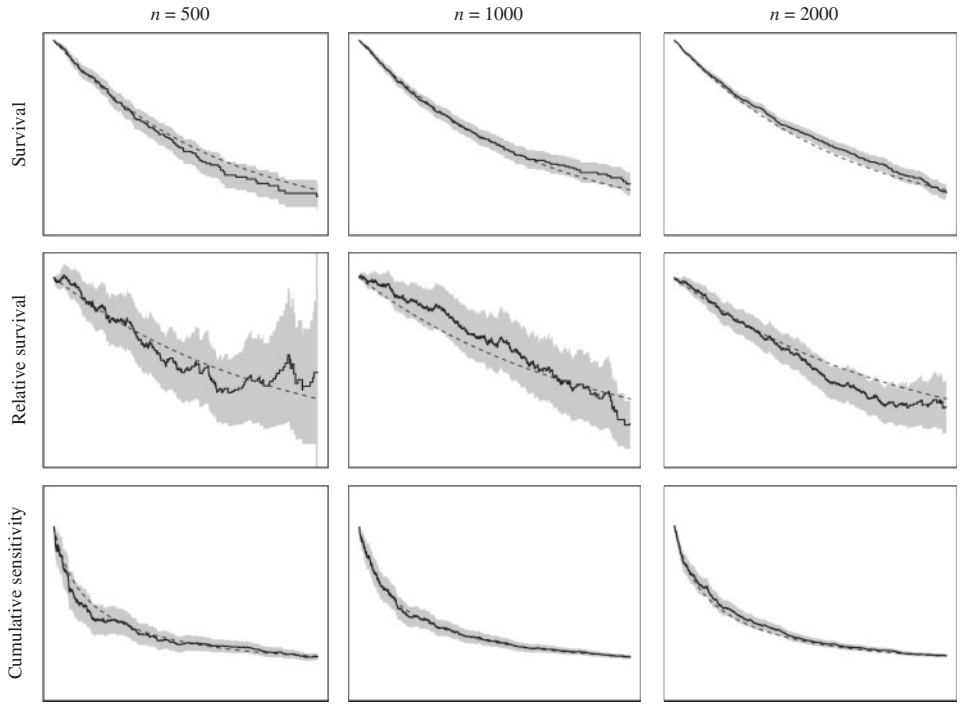

Fig. 1. Plug-in estimates for selected parameters with population sizes 500, 1000 and 2000, along with 95% pointwise confidence intervals; exact parameters are represented by dashed lines. Parameters not shown as examples here are displayed in the Supplementary Material.

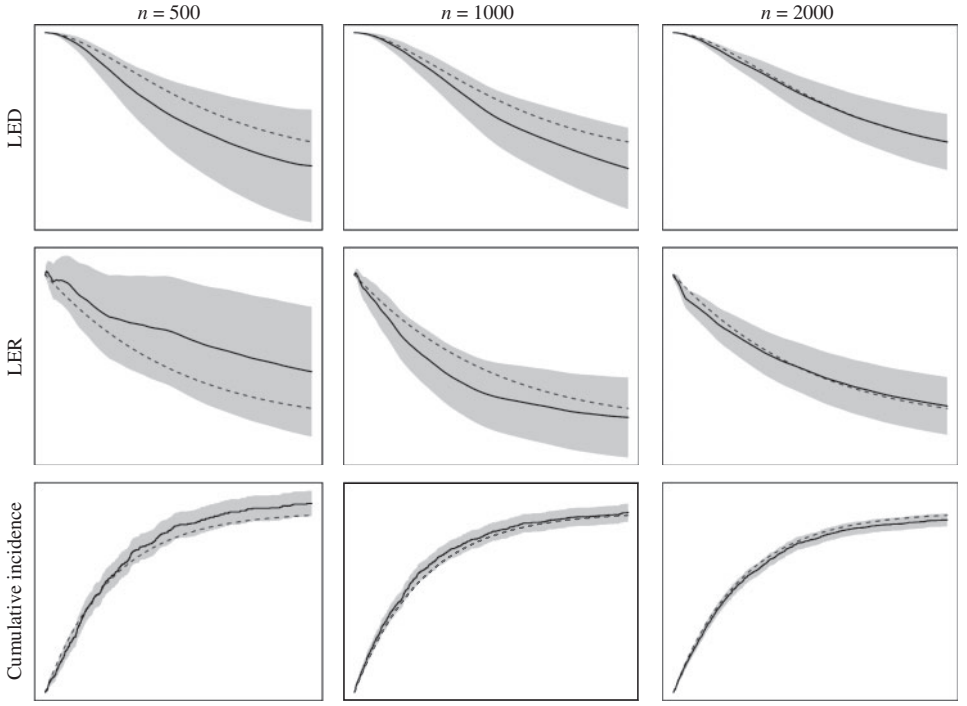

Fig. 2. Plug-in estimates for selected parameters with population sizes 500, 1000 and 2000, along with 95% pointwise confidence intervals; exact parameters are represented by dashed lines. Parameters not shown as examples here are displayed in the Supplementary Material. LED, life expectancy difference; LER, life expectancy ratio.

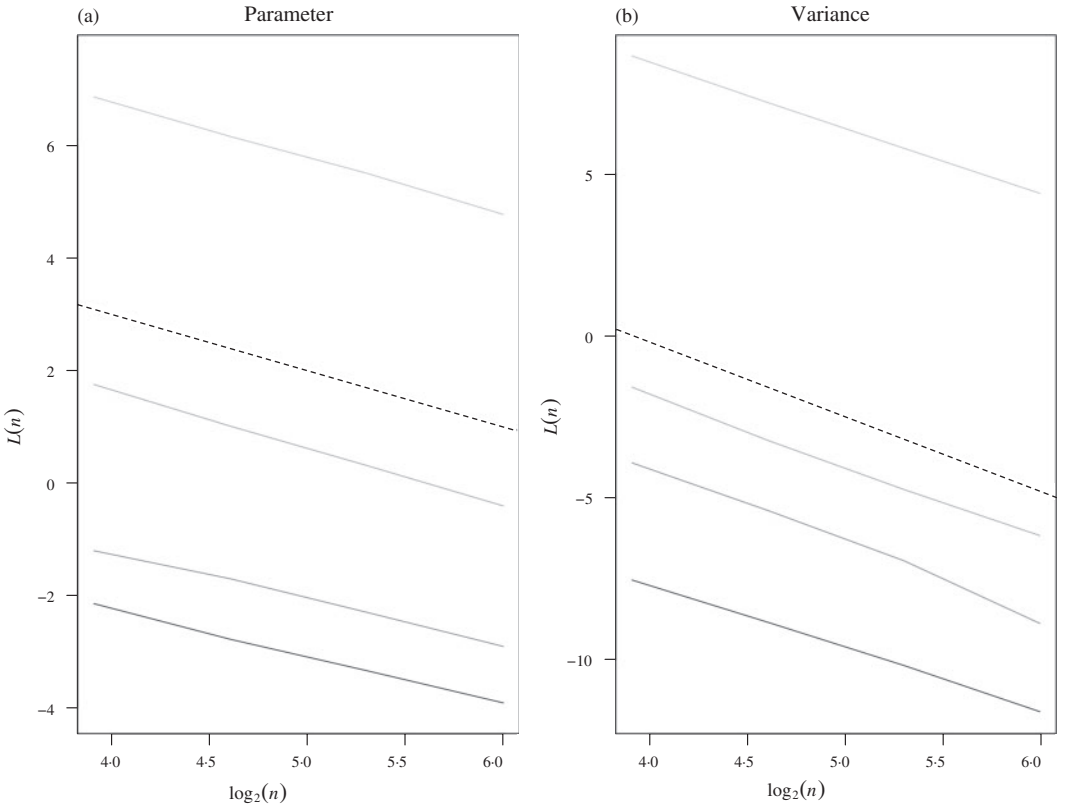

Fig. 3. Convergence of selected plug-in estimators; (a) parameter and (b) variances. We used dotted lines to indicate convergence of order 1 in (a) and 2.3 in (b). The solid lines are, from top to bottom in both panels: life expectancy difference, life expectancy ratio, survival, and cumulative incidence.

4. Performance

4·1. *Convergence*

Our assessment of convergence involves two steps. In step 1 we find close approximations to the parameters, denoted by $\tilde{X}$. In step 2 we calculate plug-in estimators for a range of sample sizes; we then check how fast the plug-in estimators converge to $\tilde{X}$ as the sample size increases.

For performing step 1 we have developed code that generates survival data for the given input hazards. Using the hazards we can calculate $\tilde{X}$ with the desired precision. For instance, knowing the hazard α we can obtain the survival $\exp(-\int \alpha_s \, \mathrm{d}s)$ by numerical integration.

For step 2 we simulate data, which enables us to estimate cumulative hazards and hence parameters specified by (3). This step is undertaken for a range of sample sizes. For each sample size n we simulate k populations, so that we obtain plug-in estimates $\hat{X}^{n,j}$ for $j = 1, \ldots, k$. We check convergence using the following L^2 criterion:

$$L(n) = \frac{1}{k} \sum_{j=1}^{k} \int_0^{\mathcal{T}} |\tilde{X}_s - \hat{X}_s^{n,j}|^2 \, \mathrm{d}s. \tag{7}$$

To assess convergence of the plug-in variance estimators, we first obtain a variance estimate from a large bootstrap sample. As in step 2 above, we simulate data and obtain k plug-in variance estimates for a range of sample sizes n; this way the same criterion (7) can be used. Convergence plots are shown in Fig. 3.

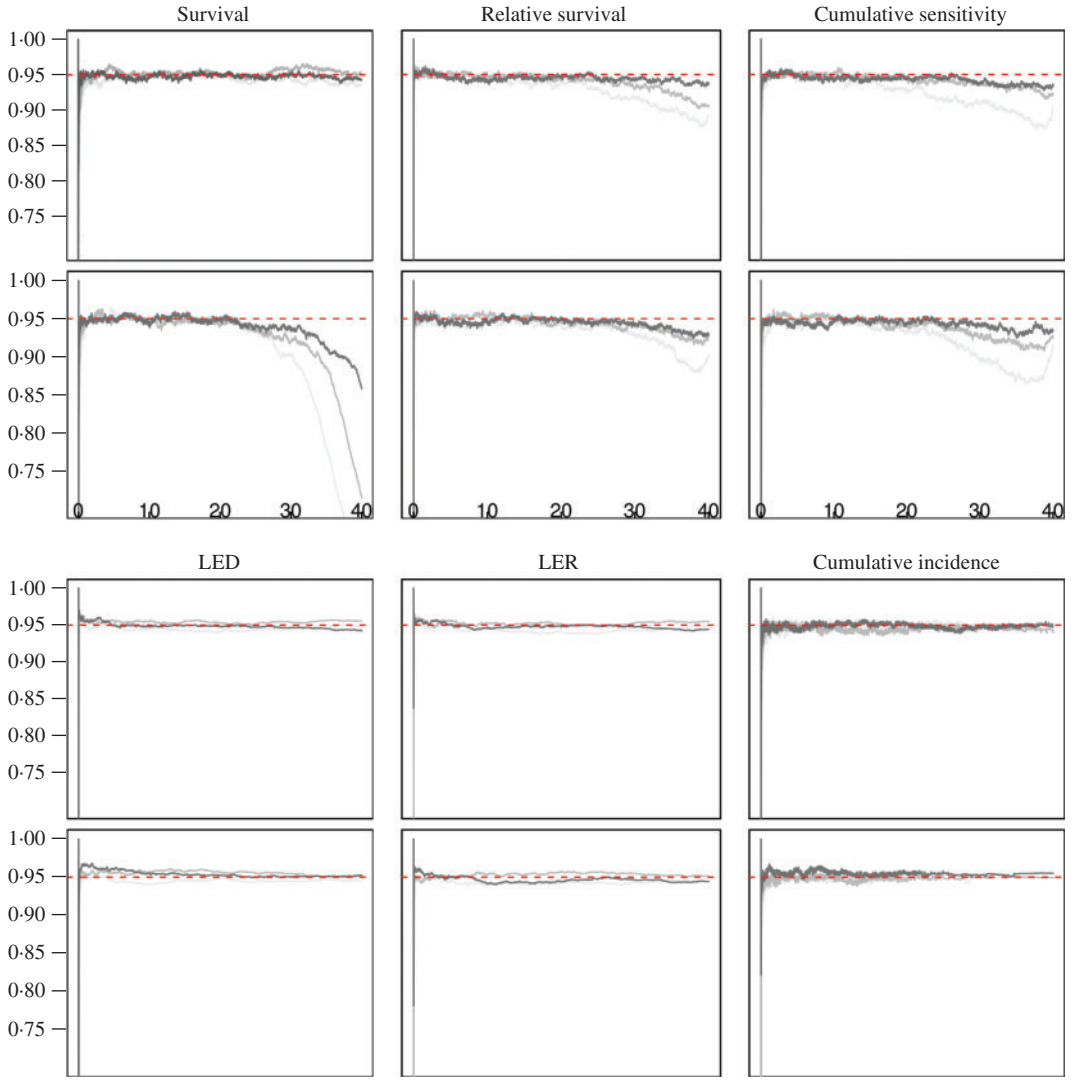

Fig. 4. Estimated mean coverage for selected parameters simulated with constant hazards (upper panels) and linearly crossing hazards (lower panels); in each panel the dashed line indicates the confidence level. $n = 500$, light grey; $n = 1000$, grey; $n = 2000$, dark grey.

Figure 3 suggests that the plug-in estimator for the life expectancy difference performs worse than the other estimators. This is because the variance depends on the time scale, since we are approximating Lebesgue integrals. The variance for the life expectancy ratio also depends on the time scale, but here the time contribution is smaller.

4·2. *Coverage*

To estimate coverage, we first obtain close approximations to the parameters, $\tilde{X}$, as in step 1 of § 4·1. Next, we simulate data to calculate plug-in estimates, which are used to obtain confidence intervals. This step is repeated until we have a large collection of confidence intervals. At time t, a Bernoulli trial is used to decide whether a confidence interval covers $\tilde{X}_t$. We estimate the expected coverage at time t by this Bernoulli probability, i.e., by the average number of confidence intervals that cover $\tilde{X}_t$.

We calculated coverage for two scenarios: constant hazards and linear (crossing) hazards over a time period $[0, \mathcal{T}]$. We selected linear hazards that were large initially but linearly decreasing, or small initially but linearly increasing, such that they crossed each other at the halfway point $\mathcal{T}/2$.

Coverage is plotted in Fig. 4, which suggests that the coverage drops below the confidence level in some scenarios. This behaviour is due to the plug-in variance estimators. The survival plug-in variance, for instance, will decrease drastically if there are events where few people are at risk, causing the survival confidence interval to be narrow thereafter. Incidentally, the Greenwood estimator would give similar performance, since it is small whenever the Kaplan–Meier curve is small.

Overall the plug-in estimators behave satisfactorily, not only when the hazards are constant but also when the hazards are linearly crossing.

Acknowledgement

All the authors were supported by the Norwegian Research Council.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of the theoretical results, along with several parameter examples and their plug-in estimators. R code for calculating the plug-in estimators is available at https://github.com/palryalen/transform.hazards.

References

Aalen, O., Borgan, O. & Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. New York: Springer.

Aalen, O. O., Cook, R. J. & Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal.* **21**, 579–93.

Andersen, P. K., Abildstrom, S. Z. & Rosthøj, S. (2002). Competing risks as a multi-state model. *Statist. Meth. Med. Res.* **11**, 203–15.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical Models Based On Counting Processes*. Springer Series in Statistics. New York: Springer.

Andersen, P. K., Geskus, R. B., de Witte, T. & Putter, H. (2012). Competing risks in epidemiology: Possibilities and pitfalls. *Int. J. Epidemiol.* **41**, 861–70.

Bradburn, M. J., Clark, T. G., Love, S. & Altman, D. (2003). Survival analysis part II: Multivariate data analysis – an introduction to concepts and methods. *Br. J. Cancer* **89**, 431–6.

Chen, L., Lin, D. Y. & Zeng, D. (2012). Predictive accuracy of covariates for event times. *Biometrika* **99**, 615–30.

Dehbi, H.-M., Royston, P. & Hackshaw, A. (2017). Life expectancy difference and life expectancy ratio: Two measures of treatment effects in randomised trials with non-proportional hazards. *Br. Med. J.* **357**, j2250.

Gill, R. D., Wellner, J. A. & Præstgaard, J. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1) [with Discussion and Reply]. *Scand. J. Statist.* **16**, 97–128.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* **7**, 498–501.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21**, 13–5.

Hernán, M. A. & Robins, J. M. (2017). *Causal Inference*. Boca Raton, Florida: CRC Press.

Jacod, J. & Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Berlin: Springer, 2nd ed.

Lange, T. & Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology* **22**, 575–81.

Martinussen, T. & Vansteelandt, S. (2013). On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Anal.* **19**, 279–96.

Martinussen, T., Vansteelandt, S., Gerster, M. & Hjelmborg, J. v. B. (2011). Estimation of direct effects for survival data by using the Aalen additive hazards model. *J. R. Statist. Soc.* B **73**, 773–88.

Nygård, M., Røysland, K., Campbell, S. & Dillner, J. (2014). Comparative effectiveness study on human papillomavirus detection methods used in the cervical cancer screening programme. *BMJ Open* **4**, e003460.

Perme, M. P., Stare, J. & Estève, J. (2012). On estimation in relative survival. *Biometrics* **68**, 113–20.

ROBINS, J. & GREENLAND, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics* **45**, 1125–38.

ROYSTON, P. & PARMAR, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statist. Med.* **30**, 2409–21.

ROYSTON, P. & PARMAR, M. K. (2013). Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med. Res. Methodol.* **13**, 152.

STENSRUD, M. J., VALBERG, M., RØYSLAND, K. & AALEN, O. O. (2017). Exploring selection bias by causal frailty models: The magnitude matters. *Epidemiology* **28**, 379–86.

STROHMAIER, S., RØYSLAND, K., HOFF, R., BORGAN, Ø., PEDERSEN, T. R. & AALEN, O. O. (2015). Dynamic path analysis – a useful tool to investigate mediation processes in clinical survival trials. *Statist. Med.* **34**, 3866–87.

TCHETGEN TCHETGEN, E. J., WALTER, S., VANSTEELANDT, S., MARTINUSSEN, T. & GLYMOUR, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26**, 402–10.

TRINQUART, L., JACOT, J., CONNER, S. C. & PORCHER, R. (2016). Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J. Clin. Oncol.* **34**, 1813–9.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*, vol. 3. Cambridge: Cambridge University Press.

[*Received on* 26 *October* 2017. *Editorial decision on* 5 *April* 2018]