ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Fifty Years of the Cox Model

John D. Kalbfleisch[1] and Douglas E. Schaubel[2]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA; email: jdkalbfl@umich.edu

[2]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; email: douglas.schaubel@pennmedicine.upenn.edu

Annu. Rev. Stat. Appl. 2023. 10:1–23

First published as a Review in Advance on November 21, 2022

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-033021-014043

Keywords

competing risks, independent censoring, intensity and marginal rate models, partial likelihood, recurrent events, time dependent covariates

Abstract

The Cox model is now 50 years old. The seminal paper of Sir David Cox has had an immeasurable impact on the analysis of censored survival data, with applications in many different disciplines. This work has also stimulated much additional research in diverse areas and led to important theoretical and practical advances. These include semiparametric models, nonparametric efficiency, and partial likelihood. In addition to quickly becoming the go-to method for estimating covariate effects, Cox regression has been extended to a vast number of complex data structures, to all of which the central idea of sampling from the set of individuals at risk at time t can be applied. In this article, we review the Cox paper and the evolution of the ideas surrounding it. We then highlight its extensions to competing risks, with attention to models based on cause-specific hazards, and to hazards associated with the subdistribution or cumulative incidence function. We discuss their relative merits and domains of application. The analysis of recurrent events is another major topic of discussion, including an introduction to martingales and complete intensity models as well as the more practical marginal rate models. We include several worked examples to illustrate the main ideas.

1. INTRODUCTION

Fifty years ago, on March 8, 1972, David Cox presented his paper entitled "Regression Models and Life-Tables" to the Royal Statistical Society. The paper was greeted as a milestone in the analysis of time-to-event data and attracted comments from many statisticians around the world. The article has had more than 57,000 citations, not including many more references to books or subsequent articles reviewing the model and methods; has influenced research in many directions and scientific disciplines; and has stimulated a vast body of work in statistics and its applications. Making this anniversary even more notable, Sir David Cox died on January 18, 2022, at the age of 97. He was, as his obituaries note, a man of great intelligence and energy who remained very active in research even into his tenth decade. He was a modest man, with friends and admirers everywhere in the world. He was always supportive of new researchers and was very generous with his ideas and suggestions. His mark in statistics, and science more generally, is broad. David Cox authored well over 300 papers and many books in varied statistical topics. The work discussed here is only one of many innovative and fundamental contributions.

Cox (1972) deals with the analysis of right-censored failure time data where, on each individual, there are one or more explanatory variables. A failure time is right-censored if it is observed only to exceed a given value, the censoring time. In many clinical trials, subjects enter the study over a time interval and are followed to some predetermined end of study time. Of interest is the time to failure (say, disease recurrence) where the origin is the time of entry for each subject. Any subject who has not failed by the end of the study is right-censored.

There had been previous analyses of right-censored survival data using parametric models, especially exponential models, extended in order to accommodate a regression structure (e.g., Feigl & Zelen 1965, Glasser 1967), and these papers indicated how other parametric models could be accommodated. There had also been developments in nonparametric methods, including the important work of Kaplan & Meier (1958) on nonparametric estimation of the survivor function based on right-censored data. In addition, there had been nonparametric methods, especially based on the work of Mantel & Haenszel (1959), including applications to life tables by Mantel (1963). In addition, an important paper by Peto & Peto (1972), which appeared more or less concurrently with that of Cox (1972), defined, named, and explored the properties of the log rank test, which, like the Mantel–Haenszel work, has close connections to the Cox model and its analysis.

The article by Cox (1972) represented a quantum leap from earlier work and presented a comprehensive approach to problems with right-censored data. The work also stimulated much research in various areas. For example, this was the first example of what became known as semiparametric regression models, in which there are both parametric and nonparametric components. Many other such models have been investigated, including the accelerated failure time model and more general transformation models. The concept of nonparametric efficiency also stemmed from early work on the efficiency of the Cox model compared with parametric submodels. In a second, related paper, Cox (1975) introduced the method of partial likelihood, a powerful approach to analyzing data from the Cox model, and this approach has been widely applied in many application areas. Research into martingale theory and counting processes was also largely stimulated by this work. The sequential view of a survival process implicit in the hazard function and the partial likelihood analysis was also instrumental in extensions to more general life history processes, including work on competing risks, recurrent events, and compartment models.

In this article, we review the 1972 Cox paper with a view to some of the ways it influenced further research and applications. The presentation is intended to be accessible to someone with a master's-level statistical background, or a person in a related discipline who would like to gain an understanding of the methods and their applications. However, we hope that the article will also

be of interest to those familiar with Cox regression, through worked examples and perhaps some new ways to think about old topics.

2. THE 1972 PAPER

2.1. The Continuous Model

The Cox model is a regression model for the hazard (or force of mortality) function. If $T > 0$ is the failure time and $\boldsymbol{Z}' = (Z_1, \ldots, Z_p)$ is a vector of p measured covariates, the Cox model specifies that the hazard function at time t is

$$\lambda(t; \boldsymbol{Z}) = \lim_{h \to 0^+} \mathrm{P}\{T \in [t, t+h) \mid T \geq t, \boldsymbol{Z}\}/h = \exp(\boldsymbol{Z}'\boldsymbol{\beta})\lambda_0(t), \quad t > 0, \qquad 1.$$

where $\boldsymbol{\beta}$ is a vector of regression parameters and $\lambda_0(t) = \lambda(t; \mathbf{0})$ is an unspecified baseline hazard function.

With time-independent covariates, the survivor function corresponding to Equation 1 is

$$\begin{aligned} S(t; \boldsymbol{Z}) &= \mathrm{P}(T \geq t \mid \boldsymbol{Z}) = \exp\left\{-\exp(\boldsymbol{Z}'\boldsymbol{\beta})\Lambda_0(t)\right\} \\ &= S_0(t)^{\exp(\boldsymbol{Z}'\boldsymbol{\beta})}, \quad t > 0, \end{aligned} \qquad 2.$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)\mathrm{d}u$ is the cumulative baseline hazard function and $S_0(t) = S(t; \mathbf{0})$ is the baseline survivor function.

Cox allowed the covariate vector to contain time-dependent elements, in which case the covariate vector is written $\boldsymbol{Z}(t)$. When $\boldsymbol{Z}$ comprises only fixed covariates, the model is one of proportional hazards, and is often referred to as the proportional hazards model in the literature. This name, however, is too restrictive, in that the extension to time-dependent covariates is important and leads to models where the hazards are typically nonproportional. We prefer the name Cox model or relative risk model to reflect this broader applicability. The latter term is descriptive in that $\exp(\boldsymbol{Z}'\boldsymbol{\beta})$ (or a time-dependent version of it) is referred to as a relative risk and describes the ratio of the hazard rate at time t with covariate value $\boldsymbol{Z}$ to that at $\boldsymbol{Z} = \mathbf{0}$. In a related idea, it is also possible to think of the regression parameter $\boldsymbol{\beta}$ as time dependent, but we focus on covariates.

In his paper, Cox considers two examples. In the first, there is a single covariate Z_1, which is a treatment indicator, and in the second, there is an additional time-dependent covariate, $Z_2(t) = Z_1 t$. This was used to provide a test of the simpler proportional hazards model—i.e., a test of $\beta_2 = 0$. This, or similar interactions with time, could also be used to describe a time-dependent relative risk. In this case, the log of the relative risk is $\boldsymbol{Z}(t)'\boldsymbol{\beta} = Z_1\beta_1 + Z_1 t\beta_2$, and relative to the baseline hazard, the relative risk in the treatment group is decreasing, constant, or increasing for β_2 respectively less than, equal to, or greater than 0. **Figure 1** illustrates the flexibility in a model for a single treatment indicator.

Time-dependent covariates arise frequently in this review, and their existence, like right censoring and left truncation, motivates the special study of failure time data. As in Kalbfleisch & Prentice (2002), time-dependent covariates fall into one of two classes:

- External covariates are external to the failure time process and might be functions of fixed covariates and time, like $Z_2(t)$ in the previous paragraph, or processes like weather or air pollution that take their values entirely separately from the failure time process.
- Internal covariates are internal to the mechanisms under consideration, like repeated measurements of blood pressure or white blood cell counts, and are potentially affected by factors that also affect failure.

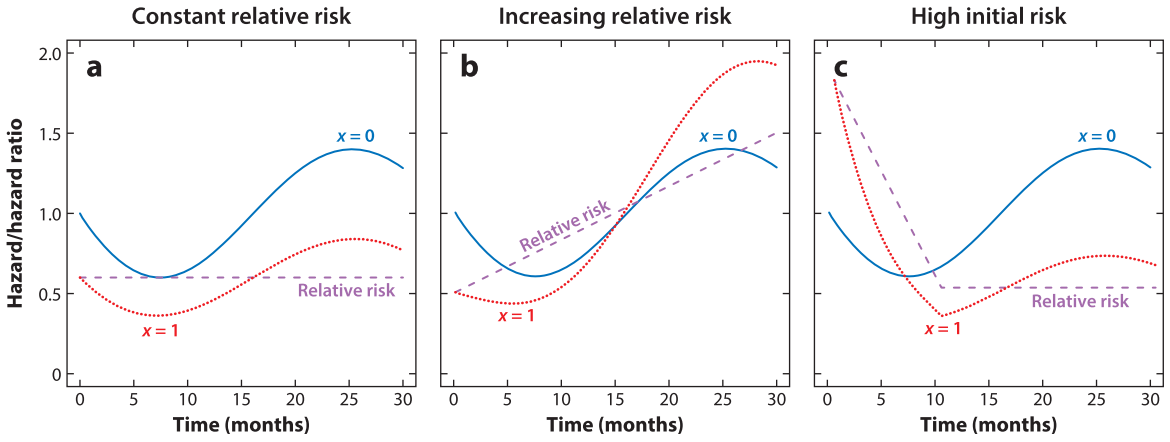

Figure 1

Hazard ratios in the Cox model with a single treatment indicator. The solid curve represents the baseline hazard, $\lambda_0(t)$ and the broken curve is the hazard for the treatment group. The curve marked relative risk is the relative risk function, which is constant in panel *a*, is increasing in panel *b*, and indicates an early elevated risk due to treatment followed by an ongoing benefit in panel *c*. Figure adapted with permission from Kalbfleisch & Prentice (2002).

In many instances, it is important to treat these two covariate types differently. For example, it is possible to define a survivor function

$$S(t) = \exp\left[-\int_0^t \exp\{\boldsymbol{Z}(u)'\boldsymbol{\beta}\}\lambda_0(u)\mathrm{d}u\right] \qquad 3.$$

when the covariates are external, but not when they are internal, since the observation of an internal covariate at time t typically requires the survival of the individual under study to time t. These distinctions are similar to those for endogenous and exogenous variables in economics.

For much of this article, we write $\boldsymbol{Z}$ as a vector of time-fixed covariates, although this can be replaced by $\boldsymbol{Z}(t)$, either internal or external, when considering hazard functions, which are our main focus. The only exception to this is the subdistribution hazard in the analysis of competing risks, where time-dependent variables must be external.

2.2. The Discrete Model

Cox also developed a discrete version of this model in which the failure time is restricted to times $u_1, u_2, \ldots$, where $u_0 = 0 < u_1 < u_2 < \cdots$. In this case, the hazard at u_i is $\lambda_i(\boldsymbol{Z}) = \mathrm{P}(T = u_i \mid T \geq u_i, \boldsymbol{Z})$, where

$$\frac{\lambda_i(\boldsymbol{Z})}{1-\lambda_i(\boldsymbol{Z})} = \exp(\boldsymbol{Z}'\boldsymbol{\beta})\frac{\lambda_{0i}}{1-\lambda_{0i}}, \quad i = 1, 2, \ldots, \qquad 4.$$

and where λ_{0i} is the discrete baseline hazard. This is a binary logistic model for the failure probability at each time u_i. The survivor function is

$$S(t; \boldsymbol{Z}) = \mathrm{P}(T \geq t \mid \boldsymbol{Z}) = \prod_{j|u_j<t}\{1-\lambda_j(\boldsymbol{Z})\}, \qquad t > 0. \qquad 5.$$

Equation 5 follows by noting that, to survive to any $t > 0$, the individual must survive each $u_j < t$ with probability $1 - \lambda_j(z)$. Cox used this model in order to extend the analysis he proposed to the discrete or mixed case. He also showed that the discrete and continuous cases can be elegantly combined using product integrals, as is reviewed in the next subsection.

2.3. Mathematical Aside: The Product Integral

Consider a continuous or discrete random failure time T and let $\Lambda(t)$ be its cumulative hazard function. In the continuous case, we have $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$, where $\lambda(t)$ is the hazard function, and in the discrete case,

$$\Lambda(t) = \sum_{j|u_j<t} \lambda_j,$$

where $\lambda_j = \mathrm{P}(t = u_j \mid T \geq t)$. We interpret $\mathrm{d}\Lambda(t)$ as $\lambda(t)\mathrm{d}t$ in the continuous case and as $\Lambda(t) - \Lambda(t^-)$ in the discrete case. Thus, in the discrete case, $\mathrm{d}\Lambda(u_j) = \lambda_j$ for $j = 1, 2, \ldots$ and $\mathrm{d}\Lambda(t) = 0$ elsewhere.

The survivor function for T can be written in either case as the product integral

$$S(t) = \mathcal{P}_0^t\{1 - \mathrm{d}\Lambda(u)\},$$

which reduces to $\exp\{-\Lambda(t))\}$ in the continuous case and to $\prod_{u_j<t}(1 - \lambda_j)$ in the discrete case. If T has both discrete and continuous parts, the product integral is still valid.

In a manner similar to the ordinary Stieltjes integral (Kalbfleisch & Prentice 2002, pp. 398–401), the product integral can be defined as a limit,

$$\mathcal{P}_0^t\{1 - \mathrm{d}\Lambda(u)\} = \lim_{N\to\infty} \prod_{j=1}^{N} \left[1 - \{\Lambda(x_j) - \Lambda(x_{j-1}\}\right],$$

where $x_0 = 0 < x_1 < \cdots < x_N = t$ and the limit is taken as $\max(x_j - x_{j-1}) \to 0$. This formulation illustrates that in both the discrete and continuous cases, survival to time t involves surviving a sequence of Bernoulli trials with varying probabilities of success. In many ways, this sequential view of survival data underlies the novelty of the Cox model and its analysis.

2.4. Estimation of $\boldsymbol{\beta}$ and $\Lambda_0(t)$

The analysis of the continuous model (Equation 1) is based on a conditional likelihood contribution at each failure time. Consider a sample of n individuals in which k individuals fail and the remainder are right-censored. The data can be specified as $(t_i, \delta_i, \boldsymbol{Z}_i), i = 1, \ldots, n$, where $\delta_i = 1$ or 0, respectively, indicates that the individual has failed or is right-censored at t_i. Without loss of generality, let i be the label of the individual who is the ith to fail ($i = 1, \ldots, k$) and suppose that the remaining individuals are censored. It follows that $\delta_i = 1$, $i = 1, \ldots, k$ and $\delta_i = 0$, $i = k + 1, \ldots, n$, and $t_1 < \cdots < t_k$ are the k ordered times of failure. Let $R(t) = \{\ell : t_\ell \geq t\}$ represent the risk set at time t^-. Provided that the hazard in Equation 1 applies to all individuals in the risk set, the probability that individual i fails at time t_i given those at risk, $R_i = R(t_i)$, and that a failure occurs at t_i, is

$$\frac{\lambda(t_i; \boldsymbol{Z}_i)}{\sum_{\ell\in R_i} \lambda(t_i; \boldsymbol{Z}_\ell)} = \frac{\exp(\boldsymbol{Z}_i'\boldsymbol{\beta})}{\sum_{\ell\in R_i} \exp(\boldsymbol{Z}_\ell'\boldsymbol{\beta})}, \qquad 6.$$

from which the baseline hazard has been eliminated. Cox argued that when $\lambda_0(t)$ is completely unknown, there is no information about $\boldsymbol{\beta}$ from intervals without failures, since any such intervals can be explained by taking $\lambda_0(t)$ near zero. There is a term like Equation 6 at each t_i ($i = 1, \ldots, k$), and the logarithm of the product across all failure times gives the log conditional likelihood,

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{k} \left[\boldsymbol{Z}_i'\boldsymbol{\beta} - \log\left\{ \sum_{\ell\in R_i} \exp(\boldsymbol{Z}_\ell\boldsymbol{\beta}) \right\} \right]. \qquad 7.$$

Let $S_0(\boldsymbol{\beta}) = \sum_{R_i} \exp(\boldsymbol{Z}'_\ell \boldsymbol{\beta})$, $\boldsymbol{S}_1(\boldsymbol{\beta}) = \sum_{R_i} \boldsymbol{Z}_\ell \exp(\boldsymbol{Z}'_\ell \boldsymbol{\beta})$, $\boldsymbol{S}_2(\boldsymbol{\beta}) = \sum_{R_i} \boldsymbol{Z}_\ell \boldsymbol{Z}'_\ell \exp(\boldsymbol{Z}'_\ell \boldsymbol{\beta})$. The score and observed information in Equation 7 are

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\beta}) = \sum_{i=1}^{k} \boldsymbol{Z}_i - S_1(\boldsymbol{\beta})/S_0(\boldsymbol{\beta}) \quad \text{and} \tag{8.}$$

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = -\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \log L(\boldsymbol{\beta}) = S_2(\boldsymbol{\beta})/S_0(\boldsymbol{\beta}) - S_1(\boldsymbol{\beta}) S_1(\boldsymbol{\beta})'/S_0(\boldsymbol{\beta})^2. \tag{9.}$$

Although each of the individual terms in Equation 6 is a legitimate conditional likelihood, their product has no simple probabilistic interpretation. However, Cox argued that the usual likelihood asymptotic properties would apply, so that the estimate of $\boldsymbol{\beta}$ would be approximately multivariate normal with a covariance matrix estimated by $\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}})^{-1}$, while noting that there would need to be assumptions on censoring and covariates to formalize the argument.

The same conditional argument was used to estimate $\boldsymbol{\beta}$ in the discrete logistic model (Equation 4), which in this case corresponds to a log odds rather than a log relative risk. Suppose that the m_i individuals in the set $D(t_i)$ fail at time t_i. The probability that those specific items fail at t_i, given the risk set and the fact that m_i individuals fail, again eliminates the hazard components. The corresponding conditional likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp(\boldsymbol{s}_i \boldsymbol{\beta})}{\sum_{\boldsymbol{\ell} \in R_{m_i}(t_i)} \exp(\boldsymbol{s}_\ell \boldsymbol{\beta})}, \tag{10.}$$

where $\boldsymbol{s}_i = \sum_{\ell \in D(t_i)} \boldsymbol{Z}_\ell$, $R_{m_i}(t_i)$ is the set of all subsets of size m_i from the risk set R_i, and $\boldsymbol{s}_{\boldsymbol{\ell}} = \sum_{j \in \boldsymbol{\ell}} \boldsymbol{Z}_j$. This likelihood has the same form as before, except that the denominator sum can contain very many terms when m_i is large. Thus, the computation can be difficult.

Cox used the discrete logistic model in order to incorporate ties in the partial likelihood as above and also to allow estimation of the underlying baseline hazard or the associated cumulative hazard, $\Lambda_0(t)$. He noted that, as with the Kaplan & Meier (1958) estimator, the maximum likelihood estimate would have mass points only at the observed failure times. At t_i, we have individuals in R_i at risk and suppose that individuals in $D(t_i)$ fail. Let λ_{ij} be the hazard for the jth individual at risk at time t_i. The likelihood contribution at that time would then be

$$L(\lambda_{0i}) = \prod_{j \in D(t_i)} \lambda_{ij} \prod_{j \in R_i - D(t_i)} (1 - \lambda_{ij}), \tag{11.}$$

where λ_{0_i} is the baseline hazard at t_i (see Equation 4). For a given $\boldsymbol{\beta}$, the maximum likelihood estimate, $\widehat{\lambda}_{0i}$, can be obtained analytically if there is a single value at t_i and from Newton's method more generally. The corresponding estimate of the cumulative hazard is

$$\widehat{\Lambda}_0(t) = \sum_{j \le t} \widehat{\lambda}_{0j}.$$

In his paper, Cox stressed several points:

- The model is more general than proportional hazards, and time-dependent variables are important.
- Fully parametric models might sometimes, or even often, be preferred.
- The efficiency of the Cox model compared with parametric models needs to be assessed.

2.5. Example 1: Analysis of Graft Failures Among Liver Transplant Patients

We analyzed time to graft failure, defined as the time between liver transplantation and the earlier of transplant failure or death. The study population ($n = 20{,}443$) was composed of adult patients

Table 1 Maximum likelihood estimates $\widehat{\beta}$ and standard errors for the Cox model with nonproportional hazards for time to graft failure among liver transplant patients

Covariate	Reference group	$\widehat{\beta}$	SE($\widehat{\beta}$)	exp($\widehat{\beta}$)	p-Value
Status 1 × $I(t \leq 30)$	Chronic liver disease ($t \leq 30$)	1.016	0.154	2.763	<0.001
Status 1× $I(t > 30)$	Chronic liver disease ($t > 30$)	−0.066	0.936	0.154	0.668
Age/5	Per 5-year increase	0.052	0.009	1.053	$<10^{-4}$
Years on waitlist	Per year	0.035	0.012	1.035	0.003
Female	Male	−0.094	0.037	0.91	0.011
Diabetic	Non-diabetic	0.168	0.037	1.183	$<10^{-4}$

I is the indicator function. Abbreviation: SE, standard error.

receiving a deceased-donor liver transplant between 2015 and 2019. The Cox model included pretransplant covariates listed in **Table 1**, plus others not listed for brevity. Patients in status 1, an indicator of acute liver failure, are given top priority in the US liver allocation rules, and status 1 is of primary interest. The relative risk, or hazard ratio (HR), compares transplanted status 1 patients to transplanted patients with chronic liver disease, having adjusted for all remaining covariates. Let Z_1 be the indicator of status 1 and let $\boldsymbol{Z}_2$ contain the remaining adjustment covariates. To investigate the HRs for $Z_1 = 1$ versus $Z_1 = 0$ over time, we fit a stratified model. For the strata defined by $Z_1 = j$, the hazard function is

$$\lambda(t; j, \boldsymbol{Z}_2) = \exp(\boldsymbol{Z}_2'\boldsymbol{\beta}_2)\lambda_{0j}(t), \quad j = 0, 1,$$

so that each stratum has its own baseline hazard function. From this model, estimates of the cumulative baseline hazards $\Lambda_{0j}(t)$ are obtained and plotted in **Figure 2*a***. The slope over an interval of the cumulative hazard gives an estimate of the average hazard over that interval. One can see that, after an initial period in which the failure rate is much higher for status 1, the hazards are similar for status 1 and chronic liver disease patients.

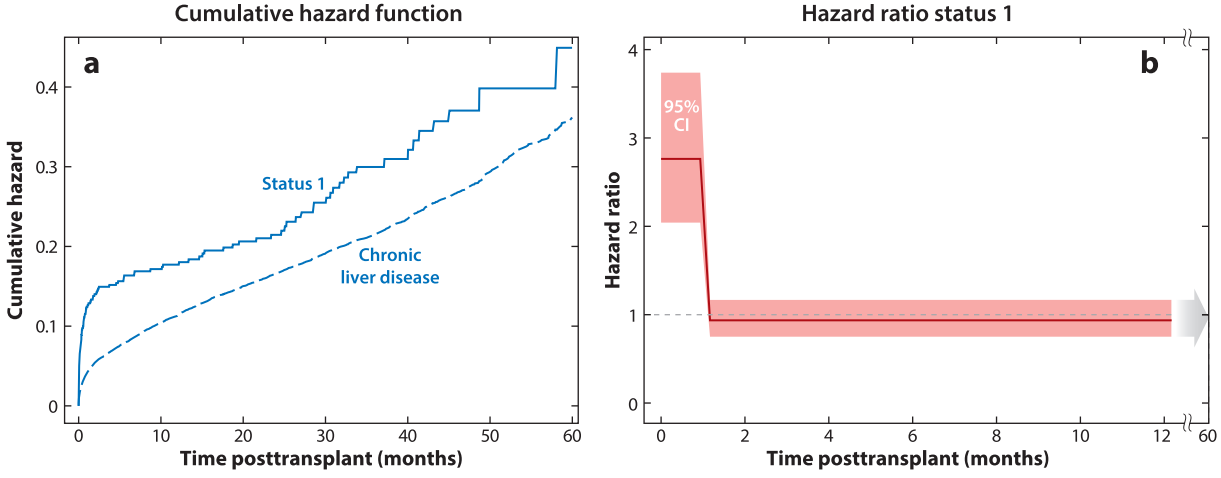

Figure 2

(*a*) Estimated baseline cumulative hazards for status 1 (*solid blue line*) and chronic liver disease (*broken blue line*) transplants. (*b*) HR (*solid red line*) and 95% CI (*light red field*) for status 1 versus chronic liver disease over follow-up time. Note that the HR point estimate and 95% CI lines extend out to 60 months. Abbreviations: CI, confidence interval; HR, hazard ratio.

In **Table 1**, we show the results of a model that allows the effect of status 1 to depend on time posttransplant. The fitted model has the form

$$\lambda(t) = \lambda_0(t) \exp\{Z_1 I(t \leq 30)\beta_0 + Z_1 I(t > 30)\beta_1 + \boldsymbol{Z}_2'\boldsymbol{\beta}_2\},$$

where $I(\cdot)$ is the indicator function and the first two covariates are time dependent. The parameter β_0 measures the effect of status 1 during the first 30 days posttransplant, and $\exp(\widehat{\beta}_0) = 2.76$ indicates a substantial and significant ($p < 0.001$) 176% increase in the graft failure rate for status 1 patients relative to otherwise similar transplanted patients with chronic liver disease. In contrast, $\widehat{\beta}_1$ indicates that the HR for status 1 after the 30 days is approximately 1 and nonsignificant ($p = 0.69$). These patterns are illustrated in **Figure 2**. **Table 1** also shows that there is a strong effect of age and diabetes on graft failure.

The preceding paragraph illustrates one method for evaluating the proportional hazards assumption (i.e., expand the model to include nonconstant effects, then test the pertinent parameters). There are other commonly-used methods to evaluate model assumptions. Proportionality with respect to a given covariate can be checked by stratifying on the covariate (assuming it to be categorical), then plotting the estimated log cumulative baseline hazards against time. Under proportional hazards, these plots should be approximately parallel and the degree to which they are not reflects the degree to which the proportionality assumption is violated. For discrete or continuous predictors, plots of the Schoenfeld (1982) residuals over time are sometimes useful, with patterns indicating lack of proportional hazards (e.g., Grambsch & Therneau 1994).

3. JUSTIFYING THE COX LIKELIHOOD

3.1. Marginal Likelihood and Maximum Likelihood

Several efforts were made to justify the likelihood given in Equation 7. In the uncensored case, Kalbfleisch & Prentice (1973) showed that the Cox likelihood corresponding to Equation 7 is the likelihood arising from the marginal distribution of the ranks based on continuous data from the model in Equation 1. They also suggested a censored data version of the rank statistic that gave rise to the likelihood more generally. This approach showed clearly why the score tests from Equation 7 gives rise to a rank statistic, but it did not allow incorporation of time-dependent covariates. Breslow (1974) assumed the baseline hazard to be a step function with discontinuities at the observed failure times and moved censoring times to the immediately preceding failure time. He showed that, in this model, the profile likelihood of $\boldsymbol{\beta}$ (maximizing over the baseline likelihood components) again gave the Cox likelihood, and this also applied to time-dependent covariates. This had the disadvantage, however, that the model itself was a function of the data, which seems unsatisfactory, at least from a theoretical point of view. A few years later, Bailey (1984) showed that nonparametric maximization of the full likelihood gave estimators of the regression coefficients and survival probabilities that were asymptotically equivalent to those described above.

None of these approaches yields the tied likelihood that Cox obtained from the discrete logistic model. The marginal likelihood approach involved breaking the tied values in all possible ways and summing the untied likelihood contributions. This reduces to the product over i of a sum over $m_i!$ cases at t_i, where m_i is the number of ties at t_i. This has very good properties for estimating $\boldsymbol{\beta}$ when ties are due to grouping the continuous model, but it can be computationally cumbersome. The Breslow approach led to a computationally simpler expression,

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\exp(s_i\boldsymbol{\beta})}{\{\sum_{\ell \in R_i} \exp(\boldsymbol{Z}_\ell'\boldsymbol{\beta})\}^{m_i}}, \qquad 12.$$

which replaces a sum over permutations without replacement by a sum over permutations with replacement. This can lead to considerable bias if the tied data arise from grouping the continuous model. Prentice & Kalbfleisch (2003) showed that the estimate from Equation 12 also arises in a martingale analysis of the discrete relative risk model,

$$\mathrm{P}(T = u_j \mid t \geq u_j, \boldsymbol{Z}) = \lambda_{0j} \exp(\boldsymbol{Z}'\boldsymbol{\beta}), \qquad 13.$$

and gave variance estimates that account for discreteness. Efron (1977) viewed Equation 12 as an approximation to the likelihood proposed in the marginal likelihood framework and suggested a better approximation.

In software for the Cox model, these four approaches to ties are referred to as logistic, exact, Breslow, and Efron. Unless the ties are numerous, these approaches are all quite similar. We recommend the use of the exact or the Efron approximation if the ties are numerous and one wishes to estimate $\boldsymbol{\beta}$ in the continuous model. If the data are essentially discrete, so that ties occur frequently and could occur at any observed time, the discrete relative risk model (Equation 13) may be more appropriate. In this case, the Breslow approximation is preferred, although the variance estimate should be corrected.

3.2. Partial Likelihood

The most satisfactory justification of the likelihood in Equation 7 was given in a second fundamental paper (Cox 1975), which defined partial likelihood and illustrated its application to the Cox model.

Suppose that the vector of response variables $\boldsymbol{Y}$ has a distribution that depends on two parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ is a nuisance parameter, typically of high dimension. Through a 1:1 transformation, $\boldsymbol{Y}$ can be rewritten as $\{\boldsymbol{A}_1, \boldsymbol{B}_1, \ldots, \boldsymbol{A}_m, \boldsymbol{B}_m\}$, where $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are random vectors, with $i = 1, \ldots, m$. Let $\boldsymbol{A}^{(j)} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_j)$ and $\boldsymbol{B}^{(j)} = (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_j)$, for $i = 1, \ldots, m$. Suppose further that the joint density of $(\boldsymbol{A}^{(m)}, \boldsymbol{B}^{(m)})$ can be written

$$\prod_{j=1}^{m} f(\boldsymbol{b}_j \mid \boldsymbol{a}^{(j-1)}, \boldsymbol{b}^{(j-1)}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \times \prod_{j=1}^{m} f(\boldsymbol{a}_j \mid \boldsymbol{a}^{(j-1)}, \boldsymbol{b}^{(j)}; \boldsymbol{\beta}), \qquad 14.$$

where $\boldsymbol{a}^{(0)} = \boldsymbol{b}^{(0)} = 0$. The second term in this expression is the partial likelihood, $L_P(\boldsymbol{\beta})$, based on $\{\boldsymbol{A}_j\}$ in the sequence $\{\boldsymbol{A}_j, \boldsymbol{B}_j\}$. In some instances, it can be argued that the ignored term in the likelihood has little or no information on $\boldsymbol{\beta}$ when $\boldsymbol{\alpha}$ is unknown. In others, the partial likelihood might be used because of its relative simplicity.

The corresponding score function is

$$\begin{aligned} U_P(\boldsymbol{\beta}) &= \sum_{j=1}^{m} \partial \log f(\boldsymbol{A}_j \mid \mathcal{H}_j; \boldsymbol{\beta})/\partial \boldsymbol{\beta} \\ &= \sum_{j=1}^{m} U_j(\boldsymbol{\beta}), \end{aligned}$$

say, where $\mathcal{H}_j = (\boldsymbol{A}^{(j-1)}, \boldsymbol{B}^{(j)})$ represents the history of the process. By conditioning on $\mathcal{H}_j$, Cox showed that under usual regularity conditions, $U_j(\boldsymbol{\beta})$ has a mean of zero and that $E\{U_i(\boldsymbol{\beta})U_j(\boldsymbol{\beta})'\} = \mathbf{0}$, for all j and $i < j$. Thus, the entries in the total score $U_P(\boldsymbol{\beta})$ are uncorrelated, and furthermore, $E\{\boldsymbol{I}_P(\boldsymbol{\beta})\} = \mathrm{var}\{U_P(\boldsymbol{\beta})\}$, where $\boldsymbol{I}_P(\boldsymbol{\beta})$ is the observed information based on the partial likelihood.

These are the building blocks in the theory of maximum likelihood estimation. Provided a central limit theorem applies to the total score and the regularity conditions of maximum likelihood apply, then the usual asymptotic results would hold and inference could be based on

score statistics, likelihood ratios, or maximum likelihood estimates with the usual asymptotic properties. Thus, for example, the partial likelihood estimator, $\widehat{\boldsymbol{\beta}}$, would be asymptotically normal, with a covariance matrix estimated by $\boldsymbol{I}_P(\widehat{\boldsymbol{\beta}})^{-1}$.

As applied to the continuous Cox model, A_j is the event that the individual j fails at time t_j, whereas $\boldsymbol{B}_j$ contains the information on censoring and covariates over the interval $(t_{(j-1)}, t_{(j)}]$ as well as the information that an individual fails at $t_{(j)}$. The application to the discrete case is the same, except the event A_j specifies which individuals fail at $t_{(j)}$ and $\boldsymbol{B}_j$ specifies all the previous observations as before and the event that exactly m_i individuals fail at $t_{(j)}$. The censoring must be independent so that the failure rate of the uncensored items at time t is as given in the model. This analysis allows time-dependent covariates of any type and requires only that the censoring be independent.

For independent censoring we require that surviving individuals at each time t must have failure rate $\lambda_0(t)\exp(\boldsymbol{Z}'\boldsymbol{\beta})$ as in the model. For this to hold, the probability an individual is censored at any given time t can depend on any variables in the history of the process up until this point or on external random mechanisms. This excludes anticipatory censoring, which occurs when individuals are censored when they are perceived to be at high or low risk of failure based on an internal time-dependent covariate, such as a measure of an individual's general health condition, that is not included in the model. In reliability applications, type II censoring is sometimes used, whereby individuals at risk are censored on the occurrence of the rth failure, where r is prespecified and all individuals are placed on study at the same time. This is also a form of independent censoring, as is any scheme of censoring depending on the number of previously observed failures of individuals under study. Some authors refer to this as noninformative censoring, noting the similarity to noninformative stopping rules. Independent censoring is a good choice of term since it notes the fact that at time t, and given the history up to that point, the failure and censoring outcomes are independent.

In these arguments, Cox had essentially shown that the score process from the Cox model is a discrete martingale with respect to the filtration defined by $\mathcal{H}_i$ and a martingale central limit theorem could provide the desired asymptotic results. Kalbfleisch & Prentice (2002, pp. 168–69) outline the argument. In this way, the partial likelihood idea was closely related to later work on martingales, which we discuss in Section 7.

4. EFFICIENCY OF THE PARTIAL LIKELIHOOD

As noted earlier, Cox raised the question of the efficiency of the partial likelihood compared with parametric models. Efron (1977) and Oakes (1977) presented different approaches to this question and arrived at similar conclusions.

Efron compared the expected (Fisher) information on $\boldsymbol{\beta}$ in the partial likelihood for the Cox model to that in a parametric submodel in which

$$\lambda_0(t) = \exp\{w_1(t)\alpha_1 + \cdots + w_q(t)\alpha_q\}, t > 0, \qquad 15.$$

where the $w_i(t)$ $(i = 1, \ldots, q)$ are specified functions, typically with $w_1(t) = 1, t > 0$. For example, a Weibull model is specified by $q = 2$, $w_1(t) = 1$, and $w_2(t) = \log t$. The parametric model becomes richer and broader in its coverage as q increases.

Efron's and Oakes's analyses yield straightforward calculation of the asymptotic relative efficiency of the Cox analysis given some particular parametric models. For example, in the two-sample case with no censoring and $\mathrm{P}(Z = 1) = \mathrm{P}(Z = 0) = 0.5$, the asymptotic relative efficiency at $\beta = 0$ is 1.0, but the efficiency falls off as $|\beta|$ increases. For the exponential model, the asymptotic relative efficiency at $\boldsymbol{\beta} = 1$ or -1 is approximately 82%, and for the Weibull model,

it is approximately 95%. Let $\boldsymbol{e}(\boldsymbol{\beta}, t)$ be the probability limit of $\boldsymbol{S}_1(\boldsymbol{\beta}, t)/S_0(\boldsymbol{\beta}, t)$ (see Equation 8). Efron showed that if linear combinations of the $w_i(t)$s for a given q closely approximate the components of $\boldsymbol{e}(\boldsymbol{\beta}, t), t \in (0, \tau]$, then the partial likelihood analysis would be nearly fully efficient for that $\boldsymbol{\beta}$. Begun et al. (1983) further developed these ideas to define nonparametric efficiency.

5. ALTERNATIVE MODELS

Overall the efficiency results are encouraging, but the loss of efficiency as $\boldsymbol{\beta}$ departs from $\mathbf{0}$ is not seen in parametric regression models. Furthermore, the Cox model is a regression model for the rate or intensity function and only indirectly specifies the relationship between the covariates and the outcome T.

On several occasions, Cox commented that there should be a larger role for parametric models (e.g., Cox 1972, Reid 1994). Parametric models, when they apply, provide a clear link between the covariates and the failure time. Parametric methods are discussed in several books, and good statistical software is available in SAS, R, and Stata. Log-linear models (i.e., linear in $\log T$) cover many standard cases and are also referred to as accelerated time to failure models, since the covariates act multiplicatively on T. Researchers have also considered a semiparametric accelerated time to failure model in which the distribution of the error term is unspecified. References here include the works of Prentice (1978), Louis (1981), Tsiatis (1990), Lin & Ying (1995), Kalbfleisch & Prentice (2002), Jin et al. (2003) and Jin et al. (2006). This work has developed computational methods so that the semiparametric accelerated time to failure model is a competitor to the Cox model. Time-dependent covariates in this model were introduced by Cox & Oakes (1984) and developed further by Lin & Ying (1995), although they are more difficult to interpret than in the Cox model. An extensive review of alternative failure time models is provided by Zeng & Lin (2007).

6. COMPETING RISKS

6.1. Background

In many instances, there may be more than one failure type involved in a time to failure model. It is typical to refer to these as causes of failure, although it should be kept in mind that in most applications, no causal relationship is under study. The data consist of the time to failure and the associated cause (or type) of failure.

The classical approach to competing risks postulates a multivariate survival model with a separate time to failure for each of cause of failure. This has been referred to as the alarm clock model in that the individual is assumed to have an alarm clock for each cause, and the individual fails when the first alarm rings. Thus, we observe the minimum time to failure among competing risks and the associated cause. This model is appealing in some ways and may have a strong physical basis in some engineering applications, but it bears little resemblance to real biological applications. Although Crowder (2001), Pintilie (2006), and others take this approach, we do not discuss this further. We instead consider analyses based on models for cause-specific hazards and subdistribution hazards.

As before, T represents the time to failure, and we let a random variable J, taking values in $\{1, \ldots, m\}$, index the m possible causes of failure. The cause-specific hazard function for the jth cause is

$$\lambda_j(t; \mathbf{Z}) = \lim_{h \to 0} \mathrm{P}\{t \le T < t + h, J = j \mid T \ge t, \mathbf{Z}\}, \quad t > 0, \quad j = 1, \ldots, m. \tag{16.}$$

The causes of failure are assumed to be mutually exclusive, so that only one cause of failure can occur. Thus, the total hazard is $\lambda(t; \mathbf{Z}) = \sum_{j=1}^{m} \lambda_j(t; \mathbf{Z})$, and the survivor function is

$$S(t; \mathbf{Z}) = \exp\{-\Lambda(t; \mathbf{Z})\} = \exp\left\{-\sum_{j=1}^{m} \Lambda_j(t; \mathbf{Z})\right\}, \quad t > 0, \tag{17.}$$

where $\Lambda_j(t; \mathbf{Z}) = \int_0^t \lambda_j(u; \mathbf{Z})\mathrm{d}u$ is the jth cumulative cause-specific hazard. It is important to note that $\exp\{-\Lambda_j(t)\}$ is not the survivor function corresponding to the jth cause, because the cause-specific hazard is the rate of failure in the presence of all potential causes of failure, and one cannot infer a distribution for the jth cause of failure in isolation. The causes of failure are mutually exclusive, but they are not assumed to be independent.

The cumulative incidence or subdistribution function for cause j is

$$F_j(t; \mathbf{Z}) = \mathrm{P}(T < t, J = j \mid \mathbf{Z}) = \int_0^t \lambda_j(u; \mathbf{Z}) S(u; \mathbf{Z})\mathrm{d}u, \quad t > 0, \quad j = 1 \ldots, m. \tag{18.}$$

This is easily seen since the integrand, $S(u; \mathbf{Z}) \times \lambda_j(u; \mathbf{Z})\mathrm{d}u$, is the probability of surviving to time u^- and then failing in $[u, u + \mathrm{d}u)$. Note that $1 - S(t; \mathbf{Z}) = \sum_{j=1}^{m} F_j(t; \mathbf{Z})$.

6.2. Cox Model for the Cause-Specific Hazard Function

Parametric or other semiparametric models could be used to model the cause-specific hazard in Equation 16, but as proposed by Prentice et al. (1978), it is common to consider Cox models

$$\lambda_j(t : \mathbf{Z}) = \lambda_{0j}(t) \exp(\mathbf{Z}'\boldsymbol{\beta}_j), \quad t > 0, \quad j = 1, \ldots, m. \tag{19.}$$

Here, each cause is assumed to have its own baseline hazard function, $\lambda_{0j}(t)$, and regression parameters, $\boldsymbol{\beta}_j$. Different components of $\mathbf{Z}$ in different failure types can be accommodated by appropriately assigning value zero to components of the $\boldsymbol{\beta}_j$s.

In this model, $\exp(\mathbf{Z}'\boldsymbol{\beta}_j)$ is the relative risk of a cause j failure for an individual with covariate vector $\mathbf{Z}$ compared with a standard with hazard $\lambda_{0j}(t)$. This is in terms of the cause j failure rate in the presence of all other risks, which is the failure rate actually observed.

The partial likelihood analysis of Section 3.2 extends to this case. Suppose that the failure at time t_i is a cause j_i failure and that censoring is independent. Then, conditioning on the risk set at t_i^- and the fact that a cause j_i failure occurs at time t_i, we obtain the partial likelihood contribution

$$L_{ij} = \frac{\exp(\mathbf{Z}_i'\boldsymbol{\beta}_{j_i})}{\sum_{\ell \in R_i} \exp(\mathbf{Z}_\ell'\boldsymbol{\beta}_{j_i})}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, m. \tag{20.}$$

The total partial likelihood is $\prod_{j=1}^{m} \prod_{i=1}^{k} L_{ij}$, which can be seen to factor into separate partial likelihoods for each cause. Furthermore, the partial likelihood for $\boldsymbol{\beta}_j$ is formally the same as the partial likelihood obtained if j were the only cause of failure and items that fail due to other causes were censored at their failure times. Thus, software to fit the Cox model can be used to estimate $\boldsymbol{\beta}_j$ for each j. The cumulative baseline cause-specific hazard, $\Lambda_{0j}(t)$, can also be estimated using the same arguments as before, again treating failure by other causes as censored. In addition, it is straightforward to obtain estimates of the baseline (or any) subdistribution function at a given $\mathbf{Z}$ using Equation 18.

This formal connection to calculation, taking other risks as censored, has led many users to think that the cause-specific hazard approach has an underlying assumption of independence between the risks, but this is not the case.

In this model, all risks are modeled simultaneously, and it is natural to refer to effects on the various causes of death. This model allows immediate extension to time-dependent covariates. Subdistribution functions are not useful with internal covariates. On the other hand, such covariates are often useful with cause-specific hazard functions in exploring causes of risks and can, in some instances, be used to evaluate dependence between risks as, for example, when an internal covariate is associated with two causes of failure (Prentice et al. 1978). In some instances, such internal covariates can be used to account for dependent censoring by inverse weighting (see, for example, Robins & Rotnitzky 1992, Robins & Finkelstein 2000).

6.3. Cox Models for the Subdistribution Hazard

The hazard function corresponding to $F_j(t; \boldsymbol{Z})$ is

$$\begin{aligned}\tilde{\lambda}_j(t; \boldsymbol{Z}) &= -\frac{\partial}{\partial t}\log\{1 - F_j(t; \boldsymbol{Z})\} \\ &= \lim_{h\to 0} P\{T \in [t, t+h), J = j \mid T \geq t \text{ or } T < t, J \neq j; \boldsymbol{\beta}\} \end{aligned} \quad 21.$$

and is called the subdistribution hazard for the jth cause. The conditioning event at time t gives rise to an associated unusual risk set comprising all individuals who have not failed and also individuals who failed by other causes prior to time t. Gray (1988) first defined the subdistribution hazard, and in a widely cited paper, Fine & Gray (1999) considered a Cox model for Equation 21,

$$\tilde{\lambda}_j(t; \boldsymbol{Z}) = \exp(\boldsymbol{Z}'\tilde{\boldsymbol{\beta}}_j)\tilde{\lambda}_0(t), \quad t > 0. \quad 22.$$

With time-independent covariates, this model is equivalent to a transformation model for the subdistribution function,

$$F_j(t; \boldsymbol{Z}) = F_0(t)^{\exp(\boldsymbol{Z}'\tilde{\boldsymbol{\beta}}_j)}, \quad t > 0. \quad 23.$$

For external time-dependent covariates, the model for the subdistribution function is typically complicated, and interpretation must be with reference to the subdistribution hazard. For internal time-dependent covariates, the subdistribution function is not defined and analysis based on Equation 22 is unavailable.

In the uncensored case, or in the censored case where times of censoring are known for items that fail, a partial likelihood based on Equation 22 is available by restricting information to the unusual subdistribution risk sets $\tilde{R}(t; j) = \{i : t_i \geq t \text{ or } t_i < t, J_i \neq j\}$, in which individuals who die of a cause other than j before time t are included. More general random censoring that is independent of $\boldsymbol{Z}$ can also be accommodated by inverse weighting, as described by Fine & Gray (1999). In contrast, the cause-specific hazard models (see Equation 19) provide valid inferences even if the censoring depends on $\boldsymbol{Z}$.

For the subdistribution hazards, each cause must be separately modeled, and the models are inconsistent with each other except in very special cases. The cause-specific hazard concentrates on examining and describing mechanisms leading to the failure patterns, whereas the subdistribution hazard concentrates on modeling the results without attention to the mechanisms involved. Estimates of the subdistribution can be important in anticipating resource requirements, for example, when one of the failure types requires hospitalization.

6.4. Example 2: Competing Risks Modeling of Chronic Liver Disease Patients

To illustrate the use of competing risks models, we studied $n = 51{,}092$ (age $\geq$18) patients waitlisted for deceased-donor liver transplantation during 2015–2019. Since there are not nearly enough

Table 2 Competing risks analysis of pretransplant liver disease patients: HRs (*p*-values)

Covariate	Cause-specific HR		Subdistribution HR	
	Death	Transplant	Death	Transplant
MELD score	1.09 (<0.001)	1.10 (<0.001)	1.01 (0.96)	1.08 (<0.001)
Age/5	1.25 (<0.001)	1.02 (<0.001)	1.21 (0.724)	0.99 (<0.001)
Black	0.87 (0.004)	0.96 (0.090)	0.92 (0.85)	0.99 (0.63)
Hispanic	0.98 (0.55)	0.94 (0.001)	1.03 (0.98)	0.93 (0.002)
Asian	0.82 (0.004)	0.95 (0.11)	0.84 (0.77)	0.96 (0.24)
Working	0.75 (<0.001)	1.16 (<0.001)	0.69 (0.83)	1.16 (<0.001)
Albumin	0.72 (<0.001)	1.04 (<0.001)	0.70 (0.77)	1.05 (<0.001)

Abbreviations: HR, hazard ratio; MELD, Model of End-stage Liver Disease.

donors relative to the number of patients in need of a liver transplant, medically suitable patients are registered on a transplant waiting list. In our analysis, we consider the two methods of leaving the waitlist (death and liver transplant) to be competing risks.

The data included 7,099 deaths prior to transplant and 27,311 liver transplants, which were 14% and 53% of the study population, respectively. We fitted cause-specific hazard models and subdistribution hazard models. Results for only a small subset of the covariates included in the two hazard models are reported in **Table 2**. The Model of End-stage Liver Disease (MELD) score is a significant predictor of the cause-specific hazard for both death (9% increase in death rates per unit increase in MELD score) and liver transplantation (10% increase). Interestingly, the effect of MELD on the subdistribution hazard of death is comfortably nonsignificant, with HR = 0.99. The increases in the cause-specific hazard for death (HR = 1.09) and liver transplant (HR = 1.10) cancel each other in evaluating the subdistribution hazards. The subdistribution HR for death is affected by the cause-specific hazards for both death and liver transplant, and the subdistribution risk sets do not account for the fact that, as MELD increases, patients are increasingly likely to be transplanted and hence not truly at risk for pretransplant death. The subdistribution HR for MELD is also attenuated (HR = 1.04 compared with HR = 1.10 for the cause-specific model). Liver transplantation has a greater impact on the death subdistribution hazard than death has on the liver transplantation subdistribution hazard, owing to the fact that approximately four times as many liver transplants as pretransplant deaths are observed.

Black patients and White patients are not significantly different with respect to the subdistribution HRs for both death and liver transplant. However, Black patients have a significantly lower cause-specific death hazard (HR = 0.87, $p = 0.4\%$). This reflects the fact that, having adjusted for other covariates, Black patients have a lower observed rate of death in the group of patients who are truly at risk for transplant. Differences like this could lead to incorrect conclusions on equity issues if the subdistribution hazards are used uncritically.

6.5. Comparison of Approaches and Recommendations

The cause-specific hazard approach has many advantages in terms of modeling, transparency, and flexibility. It also seems more appropriate in scientific investigations when interest focuses on

understanding mechanisms and not solely on the distribution of the final outcomes. In recent years, the subdistribution hazard has been widely viewed as the appropriate approach to competing risks. As we hope is apparent, when interest is on mechanisms, the subdistribution approach can miss important aspects of the data's genesis.

7. ANALYSIS OF RECURRENT EVENTS

The Cox model is also widely used in the analysis of recurrent events such as repeated hospitalizations or traffic tickets. Cook & Lawless (2007) provide detailed discussions and examples of recurrent events and their analysis. Other books with shorter discussions include those of Andersen et al. (1993), Kalbfleisch & Prentice (2002), and Martinussen & Scheike (2006).

Here, the outcome of interest involves observations on a right-continuous underlying counting process $\{\tilde{N}(t);\ t \geq 0\}$, which counts the number of events in $(0,\ t]$ with $\tilde{N}(0) = 0$ and $\tilde{N}(t) - \tilde{N}(t^-) \leq 1$ for all t. Observations are typically taken with a maximum observation time of τ, and individuals are subject to right censoring. Let $Y(t)$ be the left-continuous at-risk function, which takes the value 1 if the individual is under observation at time t^- and 0 otherwise. Let $N(t) = \int_0^t Y(u)\mathrm{d}\tilde{N}(u)$ be the observed counting process of events. The data on a sample of size n are

$$\{N_i(t), Y_i(t), \boldsymbol{Z}_i(t) : 0 \leq t \leq \tau, i = 1, \ldots, n\}. \qquad 24.$$

For this section, the covariate vector, $\boldsymbol{Z}(t)$, is allowed to have time-dependent components. The time to failure models are a special case in which $\tilde{N}_i(t) \leq 1$ for all i, t. For technical reasons, we assume that $\boldsymbol{Z}(t)$ is left-continuous with right-hand limits.

In what follows, we consider some aspects of martingale theory and define intensity models, which at each time condition on the whole history of the process, and rate models, which condition only on some aspects of the history.

7.1. Martingales and Intensity Processes for $N(t)$

A major advance came with the introduction of counting process notation and martingales (as in Aalen 1978) and the use of martingale theory to prove asymptotic properties (Andersen & Gill 1982) of the Cox partial likelihood in the analysis of both failure times and recurrent events.

Note that $\mathrm{d}\tilde{N}(t) = \tilde{N}(t) - \tilde{N}(t^-) = 1$ or 0, depending on whether or not an event occurs at time t. Let

$$\mathrm{d}\Lambda_i(t) = \mathrm{P}\{\mathrm{d}\tilde{N}_i(t) = 1 \mid \tilde{N}_i(u), \boldsymbol{Z}_i(u), 0 < u < t, i = 1, \ldots, n\}, \quad t > 0, \qquad 25.$$

so that the Cox model (Equation 1) is

$$\begin{aligned} \mathrm{d}\Lambda_i(t) &= \exp\{\boldsymbol{Z}(t)'\boldsymbol{\beta}\}\mathrm{d}\Lambda_0(t) \\ &= \exp\{\boldsymbol{Z}(t)'\boldsymbol{\beta}\}\lambda_0(t)\mathrm{d}t, \quad t > 0. \end{aligned} \qquad 26.$$

The expression on the right side of Equation 26 depends only on the current values $\boldsymbol{Z}(t)$. This is, however, without loss of generality, in that there is full flexibility in how $\boldsymbol{Z}(t)$ is defined. In most cases, event rates in a process will depend only on the previous events and covariates in that process, but the formulation in Equation 25 allows dependence on other processes as well.

For the observed data (Equation 24), the history (also called the filtration) up to and including time t is written as

$$\mathcal{F}_t = \sigma\{N_i(u), \boldsymbol{Z}_i^+(u), Y_i^+(u), i = 1, \ldots, n,\ \ 0 \leq u \leq t\}, \quad t > 0, \qquad 27.$$

where σ indicates the σ-field of events generated by the variables in the bracket. Note that

$$\mathcal{F}_{t^-} = \sigma\{N_i(u), \boldsymbol{Z}_i(u)Y_i(u), 1 = 1, \ldots, n, \quad 0 \le u < t\}, \quad t > 0, \tag{28.}$$

which specifies the information potentially available to an observer just prior to time t. Because of the left continuity of $\boldsymbol{Z}(t)$ and $Y(t)$, $\mathcal{F}_{t^-}$ includes $\boldsymbol{Z}_i(t)$ and $Y_i(t)$, but not $N_i(t)$, $i = 1, \ldots, n$. A (complete) intensity model for $\{N_i(t)\}$ is

$$\mathrm{P}\{\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t^-}\} = Y_i(t)\mathrm{d}\Lambda_i(t), \quad 0 \le t \le \tau, \tag{29.}$$

which, by comparison with Equation 25, characterizes independent censoring. Thus, individuals at risk at time t fail at the same rate as specified in the model. Rules for censoring at time t can depend on any variables in $\mathcal{F}_t$.

The intensity models described above are very flexible. For example, a component of $\boldsymbol{Z}(t)$ could be $\boldsymbol{Z}_1(t) = N(t^-)$, which allows the rate to depend on the number of previous failures. Similarly, one could include information on the number of failures other individuals have experienced and model various kinds of interdependencies of the recurrent events. Such models can be very useful in understanding how components of the process interact. Cox (1973) (in a sister paper to Cox 1972) considers models for dependencies in point processes. In another sense, these intensity models are not flexible at all, since given $\mathcal{F}_{t^-}$, the probability for $\mathrm{d}N(t)$ is determined and may depend on many features in the history that we may want to ignore.

7.2. Martingales and Counting Processes

Suppose that censoring is independent, and consider the process

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\mathrm{d}\Lambda_i(u), \quad i = 1, \ldots, n, \quad t \ge 0, \tag{30.}$$

or equivalently,

$$\mathrm{d}M_i(t) = \mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda_i(t). \tag{31.}$$

From Equation 29 and by conditioning on $\mathcal{F}_{t^-}$, we can see that both Equation 30 and Equation 31 have an expected value of 0. From Equation 30, $M_i(t)$ satisfies

$$\mathrm{E}[M_i(t) \mid \mathcal{F}_{t^-}] = 0, \quad t \ge 0, \quad \mathrm{E}\{M_i(t) \mid \mathcal{F}_s\} = M_i(s), \quad 0 \le s < t. \tag{32.}$$

A process that satisfies these two conditions is a mean-zero martingale with respect to the filtration $\mathcal{F}_{t^-}$.

Equation 31 gives a decomposition of increments in the counting process into signal plus noise: $\mathrm{d}N_i(t) = Y_i(t)\mathrm{d}\Lambda_i(t) + \mathrm{d}M_i(t)$, where the noise term, $\mathrm{d}M_i(t)$ is the random increment in the martingale. Estimates of $\mathrm{d}M_i(t)$ can therefore be viewed as residuals. These are termed the martingale residuals and sometimes used in model checking (Therneau et al. 1990). The partial likelihood argument can be extended to the recurrent event setting, and the corresponding score function (Equation 8) on data up to time t can be written

$$\boldsymbol{U}(t; \boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i(u) - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, u)\}\mathrm{d}N_i(t), \tag{33.}$$

where

$$\boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, u) = \sum_{j=1}^{n} \boldsymbol{Z}_j(u) \frac{Y_j(u)\exp\{\boldsymbol{Z}_j(u)'\boldsymbol{\beta}\}}{\sum_{\ell=1}^{n} Y_\ell(u)\exp\{\boldsymbol{Z}_\ell(u)'\boldsymbol{\beta}\}}$$

is the weighted average of $\boldsymbol{Z}_j(u)$ in the risk set at time u. Equation 33 is simply a rewriting of Equation 8. It can also be shown that $\mathrm{d}N_i(t)$ in Equation 33 can be replaced with the $\mathrm{d}M_i(t)$ to give

$$\boldsymbol{U}(t;\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i(u) - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, u)\}\mathrm{d}M_i(t), \qquad 34.$$

which can be seen to be the sum of orthogonal mean-zero martingales.

A primary result of interest arises from the central limit theorem of Rebolledo (1980), which under some regularity conditions establishes that the total score vector in Equation 34, $\boldsymbol{U}(\tau;\boldsymbol{\beta})$, is asymptotically normal with a covariance matrix estimated by the second derivative of the log partial likelihood. This then yields all the usual results associated with asymptotic likelihood theory. The interested reader is referred to Andersen & Gill (1982), Andersen et al. (1993), Fleming & Harrington (1991), or Kalbfleisch & Prentice (2002).

7.3. Marginal Rate Models for $\tilde{N}(t)$

The assumptions in the previous subsection are very strong, since the model must describe the full effect of the past on the current rate of events. In most instances, one would like to relax this to a simpler marginal model with an average rate function

$$\mathrm{d}\Lambda_i^*(t) = \mathrm{P}\{\mathrm{d}\tilde{N}_i(t) = 1 \mid \boldsymbol{Z}_i(t)\}, \quad t > 0, \qquad 35.$$

which conditions only on the elements in $\boldsymbol{Z}_i(t)$ and not otherwise on the history of the process. Again, a Cox model for Equation 35 gives

$$\mathrm{d}\Lambda_i^*(t) = \exp\{\boldsymbol{Z}_i(t)\boldsymbol{\beta}\}\mathrm{d}\Lambda_0^*(t), \quad t > 0. \qquad 36.$$

If the covariate vector $\boldsymbol{Z}_i$ is not time dependent, then integrating Equation 36 gives the proportional means model,

$$\begin{aligned}\Lambda_i^*(t) &= \mathrm{E}\{\tilde{N}_i(t) \mid \boldsymbol{Z}_i\} \\ &= \Lambda_0^*(t)\exp(\boldsymbol{Z}_i\boldsymbol{\beta}), \quad t > 0.\end{aligned}$$

When $\boldsymbol{Z}_i(t)$ is an external time-dependent covariate, the model for the mean process still exists but is not so simple. When $\boldsymbol{Z}_i(t)$ is internal, this formulation in terms of a mean process model is not available, and interpretation must be in terms of the effects on the marginal rates, as in Equation 35.

Marginal rate models were first considered by Lawless & Nadeau (1995), who gave a number of examples and established asymptotic results in discrete time. Lin et al. (2000) gave more formal mathematical formulations and established asymptotic results in the continuous case. In general, even with fixed covariate vector, Equation 36 does not define a full probability model for $\{N_i(t)\}$ given $\boldsymbol{Z}_i$. Only the intensity function does that. The data will usually be subject to right censoring. We require that the censoring variable depend only on the elements of $\boldsymbol{Z}(t)$ in the model. This is a type of independent censoring, but adapted to this marginal-type model.

Since we lack the necessary complete history or filtration, methods based on martingales no longer apply. To make headway with this type of model, we need another framework, and it is common to assume that the observed variables, $\{N_i(t), Y_i(t), \boldsymbol{Z}_i(t), 0 \le t \le \tau\}$, are independent and identically distributed for $i = 1, \ldots, n$. Further analysis and inference will depend on this assumption, and the mathematical methods are based on modern empirical process theory, for which Kosorok (2008) is an excellent reference.

The score component corresponding to each failure time has mean zero, and therefore the total score from the intensity models,

$$\boldsymbol{U}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i(u) - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, u)\} \mathrm{d}N_i(u), \tag{37.}$$

still has mean zero and so is an unbiased estimating function. Furthermore, we can replace $N_i(u)$ with $M_i(u) = N_i(u) - \int_0^t Y_i(u) \exp\{\boldsymbol{Z}_i(u)'\boldsymbol{\beta}\} \mathrm{d}\Lambda_0^*(u)$, just as in the complete intensity case, to obtain

$$\boldsymbol{U}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i(u) - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, u)\} \mathrm{d}M_i(u), \quad 0 \le t \le \tau. \tag{38.}$$

Thus, the corresponding estimate is $\widehat{\boldsymbol{\beta}}$, which is formally the same estimate as would be obtained in the complete intensity case. Lin et al. (2000) show that $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}_0, t)$ is asymptotically a p-variate normal process with mean zero and a covariance function that can be estimated by

$$\widehat{\Sigma}(t) = n^{-1} \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i(u) - \boldsymbol{\mathcal{E}}(\widehat{\boldsymbol{\beta}}, u)\} \mathrm{d}\widehat{M}_i(u) \int_0^t \{\boldsymbol{Z}_i(v) - \boldsymbol{\mathcal{E}}(\widehat{\boldsymbol{\beta}}, v)\}' \mathrm{d}\widehat{M}_i(v). \tag{39.}$$

If the observation window is $(0, \tau]$, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is normal with mean zero and covariance matrix of the sandwich type estimated by

$$\mathcal{I}(\widehat{\boldsymbol{\beta}}, \tau)^{-1} \Sigma(\widehat{\boldsymbol{\beta}}, \tau) \ \mathcal{I}(\widehat{\boldsymbol{\beta}}, \tau)^{-1}, \tag{40.}$$

where, as before, $\mathcal{I}$ is the negative of the matrix of derivatives of the estimating function, $U(\boldsymbol{\beta}, \tau)$. In this setup, $\Lambda_0^*(t)$ is estimated using

$$\widehat{\Lambda}_0^*(t) = \int_0^t \left[\sum_{i=1}^{n} Y_i(u) \exp\{\boldsymbol{Z}_i'(u)\boldsymbol{\beta}\} \right]^{-1} \mathrm{d}N_\cdot(u), \tag{41.}$$

where $N_\cdot(u) = \sum N_i(u)$. This is analogous to the Breslow estimator in Equation 12. This important result shows that the very strong assumptions underlying the complete intensity model can be greatly relaxed by using a marginal rate model with a more general variance estimate. The robust variance in Equation 40 is an option in software packages and should be used routinely in Cox model analyses.

7.4. Example 3: Modeling Emergency Department Visits

We studied emergency department visits among patients receiving a deceased-donor liver transplant between 2008 and 2017 at the University of Pennsylvania ($n = 932$). Patients begin follow-up at discharge from the hospital following transplantation and end follow-up at the earliest of death, transplant failure, five years post transplant, and the end of the observation period (December 31, 2017).

Results of our analysis are presented in **Table 3**. In total, there were 1,037 emergency department visits among the 932 patients. We fitted a proportional rates model (**Table 3**, column 1), followed by four models that adjusted progressively for the event history (columns 2–5). The proportional intensity model (column 2) yields the same parameter estimates as the proportional rates model (column 1). However, unlike the intensity model, the proportional rates model uses a robust variance estimator in order to account for the correlation between emergency department visits within each patient. The intensity model uses a martingale-based variance estimator and gives smaller, and typically incorrect, standard errors. Taking Black race as an example, under the proportional rates model, Black race is associated with a significant 59% increase in

Table 3 Estimates and SEs in recurrent event models of post–liver transplant emergency department visits

Covariate	Marginal rates model RR [SE]	Intensity model IR [SE]	Intensity model IR [SE]	Intensity model IR [SE]	Intensity model IR [SE]
Black race	1.59 [0.173]	1.59 [0.091]	1.44 [0.082]	1.42 [0.082]	1.41 [0.081]
Year of transplant	1.17 [0.019]	1.17 [0.012]	1.08 [0.013]	1.11 [0.012]	1.08 [0.012]
Cold time (hours)	1.022 [0.021]	1.022 [0.006]	1.015 [0.006]	1.015 [0.006]	1.015 [0.006]
Years on waitlist	1.051 [0.021]	1.051 [0.012]	1.028 [0.012]	1.024 [0.012]	1.019 [0.012]
Hemodialysis after transplant	1.81 [0.268]	1.81 [0.131]	1.61 [0.116]	1.41 [0.104]	1.34 [0.099]
Length of stay for transplant	1.014 [0.004]	1.014 [0.002]	1.011 [0.003]	1.011 [0.003]	1.008 [0.003]
$I\{N(t^-) > 0\}$	NA NA	NA NA	3.79 [0.227]	NA NA	NA NA
$N(t^-)$	NA NA	NA NA	NA NA	1.27 [0.11]	1.72 [0.043]
$N(t^-)^2$	NA NA	NA NA	NA NA	NA NA	0.97 [0.002]

The intensity models in columns 2–5 are distinguished by the variables each model conditions on as indicated in the table. SEs are shown in brackets. Abbreviations: IR, intensity ratio; NA, not applicable; RR, rate ratio, SE, standard error.

the rate of emergency department visits, relative to Whites (the reference category). Adjusting for a 0/1 time-dependent indicator of whether or not a prior emergency department visit had occurred, the Black effect decreases to an intensity ratio of 1.44. Examination of the last two columns of **Table 4** reveals that greater adjustment for event history results in further attenuation of the estimated effect of Black race (i.e., decreasing intensity ratios). This phenomenon is also present to a greater (e.g., hemodialysis after transplant) or lesser (e.g., cold ischemia time) extent for the remaining covariates. Adjustment for the event history provides information regarding the association among the events within-patient. For instance, the rate of emergency department visits is 3.79 times as great after a patient has experienced their first post–liver transplantation visit (column 2). Each emergency department visits increases the rate of the next emergency department visits by about 27% (column 3).

8. EXTENSIONS OF THE COX MODEL

In the above sections, we covered the application of Cox regression to survival data and methods for accommodating recurrent events. In a data structure frequently encountered in medical and epidemiological studies, the recurrent event sequence (e.g., hospitalizations) may be stopped by a terminal event (e.g., death). This structure has often been labeled recurrent/terminal event data. Cook & Lawless (1997), in one of the first papers to map out potential methodology in the recurrent/terminal arena, initiated a plethora of associated research in the decade that followed. The Cox-type model (nonparametric baseline, multiplicative covariate effects) has been used in many recurrent/terminal event methods. For example, Ghosh & Lin (2002) proposed a model for $\mu(t) = \mathrm{E}\{N(t) \mid \boldsymbol{Z}\} = \mu_0(t)\exp(\boldsymbol{Z}'\boldsymbol{\beta})$, where $\mu(t)$ is the marginal mean number of events averaging over the death distribution. Various forms of conditional modeling have also been proposed. For example, Ye et al. (2007) proposed a two-stage approach for the models $\mathrm{d}R(t) = \mathrm{E}\{\mathrm{d}N(t) \mid \boldsymbol{Z}, Q, D > t\}$ and $\lambda(t \mid \boldsymbol{Z}, Q)$, where Q is a common gamma frailty connecting

the two rates. Cox-type models were used for both $dR(t)$ and $\lambda(t)$. In the recurrent/terminal event setting, marginal methods are mostly of interest for public-health-related interpretations, whereas modeling the terminal event hazard and recurrent event rate given survival is best suited to studies of mechanisms. This is somewhat analogous to the competing risks setup.

Another important extension of Cox regression has been to the development of cost-efficient sampling designs. In many epidemiological studies, the collection and/or measurement of covariates can be very costly in terms of time and resources. For example, in-person interviews, collection and subsequent analysis of blood samples, and medical chart reviews are cost intensive. For the study of outcomes that occur rather infrequently, the collection of covariate data on all subjects seems unnecessary. This is especially true when Cox regression is considered for the analysis. A heuristic motivation for partial likelihood itself is that most of the information on the regression parameter lies at the death times. With this in mind, it is not surprising that Cox regression extends so readily to cost-efficient sampling designs. Prentice (1986) proposed the case-cohort design, wherein covariate information is collected on all deaths and a randomly chosen subcohort of subjects. A pseudo–partial likelihood procedure was proposed, with a score function of the same form as in Cox regression,

$$\sum_{i=1}^{n} \int_0^{\tau} \{\boldsymbol{Z}_i - \widehat{\boldsymbol{\mathcal{E}}}(\boldsymbol{\beta}, t)\} \mathrm{d}N_i(t), \qquad 42.$$

but with the full-data $\boldsymbol{\mathcal{E}}(\boldsymbol{\beta}, t)$ replaced with $\widehat{\boldsymbol{\mathcal{E}}}(\boldsymbol{\beta}, t)$, which is based on the subject who died at time t plus the at-risk subjects from the subcohort. Self & Prentice (1988) proved that the regression parameter estimator was consistent and asymptotically normal. Chen & Lo (1999) later proposed case-cohort estimators that offered improved efficiency over the original Prentice (1986) estimator, provided that the total number of cases and controls in the full cohort was known. Another frequently adopted subsampling technique is the nested case-control study (see Lubin & Gail 1984, Goldstein & Langholz 1992). Under a nested case-control design, subjects are sampled at the time of each case's death. Chen (2001) subsequently proposed generalized case-cohort sampling and related Cox-type estimators. The generalized case-cohort class encompassed the case-cohort, the nested case control, and even the traditional case-control sampling designs, all through weighted estimating functions. Much recent attention has been given to two-stage sampling schemes, in which data are collected on a cohort for all outcomes and covariates except for one or more expensive covariates. The outcomes observed in the first stage are used to guide selection at the second stage. Tao et al. (2020) investigated optimal designs and gave an extensive bibliography.

9. DISCUSSION

Cox regression has had an enormous impact, both methodologically and in biomedical applications. The Cox model has found important applications in many disciplines outside the health sciences, including economics, sociology, and political science. The Cox model and methods are now so ingrained in scientific research that alternative approaches typically have to be accompanied with justification for not using the standard.

Here, we sought to capture some important aspects of the Cox model and to demonstrate its application to a variety of data structures. Space limitations prevent us from covering many interesting areas involving the Cox model, certainly among them being causal inference. Martinussen (2022) gives an enlightening treatment of the topic that highlights an apparent conflict with intuition, as the widely used relative risk in the Cox model does not satisfy the conditions for a causal parameter.

Over the past 50 years, the Cox model has been intrinsic to research in many application areas and has stimulated much methodological development leading to important new areas of investigation. It is now a workhorse of applied statistics, like generalized linear models or the analysis of variance. It is plausible and maybe even likely that at its centennial in 2072, the Cox model will still be a force in statistics and data science.

SUMMARY POINTS

1. Cox's original paper was an inspired combination of innovation and intuition. The intuition was formalized in his 1975 paper.
2. Failure time problems arise in many areas, and the work has found applications in many disciplines.
3. Competing risks exemplify the use of the Cox model. Models based on cause-specific hazards are most useful in understanding the mechanisms of the process, whereas models based on the subdistribution function focus on final outcomes. Approaches should be chosen depending on the questions being posed.
4. Marginal rate models greatly relax the strong assumptions in intensity models. The robust variance estimate should be routinely used in applications.
5. The partial likelihood approach of examining outcomes from a risk set at each event time has led to numerous extensions to cost-efficient sampling designs.
6. Although the Cox model has become part of the scientific fabric, the relative risk parameter, so often relied upon, is not a causal parameter as currently formulated in causal inference. This conflict requires some resolution.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health grant R01-DK070869. The authors wish to thank the United Network for Organ Sharing (UNOS) for access to the UNOS data, as well as Dr. Tess Bittermann for access to the University of Pennsylvania liver transplant data.

LITERATURE CITED

Aalen O. 1978. Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Stat.* 6:534–45

Andersen PK, Borgan Ø, Gill R, Keiding N. 1993. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag

Andersen PK, Gill R. 1982. Cox's regression model for counting processes: a large sample study. *Ann. Stat.* 10:1100–20

Bailey K. 1984. Asymptotic equivalence between the Cox estimator and the general M-estimators of regression and survival parameters in the Cox model. *Ann. Stat.* 12:730–36

Begun J, Hall W, Huang W-M, Wellner J. 1983. Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Stat.* 11:432–52

Breslow N. 1974. Covariance analysis of censored survival data. *Biometrics* 30:89–99
Chen K. 2001. Generalized case-cohort sampling. *J. R. Stat. Soc. Ser. B* 63:791–809
Chen K, Lo S-H. 1999. Case-cohort and case-control analysis with Cox's model. *Biometrika* 86:755–64
Cook RJ, Lawless JF. 1997. Marginal analysis of recurrent events and a terminating event. *Stat. Med.* 16:911–24
Cook RJ, Lawless JF. 2007. *The Statistical Analysis of Recurrent Events*. New York: Springer-Verlag
Cox DR. 1972. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* 34:187–220
Cox DR. 1973. The statistical analysis of dependencies in point processes. In *Symposium on Point Processes*, ed. PAW Lewis, pp. 55–66. New York: Wiley
Cox DR. 1975. Partial likelihood. *Biometrika* 62:269–76
Cox DR, Oakes D. 1984. *The Analysis of Survival Data*. London: Chapman and Hall
Crowder M. 2001. *Classical Competing Risks*. London: Chapman and Hall/CRC
Efron B. 1977. The efficiency of Cox's likelihood function for censored data. *J. Am. Stat. Assoc.* 72:557–65
Feigl P, Zelen M. 1965. The efficiency of Cox's likelihood function for censored data. *Biometrics* 21:826–38
Fine JP, Gray RJ. 1999. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* 94:496–509
Fleming T, Harrington D. 1991. *Counting Processes and Survival Analysis*. New York: Wiley
Ghosh D, Lin DY. 2002. Marginal regression models for recurrent and terminal events. *Stat. Sin.* 12:663–88
Glasser M. 1967. Exponential survival with covariance. *J. Am. Stat. Assoc.* 62:561–68
Goldstein L, Langholz B. 1992. Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann. Stat.* 20:1903–28
Grambsch P, Therneau T. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81:515–26
Gray RJ. 1988. A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Stat.* 16:1141–54
Jin Z, Lin DY, Wei LJ, Ying Z. 2003. Rank-based inference for the accelerated failure time model. *Biometrika* 90:341–53
Jin Z, Lin DY, Ying Z. 2006. On least squares regression with censored data. *Biometrika* 93:141–61
Kalbfleisch JD, Prentice RL. 1973. Marginal likelihoods based on Cox's regression and life model. *Biometrika* 60:267–78
Kalbfleisch JD, Prentice RL. 2002. *The Statistical Analysis of Failure Time Data*. New York: Wiley. 2nd ed.
Kaplan E, Meier P. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–81
Kosorok M. 2008. *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer
Lawless JF, Nadeau C. 1995. Some simple robust methods for the analysis of recurrent events. *Technometrics* 37:158–68
Lin DY, Wei LJ, Yang I, Ying Z. 2000. Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B* 62:711–30
Lin DY, Ying Z. 1995. Semiparametric inference for the accelerated life model with time dependent covariates. *J. Stat. Plan. Inference* 44:47–63
Louis TA. 1981. Nonparametric analysis of an accelerated failure time model. *Biometrika* 68:381–90
Lubin JH, Gail MH. 1984. Biased selection of controls for case-control analyses of cohort studies. *Biometrics* 40:63–75
Mantel N. 1963. Chi-square tests with one degree of freedom; extensions of the Mantel–Haenszel procedure. *J. Am. Stat. Assoc.* 58:690–700
Mantel N, Haenszel W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* 22:719–48
Martinussen T. 2022. Causality and the Cox regression model. *Annu. Rev. Stat. Appl.* 9:249–59
Martinussen T, Scheike T. 2006. *Dynamic Regression Models for Survival Data*. New York: Springer
Oakes D. 1977. The asymptotic information in censored survival data. *Biometrika* 64:487–93
Peto R, Peto J. 1972. Asymptotically efficient rank invariant test procedures. *J. R. Stat. Soc. Ser. A* 135:185–207
Pintilie M. 2006. *Competing Risks: A Practical Perspective*. New York: Wiley
Prentice RL. 1978. Linear rank tests with right censored data. *Biometrika* 65:167–79

Prentice RL. 1986. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73:1–11

Prentice RL, Kalbfleisch J. 2003. Mixed discrete and continuous Cox regression model. *Lifetime Data Anal.* 9:195–210

Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. 1978. The analysis of failure time in the presence of competing risks. *Biometrics* 34:541–54

Rebolledo R. 1980. Central limit theorems for local martingales. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* 51:269–86

Reid NR. 1994. A conversation with Sir David Cox. *Stat. Sci.* 9:439–55

Robins J, Finkelstein D. 2000. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 56:779–88

Robins J, Rotnitzky A. 1992. A recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*, ed. NP Jewell, K Dietz, VT Farewell, pp. 297–331. Boston: Birkhäuser

Schoenfeld D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika* 69:239–41

Self S, Prentice RL. 1988. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Stat.* 16:64–81

Tao R, Zeng D, Lin DY. 1984. Optimal designs of two-phase studies. *J. Am. Stat. Assoc.* 115:1946–59

Therneau T, Grambsch P, Fleming T. 1990. Martingale-based residuals for survival models. *Biometrika* 77:147–60

Tsiatis A. 1990. Estimating regression parameters using linear rank tests for censored data. *Ann. Stat.* 18:354–72

Ye Y, Kalbfleisch J, Schaubel D. 2007. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 63:78–87

Zeng D, Lin DY. 2007. Efficient estimation for the accelerated failure time model. *J. Am. Stat. Assoc.* 102:1387–96