Lifetime Data Anal (2007) 13:471–480
DOI 10.1007/s10985-007-9048-y

On the Breslow estimator

D. Y. Lin

Received: 5 April 2007 / Accepted: 16 July 2007 / Published online: 2 September 2007

Abstract In his discussion of Cox's (1972) paper on proportional hazards regression, Breslow (1972) provided the maximum likelihood estimator for the cumulative baseline hazard function. This estimator is commonly used in practice. The estimator has also been highly valuable in the further development of Cox regression and semiparametric inference with censored data. The present paper describes the Breslow estimator and its tremendous impact on the theory and practice of survival analysis.

Keywords Cox model · Maximum likelihood · Partial likelihood · Proportional hazards · Semiparametric inference · Survival data

1 Introduction

In 1972, Sir David Cox read the paper "Regression models and life tables" to the Royal Statistical Society. In this seminal paper, Cox (1972) presented the proportional hazards model, which specifies that the conditional hazard function of failure time given a set of covariates is the product of an unknown baseline hazard function and an exponential regression function of covariates. Because the baseline hazard function is arbitrary, standard parametric likelihood does not apply. Cox (1972) suggested to estimate the regression parameters by a "conditional likelihood," which does not involve the nuisance baseline hazard function.

In the discussion of Cox's paper, several people questioned Cox's conditional likelihood. Kalbfleisch and Prentice (1972) pointed out that it is rather the marginal likelihood of the ranks. Breslow (1972) suggested an alternative estimation approach which

D. Y. Lin (✉)
Department of Biostatistics, University of North Carolina, CB#7420, Chapel Hill, NC 27599-7420, USA
e-mail: lin@bios.unc.edu.

yields Cox's estimator for the vector of regression parameters as well as an estimator for the cumulative baseline hazard function. His derivation also led to an estimator of the baseline survival function that is simpler than the one given by Cox (1972). In modern language, the approach taken by Breslow (1972) is the nonparametric maximum likelihood estimation (NPMLE) (Andersen et al. 1992, pp. 221–229; pp. 481–483; Kalbfleisch and Prentice 2002, pp. 114–128).

In response to the discussion of his paper, Cox (1975) introduced the concept of partial likelihood and showed that the "conditional likelihood" he originally called is in fact a partial likelihood. The concept of partial likelihood has little usage outside of the Cox regression model. By contrast, the NPMLE approach and the resulting Breslow estimator have far-reaching implications.

The Breslow estimator for the cumulative baseline hazard function has been implemented in all major statistical software packages. This has facilitated the use of the estimator in scientific studies. Simple transformations can be applied to the Breslow estimator to provide estimation of the baseline and conditional survival functions.

In addition to its practical impact, the Breslow estimator has greatly influenced survival analysis research in at least two major ways. First, one can estimate the cumulative baseline hazard function for many complex problems, such as correlated failure times and missing/mismeasured covariates, by mimicking the Breslow estimator. Second, one can use Breslow's NPMLE approach to estimate both the finite-dimensional and infinite-dimensional parameters in a wide variety of situations, such as frailty models and linear transformation models.

In the next section, I will describe the Cox model and the Breslow estimator. In Sects. 3–5, I will demonstrate how the Breslow estimator is used in the analysis of correlated failure time data, in two-phase studies and in the analysis of transformation models. Some concluding remarks are made in Sect. 6.

2 Cox model and Breslow estimator

The Cox (1972) proportional hazards model specifies that the hazard function of failure time T, given a vector of possibly time-dependent covariates $Z(\cdot)$, takes the form

$$\lambda(t|Z) = e^{\beta^{\mathrm{T}} Z(t)} \lambda_0(t), \tag{1}$$

where β is a vector of unknown regression parameters, and $\lambda_0(\cdot)$ is an arbitrary baseline hazard function. For a random sample of size n, the data consist of $\{\widetilde{T}_i, \Delta_i, Z_i(\cdot)\}$ $(i = 1, \ldots, n)$, where $\widetilde{T}_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, C_i is the censoring time for the ith subject, and $I(\cdot)$ is the indicator function. Cox (1972, 1975) proposed to estimate β by the partial likelihood

$$PL(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta^{\mathrm{T}} Z_i(\widetilde{T}_i)}}{\sum_{j \in \mathcal{R}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)}} \right\}^{\Delta_i},$$

where $\mathcal{R}_i = \{j : \widetilde{T}_j \geq \widetilde{T}_i\}$. The corresponding score function is

$$U(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i(\widetilde{T}_i) - \frac{\sum_{j \in \mathcal{R}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)} Z_j(\widetilde{T}_i)}{\sum_{j \in \mathcal{R}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)}} \right\}. \tag{2}$$

The maximum partial likelihood estimator $\widehat{\beta}$ is the maximizer of $PL(\beta)$ or equivalently the solution to $U(\beta) = 0$.

The cumulative baseline hazard function is given by $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Breslow (1972) suggested to estimate β and Λ_0 in the maximum likelihood framework. The joint likelihood for β and Λ_0 is

$$L(\beta, \Lambda_0) = \prod_{i=1}^{n} \left\{ e^{\beta^{\mathrm{T}} Z_i(\widetilde{T}_i)} \lambda_0(\widetilde{T}_i) \right\}^{\Delta_i} \exp\left\{ -\int_0^{\widetilde{T}_i} e^{\beta^{\mathrm{T}} Z_i(t)} \lambda_0(t) dt \right\}.$$

By treating $\lambda_0(\cdot)$ as piecewise constant between uncensored failure times, one can show that $L(\beta, \Lambda_0)$ is maximized simultaneously at $\widehat{\beta}$ and

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \frac{I(\widetilde{T}_i \leq t)\Delta_i}{\sum_{j \in \mathcal{R}_i} e^{\widehat{\beta}^{\mathrm{T}} Z_j(\widetilde{T}_i)}}. \tag{3}$$

The latter is the well-known Breslow estimator. The corresponding estimator for the conditional survival function under $Z = z$ is

$$\widehat{S}(t|z) = \exp\left\{ -\int_0^t e^{\widehat{\beta}^{\mathrm{T}} z(u)} d\widehat{\Lambda}_0(u) \right\}. \tag{4}$$

A number of researchers studied the asymptotic properties of $\widehat{\beta}$ and $\widehat{\Lambda}_0$ in the 1970's and 1980's. Tsiatis (1981) provided the first published proof. Andersen and Gill (1982) extended the Cox model to general counting processes and established an asymptotic theory via the elegant counting process martingale theory.

In the counting process notation, the data consist of $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}(i = 1, \ldots, n)$, where $N_i(t)$ records the number of events observed on the ith subject by time t, $Y_i(t)$ indicates whether the ith subject is at risk at time t, and $Z_i(t)$ is the set of covariates at time t. We assume that the intensity function for $N_i(t)$ is $Y_i(t)e^{\beta^{\mathrm{T}} Z_i(t)}\lambda_0(t)$, where β is a set of unknown regression parameters, and $\lambda_0(\cdot)$ is an arbitrary positive function. The partial likelihood for β is

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{e^{\beta^{\mathrm{T}} Z_i(t)}}{\sum_{j=1}^{n} Y_j(t) e^{\beta^{\mathrm{T}} Z_j(t)}} \right\}^{\Delta N_i(t)},$$

where $\Delta N_i(t) = N_i(t) - N_i(t-)$. The Breslow estimator for $\Lambda_0(t) \equiv \int_0^t \lambda_0(u)du$ is

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\sum_{j=1}^{n} Y_j(u) e^{\widehat{\beta}^{\mathrm{T}} Z_j(u)}},$$

where $\widehat{\beta}$ maximizes $PL(\beta)$.

As in the special case of the proportional hazards model, $\widehat{\beta}$ and $\widehat{\Lambda}_0(\cdot)$ are NPMLEs. The two estimators are consistent and asymptotically normal. In addition, they are asymptotically efficient (Andersen et al. 1992, Sect. VIII. 4.3). Specifically, $n^{1/2}(\widehat{\beta}-\beta)$ converges to a zero-mean normal distribution with a covariance matrix that can be consistently estimated by $\{\mathcal{I}(\widehat{\beta})/n\}^{-1}$, where $\mathcal{I}(\beta) = -\partial^2 \log PL(\beta)/\partial\beta^2$. In addition, $n^{1/2}\{\widehat{\Lambda}_0(t) - \Lambda_0(t)\}$ converges to a zero-mean Gaussian process whose covariance function at (t, s) can be consistently estimated by

$$\left\{\int_0^t \overline{Z}(\widehat{\beta}; u) d\widehat{\Lambda}_0(u)\right\}^{\mathrm{T}} \left\{\mathcal{I}(\widehat{\beta})/n\right\}^{-1} \left\{\int_0^s \overline{Z}(\widehat{\beta}; u) d\widehat{\Lambda}_0(u)\right\} + \int_0^{\min(t,s)} \frac{d\widehat{\Lambda}_0(u)}{n^{-1}\sum_{j=1}^{n} Y_j(u) e^{\widehat{\beta}^{\mathrm{T}} Z_j(u)}},$$

where

$$\overline{Z}(\beta; t) = \frac{\sum_{j=1}^{n} Y_j(t) e^{\beta^{\mathrm{T}} Z_j(t)} Z_j(t)}{\sum_{j=1}^{n} Y_j(t) e^{\beta^{\mathrm{T}} Z_j(t)}}.$$

These asymptotic results enable one to make inference about β and $\Lambda_0(\cdot)$. In the special case of the proportional hazards model, the asymptotic properties for $\widehat{S}(t|z)$ given in (4) follow from the δ-method. Simultaneous confidence bands can be constructed for the survival function estimator (Lin et al. 1994).

The Breslow estimator and the corresponding survival function estimator have been implemented in all major statistical software packages, such as SAS and S-plus, and are widely used in scientific studies. Fleming and Harrington (1991) described an interesting application to the Mayo primary biliary cirrhosis study, in which the survival function estimator given in (4) was used to understand the natural history of the disease, to counsel patients and to determine intervention strategies. Andersen et al. (1992) and Kalbfleisch and Prentice (2002) provided many other examples.

3 Correlated failure time data

The Breslow estimator also plays an essential role in the extension of Cox regression to correlated or multivariate failure time data. These data arise when each subject can potentially experience multiple events or when the subjects are sampled in clusters such that the failure times within the same cluster tend to be correlated. We shall focus on the latter situation, although the former can be handled in a very similar manner.

For $i = 1, \ldots, n$ and $l = 1, \ldots, n_i$, let T_{il} be the failure time for the lth member of the ith cluster, and let C_{il} and $Z_{il}(\cdot)$ be the corresponding censoring time and covariates. The data consist of $\{\widetilde{T}_{il}, \Delta_{il}, Z_{il}(\cdot)\}(i = 1, \ldots, n; l = 1, \ldots, n_i)$, where $\widetilde{T}_{il} = \min(T_{il}, C_{il})$, and $\Delta_{il} = I(T_{il} \leq C_{il})$. There are two major modelling approaches: frailty models and marginal models.

In the marginal modelling approach, we assume that the marginal distributions of the T_{il}'s satisfy the proportional hazards model while leaving the dependence of related failure times unspecified. To be specific, the marginal hazard function of T_{il} takes the form

$$\lambda(t|Z_{il}) = e^{\beta^{\mathrm{T}} Z_{il}(t)} \lambda_0(t),$$

where β is a set of regression parameters, and $\lambda_0(\cdot)$ is an arbitrary baseline hazard function.

Under the independence working assumption, the "partial likelihood" for β is

$$\widetilde{PL}(\beta) = \prod_{i=1}^{n} \prod_{l=1}^{n_i} \left\{ \frac{e^{\beta^{\mathrm{T}} Z_{il}(\widetilde{T}_{il})}}{\sum_{j=1}^{n} \sum_{k=1}^{n_j} I(\widetilde{T}_{jk} \geq \widetilde{T}_{il}) e^{\beta^{\mathrm{T}} Z_{jk}(\widetilde{T}_{il})}} \right\}^{\Delta_{il}}.$$

Denote the maximizer of $\widetilde{PL}(\beta)$ by $\widehat{\beta}$. In analogy with (3), the Breslow estimator for $\Lambda_0(t)$ is

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \sum_{l=1}^{n_i} \frac{I(\widetilde{T}_{il} \leq t)\Delta_{il}}{\sum_{j=1}^{n} \sum_{k=1}^{n_j} I(\widetilde{T}_{jk} \leq \widetilde{T}_{il}) e^{\widehat{\beta}^{\mathrm{T}} Z_{jk}(\widetilde{T}_{il})}}.$$

The estimators $\widehat{\beta}$ and $\widehat{\Lambda}_0$ are consistent and asymptotically normal with easily estimated variances and covariances (Lee et al. 1992; Spiekerman and Lin 1998).

In the frailty approach, we incorporate an unobserved frailty into the proportional hazards model to account for the dependence of the failure times with in each cluster. Specifically,

$$\lambda(t|Z_{il}, \xi_i) = \xi_i e^{\beta^{\mathrm{T}} Z_{il}(t)} \lambda_0(t),$$

where ξ_i is a positive random variable with density function $f(\xi;\gamma)$ indexed by parameter γ, β is a set of regression parameters, and $\lambda_0(\cdot)$ is an arbitrary baseline hazard function.

The joint likelihood for β, γ and $\Lambda_0(t) \equiv \int_0^t \lambda_0(u)du$ is

$$L(\beta, \gamma, \Lambda_0) = \prod_{i=1}^{n} \int_{\xi_i} \prod_{l=1}^{n_i} \left\{ \xi_i e^{\beta^{\mathrm{T}} Z_{il}(\widetilde{T}_{il})} \lambda_0(\widetilde{T}_{il}) \right\}^{\Delta_{il}}$$
$$\times \exp\left\{ - \int_0^{\widetilde{T}_{il}} \xi_i e^{\beta^{\mathrm{T}} Z_{il}(t)} \lambda_0(t) dt \right\} f(\xi_i; \gamma) d\xi_i.$$

Following Breslow (1972), we treat $\lambda_0(\cdot)$ as piecewise constant between uncensored failure times and maximize $L(\beta, \gamma, \Lambda_0)$ over β, γ and Λ_0 simultaneously. The resulting NPMLEs of β, γ and Λ_0 are consistent, asymptotically normal, and asymptotically efficient (Parner 1998; Zeng et al. 2007).

It is convenient to obtain the NPMLEs via the EM-algorithm. In the M-step, we solve the following equation for β

$$\sum_{i=1}^{n}\sum_{l=1}^{n_i}\Delta_{il}\left[Z_{il}(\widetilde{T}_{il}) - \frac{\sum_{j=1}^{n}\sum_{k=1}^{n_j} I(\widetilde{T}_{jk} \geq \widetilde{T}_{il})\widehat{\xi}_j e^{\beta^{\mathrm{T}} Z_{jk}(\widetilde{T}_{il})} Z_{jk}(\widetilde{T}_{il})}{\sum_{j=1}^{n}\sum_{k=1}^{n_j} I(\widetilde{T}_{jk} \geq \widetilde{T}_{il})\widehat{\xi}_j e^{\beta^{\mathrm{T}} Z_{jk}(\widetilde{T}_{il})}}\right] = 0,$$

and estimate $\Lambda_0(t)$ by

$$\sum_{i=1}^{n}\sum_{l=1}^{n_i}\frac{I(\widetilde{T}_{il} \leq t)\Delta_{il}}{\sum_{j=1}^{n}\sum_{k=1}^{n_j} I(\widetilde{T}_{jk} \geq \widetilde{T}_{il})\widehat{\xi}_j e^{\beta^{\mathrm{T}} Z_{jk}(\widetilde{T}_{il})}}, \tag{5}$$

where $\widehat{\xi_i} = E(\xi_i|\text{Data})$. In addition, we estimate γ by solving the equation

$$\sum_{i=1}^{n} E\{\partial \log f(\xi_i; \gamma)/\partial\gamma|\text{Data}\} = 0.$$

In the E-step, we evaluate the conditional expectations through numerical integration. Interestingly, (5) takes the same form as the Breslow estimator shown in (3).

4 Two-phase sampling

For large cohorts with infrequent failures and expensive covariate measurements, it is cost-effective to measure the expensive covariates on all the cases (uncensored observations) and a small subset of the controls (censored observations). There are two popular sampling strategies: under the nested case-control sampling (Thomas 1977; Prentice and Breslow 1978), a random sample of controls of fixed size m, typically in the range 1 to 5, is selected for each case; under the case-cohort sampling (Prentice 1986), a sub-cohort is randomly selected from the whole cohort to provide the controls for all the cases. Breslow-type estimators can be constructed under such two-phase sampling.

We return to the setting described in the first paragraph of Sect. 2. For nested case-control studies, the partial likelihood for β is

$$PL(\beta) = \prod_{i=1}^{n}\left\{\frac{e^{\beta^{\mathrm{T}} Z_i(\widetilde{T}_i)}}{\sum_{j\in\widetilde{\mathcal{R}}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)}}\right\}^{\Delta_i},$$

where $\widetilde{\mathcal{R}}_i$ consists of the ith subject and the m controls selected at $\widetilde{T}_i$. In light of (3), the Breslow-type estimator for $\Lambda_0(t)$ is

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \frac{I(\widetilde{T}_i \le t)\Delta_i}{(m+1)^{-1} n(\widetilde{T}_i) \sum_{j \in \widetilde{\mathcal{R}}_i} e^{\widehat{\beta}^{\mathrm{T}} Z_j(\widetilde{T}_i)}},$$

where $n(t) = \sum_{j=1}^{n} I(\widetilde{T}_j \ge t)$, and $\widehat{\beta}$ is the maximizer of $PL(\beta)$. The asymptotic properties of $\widehat{\beta}$ and $\widehat{\Lambda}_0$ were established by Goldstein and Langholz (1992) and Borgan et al. (1995). The latter authors also studied the Breslow estimator for various extensions of the nested case-control sampling.

For case-cohort studies, we estimate β and Λ_0 by mimicking (2) and (3)

$$\widetilde{U}(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i(\widetilde{T}_i) - \frac{\sum_{j \in \widetilde{\mathcal{R}}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)} Z_j(\widetilde{T}_i)}{\sum_{j \in \widetilde{\mathcal{R}}_i} e^{\beta^{\mathrm{T}} Z_j(\widetilde{T}_i)}} \right\},$$

and

$$\widetilde{\Lambda}_0(t) = \frac{\widetilde{n}}{n} \sum_{i=1}^{n} \frac{I(\widetilde{T}_i \le t)\Delta_i}{\sum_{j \in \widetilde{\mathcal{R}}_i} e^{\widetilde{\beta}^{\mathrm{T}} Z_j(\widetilde{T}_i)}},$$

where $\widetilde{n}$ is the total number of subject in the sub-cohort, $\widetilde{\mathcal{R}}_i$ consists of the sub-cohort members who are at risk at time $\widetilde{T}_i$, and $\widetilde{\beta}$ is the solution to $\widetilde{U}(\beta) = 0$. The asymptotic properties of $\widetilde{\beta}$ and $\widetilde{\Lambda}_0$ were studied by Self and Prentice (1988). More efficient case-cohort designs and estimators have been explored by Borgan et al. (2000) and Kulich and Lin (2004), among others.

The aforementioned estimators for nested case-control and case-cohort studies are not asymptotically efficient. Efficient estimation of β and Λ_0 can be achieved by adopting the NPMLE approach of Breslow (1972); see Scheike and Juul (2004), Scheike and Martinussen (2004), and Zeng et al. (2006).

Lin (2000) and Breslow and Wellner (2007) considered more general two-phase sampling. By weighting a selected subject's contribution to (2) and (3) by his/her inverse probability of selection, we can obtain consistent and asymptotically normal estimators of β and Λ_0. Efficient estimation under general two-phase sampling is an open problem.

5 Transformation models

When covariates are time-independent, model (1) can be written as

$$H(T) = -\beta^{\mathrm{T}} Z + \epsilon, \tag{6}$$

where $H(\cdot)$ is an arbitrary increasing function, and ϵ has the extreme-value distribution. With various choices of the error distribution, (6) represents a rich class of linear transformation models. The specific choice of the standard logistic error distribution yields the proportional odds model. In the presence of time-dependent covariates, we formulate transformation models through the cumulative hazard function

$$\Lambda(t|Z) = G\left\{ \int_0^t e^{\beta^{\mathrm{T}} Z(u)} d\Lambda(u) \right\}, \tag{7}$$

where $G(\cdot)$ is a specific transformation function, and $\Lambda(\cdot)$ is an arbitrary increasing function (Zeng and Lin 2006). The omission of a covariate in the proportional hazards model would yield a transformation model in the form of (7). If covariates are time-independent, then (6) and (7) are equivalent.

Except for the special case of the proportional hazards model, the partial likelihoods for (6) and (7) involve infinite-dimensional nuisance parameters and the resulting estimators are computationally intractable and statistically inefficient. By contrast, the NPMLE approach of Breslow (1972) extends naturally from the proportional hazards model to general transformation models.

The joint likelihood for β and Λ under model (7) is

$$L(\beta, \Lambda) = \prod_{i=1}^{n} \left[e^{\beta^{\mathrm{T}} Z_i(\widetilde{T}_i)} \lambda(\widetilde{T}_i) G' \left\{ \int_0^{\widetilde{T}_i} e^{\beta^{\mathrm{T}} Z_i(u)} d\Lambda(u) \right\} \right]^{\Delta_i} \times \exp\left[-G\left\{ \int_0^{\widetilde{T}_i} e^{\beta^{\mathrm{T}} Z_i(u)} d\Lambda(u) \right\} \right],$$

where $\lambda(t) = d\Lambda(t)/dt$, and $G'(x) = dG(x)/dx$. Following Breslow (1972), we treat $\lambda(\cdot)$ as piecewise constant between uncensored failure times and maximize $L(\beta, \Lambda)$ over β and Λ simultaneously. By constructing a frailty which induces the transformation function G, we can turn the numerical problem of estimating the transformation model into that of estimating the proportional hazards frailty model and apply the EM algorithm described in Sect. 3. The NPMLEs of β and Λ are consistent, asymptotically normal and asymptotically efficient (Zeng and Lin 2006).

We can define transformation models for general counting processes by extending (7) to the cumulative intensity function (Zeng and Lin 2006). Furthermore, we can extend transformation models to correlated failure time data (Zeng et al. 2007). We can again adopt Breslow's (1972) NPMLE approach to obtain efficient estimators of the model parameters.

6 Remarks

I have demonstrated, through a variety of problems, that the Breslow estimator and the NPMLE approach of Breslow (1972) have had enormous impact on survival analysis and semiparametric inference. The specific problems were chosen from my areas of research because I am more familiar with them than with others and also because I wanted to show how my own research has been influenced by this particular piece of work by Breslow (1972), which is less than one page in print. I am not aware of any other contribution to the discussion of a RSS read paper that has had such a great impact.

Norm Breslow has made many other path-breaking contributions to statistics, especially in categorical data analysis, generalized linear models, generalized linear mixed models, and case-control studies. Some of those contributions are reflected in other papers of this special issue.

Norm Breslow's contributions to statistics go well beyond his original statistical research, although that is what he takes the most pride in. He has been a highly successful collaborator in the Wilms tumor research and other scientific studies. His two monographs with Nick Day on *Statistical Methods in Cancer Research* (Breslow and Day 1980, 1987) are the standard reference on epidemiological methods. Many of Norm's former students have become extraordinary leaders in our field. Norm has given lectures and seminars around the world. He served as the department chair for many years and recently served as the President of the International Biometric Society. Those of us who have been fortunate enough to know him all admire his great passion and tremendous kindness. He is truly inspirational!

Acknowledgements It is my distinct pleasure to contribute to this special issue of LIDA in honor of Norm Breslow, whose illustrious achievements have greatly advanced the field of biostatistics and inspired generations of biostatisticians. I thank a referee for helpful comments. This work was supported by the National Institutes of Health.

References

Andersen PK, Borgan Ø, Gill RD, Keiding N (1992) Statistical models based on counting processes. Springer, New York
Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. Ann Statist 10:1100–1120
Borgan O, Goldstein L, Langholz B (1995) Methods for the analysis of sampled cohort data in the Cox proportional hazards model. Ann Statist 23:1749–1778
Borgan Ø, Langholz B, Samuelsen SO, Goldstein L, Pagoda J (2000) Exposure stratified case-cohort designs. Lifetime Data Analy 6:39–58
Breslow NE (1972) Discussion of the paper by D. R. Cox. J R Statist Soc B 34:216–217
Breslow NE, Day NE (1980) Statistical methods in cancer research: 1. The design and analysis of case-control studies. IARC Scientific Publications 32. International Agency for Research on Cancer, Lyon, France
Breslow NE, Day NE (1987) Statistical methods for cancer research: 2. The design and analysis of cohort studies. IARC Scientific Publications 82. International Agency for Research on Cancer, Lyon, France
Breslow NE, Wellner JA (2007) Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. Scand J Statist 34:86–102
Cox DR (1972) Regression models and life-tables (with discussion). J R Statist Soc B 34:187–220
Cox DR (1975) Partial likelihood. Biometrika 62:269–276
Fleming TR, Harrington DP (1991) Counting processes and survival analysis. Wiley, New York
Goldstein L, Langholz B (1992) Asymptotic theory for nested case-control sampling in the Cox regression model. Ann Statist 20:1903–1928
Kalbfleisch JD, Prentice RL (1972) Discussion of the paper by D. R. Cox. J R Statist Soc B 34:215–216
Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd ed. Wiley, Hoboken
Kulich M, Lin DY (2004) Improving the efficiency of relative-risk estimation in case-cohort studies. J Am Statist Ass 99:832–844
Lee EW, Wei LJ, Amato DA (1992) Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein JP, Goel PK (eds) In survival analysis: state of the art. Kluwer Academic, Norwell, Mass., pp 237–247.

Lin DY (2000) On fitting Cox's proportional hazards models to survey data. Biometrika 87:37–48

Lin DY, Fleming TR, Wei LJ (1994) Confidence bands for survival curves under the proportional hazards model. Biometrika 81:73–81

Parner E (1998) Asymptotic theory for the correlated gamma-frailty model. Ann Statist 26:183–214

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

Prentice RL, Breslow NE (1978) Retrospective studies and failure time models. Biometrika 65:153–158

Scheike TH, Juul A (2004) Maximum likelihood estimation for Cox's regression model under nested case-control sampling. Biostatistics 5:193–206

Scheike TH, Martinussen T (2004) Maximum likelihood estimation for Cox's regression model under case-cohort sampling. Scand J Statist 31:283–293

Self SG, Prentice RL (1988) Asymptotic distribution theory and efficiency results for case-cohort studies. Ann Statist 16:64–81

Spiekerman CF, Lin DY (1998) Marginal regression models for multivariate failure time data. J Am Statist Assoc 93:1164–1175

Thomas DC (1977) Addendum to Methods for cohort analysis: Appraisal by application to asbestos mining by FDD Liddell, JC McDonald, PC Thomas. J Roy Statist Soc Ser A 140:469–491

Tsiatis AA (1981) A large sample study of Cox's regression model. Ann Statist 9:93–108

Zeng D, Lin DY (2006) Efficient estimation of semiparametric transformation models for counting processes. Biometrika 93:627–640

Zeng D, Lin DY, Lin X (2007) Semiparametric transformation models with random effects for clustered failure time data. Stat Sin (in press)

Zeng D, Lin DY, Avery CL, North KE (2006) Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. Biostatistics 7:486–502