Event History Analysis

Niels Keiding

Department of Biostatistics, University of Copenhagen, Copenhagen DK-1014, Denmark; email: nike@sund.ku.dk

Annu. Rev. Stat. Appl. 2014. 1:333–60

First published online as a Review in Advance on October 11, 2013

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

This article's doi: 10.1146/annurev-statistics-022513-115558

Keywords

causal analysis, censoring, illness-death process, local dependence, multistate model, prevalent cohort study

Abstract

Event history analysis deals with data obtained by observing individuals over time, focusing on events occurring for the individuals under observation. Important applications are to life events of humans in demography, life insurance mathematics, epidemiology, and sociology. The basic data are the times of occurrence of the events and the types of events that occur. The standard approach to the analysis of such data is to use multistate models; a basic example is finite-state Markov processes in continuous time. Censoring and truncation are defining features of the area. This review comments specifically on three areas that are current subjects of active development, all motivated by demands from applications: sampling patterns, the possibility of causal interpretation of the analyses, and the levels and interpretation of variability.

1. INTRODUCTION

Event history analysis deals with data obtained by observing individuals over time, focusing on events occurring for the individuals under observation. Important applications are to life events of humans in demography, life insurance mathematics, epidemiology, and sociology. The interest is in modeling individual event histories, which in some disciplines is termed a microdata approach, as opposed to the aggregate-data approach. The basic data are the times of occurrence of the events and the types of events that occur. Today, the standard approach to the analysis of such data is to use multistate models; a basic example is finite-state Markov processes in continuous time.

Major progress in this field occurred between 1975 and 1990, fueled by the increased availability of sufficiently detailed databases and by remarkable technical-statistical development building directly on the "French theory" of stochastic processes (for a personal account of some of this development, see Aalen et al. 2009). Several authoritative surveys of various lengths and depths have covered this body of theory and methodology, allowing me to be fairly concise in the technical presentation here. Several monographs have also discussed the general methodology of event history analysis (often with applications primarily in biostatistics) (Kalbfleisch & Prentice 1980, 2002; Jacobsen 1982; Fleming & Harrington 1991; Andersen et al. 1993; Klein & Moeschberger 1997; Ibrahim et al. 2001; Martinussen & Scheike 2006; Aalen et al. 2008), whereas some books have concentrated more specifically on survival analysis (Cox & Oakes 1984, Therneau & Grambsch 2000, Collett 2003, Lawless 2003) and on event history analysis in the social sciences (Allison 1984, Yamaguchi 1991, Courgeau & Lelièvre 1992, Trussell et al. 1992, Blossfeld & Rohwer 1995, Mills 2011). Many articles on themes regarding event history analysis are found in the *Encyclopedia of Biostatistics* (Armitage & Colton 2005) and have been collected in a special volume (Andersen & Keiding 2006). Finally, several survey papers have also been published on event history analysis (Commenges 1999, Hougaard 1999, Andersen & Keiding 2002, Putter et al. 2007, Meira-Machado et al. 2009, Andersen & Perme 2013). In this review, I comment specifically on three areas that are still subjects of active development, all motivated by demands from applications: sampling patterns, the possibility of causal interpretation of the analyses, and the levels and interpretation of variability.

1.1. Sampling Patterns

Event history analysis has always had the intrinsic practical problem that events happen on the same timescale as that of the observer: We cannot wait until everybody has died; we may be dead by then. Therefore, censoring (incomplete observation of being at risk for events) and truncation (observation conditional on being at risk for events) are almost defining features of event history analysis. However, in practice, many other observational designs are important, and there have been recent significant developments. Notable examples come directly from practical observation patterns: intermittent observation (interval censoring) and observation around a cross section (including retrospective observation). Most methods assume that censoring is noninformative.

1.2. Event History Analysis and Causal Inference

Event history analysis incorporates time as an essential ingredient in statistical modeling; therefore, there should be ample opportunity to contribute to the current development in causal analysis. I briefly review the development in event history analysis of the concept of local dependence, which corresponds closely to the econometric concept of Granger causality. Another important recent

development is the approach initiated by J.M. Robins to handling time-dependent confounding. This area uses some tools from event history analysis, but it has also had to introduce essential new concepts to crack problems such as the healthy worker effect, which was fully described more than a century ago.

1.3. Levels and Interpretation of Variability

Most event history models in practical use still build on the Poisson distribution postulate stating that variability is constrained by the model to be given by variance = mean, implicitly making a very restrictive assumption of residual homogeneity. This problem has long been recognized in the methodological literature, but general recommendations for use in textbooks and software are not yet ready. On a different level, heterogeneity between individuals within groups may also often be fruitfully described using randomness. Starting with motivations in demography, such random variation has been studied under the label "frailty" and initially focused on probability modeling rather than statistical analysis. Yet, more work is needed before we have a full multilevel methodology in event history analysis. Both of these issues are prominent in applications to demography, which has large databases in which customary measures of uncertainty lose their relevance and individual homogeneity becomes a remotely relevant hypothesis.

1.4. Article Overview

Sections 2–4 introduce the basic multistate models, their counting process representations, and likelihood structures. The focus is on the three-state illness-death (or disability) model and the four-state model describing interactions between life history events. The most commonly used statistical models in event history analysis concern transition intensities, which are surveyed in Section 5. Section 6 exemplifies the need for targets of inference other than the transition intensities and the occurrence of non-Markov models. The final sections cover the specific issues listed above: sampling patterns in Section 7, local independence in Section 8, time-dependent confounding in Section 9, and the role of random variation in Section 10. This article ends with brief concluding remarks.

2. SURVIVAL ANALYSIS

Multistate models describe how individuals move between a finite number of states. The simplest example is the survival model with one transient state, "0: alive," and one absorbing state, "1: dead." This model is characterized by the distribution of the survival time T, representing the time from a given origin (time 0) to the occurrence of the event "death." The distribution of T may be characterized by the distribution function $F(t) = \text{Prob}(T \leq t)$ or, equivalently, by the survival function $S(t) = 1 - F(t) = \text{Prob}(T > t)$. Accordingly, $S(t)$ and $F(t)$ correspond to the probabilities of being in state 0 and 1, respectively, at time t. If every individual is assumed to be in state 0 at time 0, then $F(t)$ is also the transition probability from state 0 to state 1 for the time interval from 0 to t. In continuous time, the distribution of T may also be characterized by the hazard rate function $\alpha(t) = -\mathrm{d}\log S(t)/\mathrm{d}t$, that is, $S(t) = \exp(-A(t))$ with $A(t) = \int_0^t \alpha(u)\mathrm{d}u$.

Thus, $\alpha(\cdot)$ is the transition intensity from state 0 to state 1.

In general, event history analysis deals with inference for transition intensities and transition probabilities in multistate models. This includes estimation and hypothesis tests for these quantities and analysis of regression models where these quantities are related to (possibly time-dependent) explanatory variables observed for the individuals under study. Most frequently, multistate models

are defined by their transition intensities from which transition probabilities may or may not be derived depending on the modeling assumptions.

A typical feature of event history analysis is the inability of observing complete event histories such that, for example, by the end of the observation period, all individuals under study may not have reached an absorbing state. In survival analysis, this would correspond to individuals still being alive by the end of the study; this kind of incomplete observation is known as right censoring. Furthermore, all individuals may not have been observed from the same time origin; this kind of incomplete observation where individuals are observed only conditionally on not having reached an absorbing state by the time of initiation of the study is known as left truncation. Restricting attention to right censoring, we find that a crucial problem is whether the available incomplete data enable valid inference to be made on parameters in the multistate model for the complete data. The condition for this is known as independent right censoring, and the interpretation is that a sample observed after independent right censoring is representative for the population without censoring. As such, individuals who are censored should have neither lower nor higher risk of future events than do individuals who are not censored (see, for example, Andersen et al. 1993, ch. III; Kalbfleisch & Prentice 2002, section 6.2).

3. MULTISTATE MODELS

A multistate process is a stochastic process $(X(t), t \in \mathcal{T})$ with a finite state space $\mathcal{S} = \{1, \ldots, p\}$ and with right-continuous sample paths: $X(t+) = X(t)$. Here, $\mathcal{T} = [0, \tau]$ or $[0, \tau)$ with $\tau \leq +\infty$. The process has initial distribution $\pi_h(0) = \text{Prob}(X(0) = h)$, $h \in \mathcal{S}$, transition probabilities

$$P_{hj}(s, t) = \text{Prob}\left(X(t) = j \,\middle|\, X(s) = h, \{X(u), 0 \leq u \leq s\}\right)$$

for $h, j \in \mathcal{S}, s, t \in \mathcal{T}, s \leq t$, and transition intensities given by the derivatives

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{P_{hj}(t, t + \Delta t)}{\Delta t},$$

which we assume exist. Some transition intensities may be 0 for all t. Graphically, multistate models may be illustrated using diagrams with boxes representing the states and with arrows between the states representing the possible transitions, i.e., the nonzero transition intensities (Hoem 1976, Andersen et al. 1993). A state $h \in \mathcal{S}$ is absorbing if, for all $t \in \mathcal{T}, j \in \mathcal{S}, j \neq h, \alpha_{hj}(t) = 0$; otherwise, h is transient. The state probabilities $\pi_h(t) = \text{Prob}(X(t) = h)$ are given by

$$\pi_h(t) = \sum_{j \in \mathcal{S}} \pi_j(0) P_{jh}(0, t).$$

Notice that $P_{hj}(\cdot, \cdot)$ and, thus, $\alpha_{hj}(\cdot)$ depend on both the probability measure Prob and the history $\{X(u), 0 \leq u < s\}$, though this dependence has been suppressed in the notation. If $\alpha_{hj}(t)$ depends on the history only via the state $h = X(t)$ occupied at t, then the process is Markovian.

3.1. The Two-State Model for Survival Data

The two-state model for survival data, illustrated in **Figure 1**, has $p = 2$ states and only one possible transition from state 0 to 1. The corresponding transition intensity $\alpha_{01}(t)$ is given by the hazard rate function $\alpha(t)$, whereas $\alpha_{10}(t) = 0$ for all t; that is, state 1 is absorbing. The initial distribution is degenerate in 0 : $\pi_0(0) = 1$, and the process is Markovian.

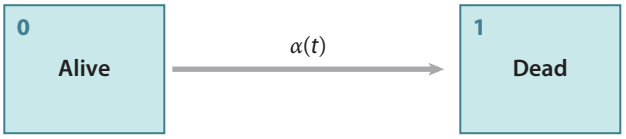

Figure 1

The two-state model for survival data.

3.2. The Competing Risks Model

The competing risks model has one transient state "0: alive" and a number k of absorbing states, with state h, $h = 1, \ldots, k$ corresponding to "death from cause h." Thus, there are $p = k+1$ states. The model is illustrated for $k = 2$ in **Figure 2**.

The transition intensities $\alpha_{0h}(t)$ for $h = 1, \ldots, k$ are given by the cause-specific hazard functions, here denoted $\alpha_h(t)$:

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}\,(\textit{Dead from cause } \ h \ \textit{ by } t + \Delta t \,|\, T \geq t)}{\Delta t},$$

where T is the survival time. The initial distribution is degenerate in 0, the only transient state of the model, i.e., $\alpha_{hj}(t) = 0$ for all $h \neq 0$ and all j. The transition probabilities are given by the survival function

$$P_{00}(0, t) = S(t) = \text{Prob}(T > t) = \exp\left(-\int_0^t \sum_{h=1}^{k} \alpha_h(u)\mathrm{d}u\right)$$

and the so-called cumulative incidence functions

$$P_{0h}(0, t) = \int_0^t S(u-)\alpha_h(u)\mathrm{d}u, \quad h = 1, \ldots, k.$$

As with the simple two-state model ($k = 1$), the competing risks model is Markovian.

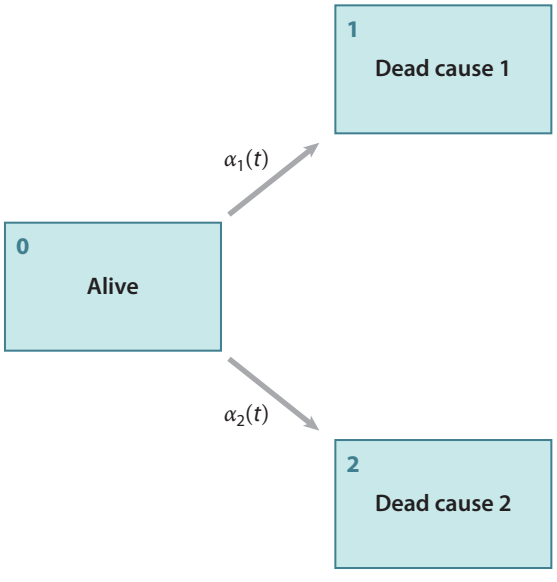

Figure 2

Competing risks model for mortality from two causes.

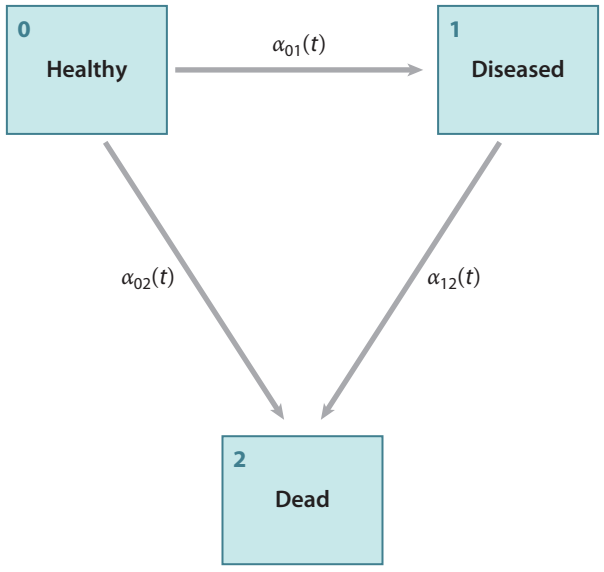

Figure 3

Unidirectional illness-death model.

3.3. The Illness-Death Model

The illness-death model is illustrated in **Figure 3**. Often, the time t is the age of the individual, and individuals will usually be assumed to be in state 0 at $t = 0$. However, individuals will not always be observed from $t = 0$ (further discussed below). The mortality $\alpha_{12}(t)$ of the diseased (the lethality) may sometimes depend on duration d since entry to state 1 in addition to the dependence on age t. [Notice that, despite the notation, $\alpha_{12}(t)$ then depends on the random time of the most recent transition into 1.] If $\alpha_{12}(t)$ does not depend on d, the process is Markovian; otherwise, it is a semi-Markov process.

The transition probabilities in this model have simple explicit expressions:

$$P_{00}(s,t) = \exp\left(-\int_s^t (\alpha_{02}(u) + \alpha_{01}(u))\mathrm{d}u\right)$$

and (in the Markovian case)

$$P_{01}(s,t) = \left(\int_s^t P_{00}(s,u-)\alpha_{01}(u)P_{11}(u,t)\mathrm{d}u\right), \tag{1}$$

where

$$P_{11}(s,t) = \exp\left(-\int_s^t \alpha_{12}(u)\mathrm{d}u\right).$$

If the lethality $\alpha_{12}(\cdot)$ depends on both age and duration, then we define

$$\alpha_{12}(t,d) = \lim_{\Delta t \to 0} \frac{\mathrm{Prob}\left(X(t+\Delta t) = 2 \,\middle|\, X(t) = 1, 0 \to 1 \text{ transition at } t-d\right)}{\Delta t},$$

and $P_{11}(u,t)$ in Equation 1 should be replaced by $\exp(-\int_u^t \alpha_{12}(s, s-u)\mathrm{d}s)$. The illness-death model is one of the most important multistate models and was discussed in early papers by Fix & Neyman (1951) and Sverdrup (1965).

4. COUNTING PROCESS REPRESENTATION: LIKELIHOOD

Assume that multistate processes $X_i(t)$ such as those described in Section 3 are observed over intervals $[0, \tau_i]$ for individuals $i = 1, \ldots, n$. Assume first that τ_i is a fixed (i.e., nonrandom) time of termination of observation for individual i. Random right censoring (see Section 2) and delayed entry are treated below. Because $X_i(t)$ is constant between transitions, it is equivalent to record $X_i(0)$, and the counting processes

$$N_{hj}^i(t) = \#\,(\text{direct transitions } h \to j \text{ in } [0, t] \text{ for } i)\,,$$

as described by the times T_{hj}^{ik} of these transitions, where

$$0 < T_{hj}^{i1} < \cdots < T_{hj}^{iN_{hj}^i(\tau_i)} \leq \tau_i.$$

Let $N_{hj}(t) = \sum_{i=1}^n N_{hj}^i(t)$. It is also useful to introduce $Y_h^i(t) = I\{X_i(t-) = h\}$ and

$$Y_h(t) = \#\,(\text{individuals "at risk" in state } \ h \ \text{ at time } t-) = \sum_{i=1}^n Y_h^i(t).$$

Note that since, for $t > \tau_i$, $N_{hj}^i(t) = N_{hj}^i(\tau_i)$ and $Y_h^i(t) = 0$, these quantities (strictly speaking, processes) can be considered as defined on $(0, \infty)$.

Conditional on the initial state $X_i(0)$ and the time-fixed covariates Z_i, the parameters of the model are the transition intensities $\alpha_{hj}^i(t)$, and the likelihood is

$$\prod_{i=1}^n \prod_{h \neq j} \prod_{k=1}^{N_{hj}^i(\tau_i)} \alpha_{hj}^i\left(T_{hj}^{ik}\right) \exp\left(-\int_0^{\tau_i} \alpha_{hj}^i(t) Y_h^i(t) \mathrm{d}t\right) \tag{2}$$

(Andersen et al. 1993). Recall from the above that the notation $\alpha_{hj}^i(t)$ represents possible dependence of the transition intensity on the whole history of the process. Thus, $\alpha_{hj}^i(t)$ may contain covariates and other random elements, as already exemplified.

Two patterns of incomplete observations in particular are easily tractable because they lead to only minor modifications of this likelihood: delayed entry, where individual i enters at some time V_i, and right censoring, where nothing is known about i after some time U_i. Both V_i and U_i may be random, although only either dependent on the previous history of the process or independent of the process. The reason for the particular tractability of these mechanisms is that the at-risk indicator $Y_h^i(t) = I\{X_i(t-) = h\}$ in the likelihood only needs to be amended to

$$Y_h^i(t) = I\left\{X_i(t-) = h, V_i < t \leq U_i\right\}.$$

Andersen et al. (1993, ch. III) gave a detailed specification of likelihood derivation and conditions on censoring patterns in a counting process framework. Commenges & Gégout-Petit (2007) gave a general mathematical discussion of likelihood for coarsened observations from multistate models.

5. STATISTICAL MODELS FOR TRANSITION INTENSITIES IN MARKOV PROCESSES

As indicated in the Introduction (Section 1), the primary purpose of event history analysis is to gain insight into the dynamics of the processes by estimating transition intensities and, perhaps, by assessing their dependence on covariates, often using various types of transition probabilities or state occupation probabilities obtained by integrating certain functions of the transition intensities.

A final purpose may be prediction, both as an illustration of the dynamics and for concrete practical use. This section considers the most important class of models—the (continuous-time) Markov process $X(t)$ on the finite state space $\mathcal{S} = \{1, \ldots, p\}$, where the dependence of $\alpha_{hj}^i(t)$ on the history $\mathcal{X}_i$ introduced at the beginning of Section 3 is via the current state of $X(t)$ (and possibly via time-fixed covariates). Observations may be left truncated and right censored (I give examples of other observational patterns in Section 7).

So far, the dominating approach to statistical modeling has been to specify the class of transition intensities $(\alpha_{hj}^i(t))$ for each individual i. The following sections focus on estimation of these intensities and the associated transition probabilities. Section 6 then lists some other targets of inference that have generated separate methodological developments.

5.1. Constant and Piecewise Constant Transition Intensities: Parametric Models

The simplest class of models is obtained by keeping the transition rates constant: $\alpha_{hj}^i(t) = \alpha_{hj}^i$. Piecewise constant intensities

$$\alpha_{hj}^i(t) = \alpha_{hj}^{i(l)}, \qquad t_{l-1}^{hj} < t \leq t_l^{hj}, \qquad \text{all } t_0 = 0$$

form the next step. This choice is of widespread use, particularly in large studies in econometrics, epidemiology, sociology, and demography (Hoem 1976, Andersen et al. 1993, Clayton & Hills 1993).

Assume first that all individuals have the same transition intensities, $\alpha_{hj}^{i(l)} = \alpha_{hj}^{(l)}$. The likelihood in Equation 2 then simplifies to

$$\prod_l \prod_{h \neq j} (\alpha_{hj}^{(l)})^{N_{hj}^{(l)}} e^{-\alpha_{hj}^{(l)} S_h^{(l)}},$$

where $N_{hj}^{(l)} = N_{hj}(t_l^{hj}) - N_{hj}(t_{l-1}^{hj})$ and

$$S_h^{(l)} = \sum_i \int_{t_{l-1}^{hj}}^{t_l^{hj}} Y_h^i(t) \mathrm{d}t.$$

This likelihood resembles the one resulting from observations of $N_{hj}^{(l)}$ events in a Poisson process with intensity $\alpha_{hj}^{(l)}$ observed over the interval $(0, S_h^{(l)})$, except that here $S_h^{(l)}$ is random. Thus, this likelihood, and even the model, is often associated with Poisson's name. Maximum likelihood estimation is elementary, yielding

$$\hat{\alpha}_{hj}^{(l)} = \frac{N_{hj}^{(l)}}{S_h^{(l)}},$$

the classical occurrence/exposure rate. Asymptotic inference may be obtained from the observed information

$$-D^2 \log L = \frac{N_{hj}^{(l)}}{\left(\alpha_{hj}^{(l)}\right)^2},$$

yielding variance estimates

$$\mathrm{Var}\left(\hat{\alpha}_{hj}^{(l)}\right) \sim \frac{\left(\alpha_{hj}^{(l)}\right)^2}{N_{hj}^{(l)}} \sim \frac{N_{hj}^{(l)}}{\left(S_h^{(l)}\right)^2}$$

and all estimators asymptotically independent.

Transition probabilities for the constant and piecewise constant Markov process models are explicit functions of the transition intensities (Chiang 1968), allowing direct plug-in maximum likelihood estimation as well as calculation of standard error estimates via the delta method. Although the piecewise constant model is often sufficient to describe the dependence of intensities on time, other possibilities exist. Certain mathematical functions of time may also generate the model, such as the Gompertz-Makeham model for mortality

$$\alpha(t) = \alpha + \beta\gamma^t.$$

However, except for mortality studies in actuarial and some demographic contexts, such parametric models are little used. One reason for this may be the development of the powerful methodology for nonparametric statistical inference, where $\alpha_{hj}(t)$ is left unspecified.

5.2. Freely Varying (Nonparametric) Transition Intensities

Assume first that the transition intensities are the same for all individuals but that they are allowed to vary freely with time: $\alpha^i_{hj} = \alpha_{hj}(t)$. Statistical inference is then conveniently phrased in terms of the counting process approach pioneered by Aalen (1975, 1978) (for a detailed exposition, see Andersen et al. 1993). Estimators (which may be given a nonparametric maximum likelihood interpretation) of the integrated intensities

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u)\mathrm{d}u$$

are obtained as the Nelson–Aalen estimators

$$\hat{A}_{hj}(t) = \int_0^t \frac{J_h(u)}{Y_h(u)}\mathrm{d}N_{hj}(u) = \sum_i \sum_{k:0<T^{ik}_{hj}<t} \frac{1}{Y_h\left(T^{ik}_{hj}\right)}, \tag{3}$$

where $J_h(u) = I\{Y_h(u) > 0\}$, with variance estimators

$$\hat{\sigma}^2\left(\hat{A}_{hj}(t)\right) = \int_0^t \frac{J_h(u)}{Y_h(u)^2}\mathrm{d}N_{hj}(u) = \sum_i \sum_{k:0<T^{ik}_{hj}<t} \frac{1}{Y_h\left(T^{ik}_{hj}\right)^2}.$$

A detailed mathematical theory based on stochastic integrals and martingales is available to study exact and asymptotic properties of these estimators. When estimates are desired of the transition intensities $\alpha_{hj}(t)$, rather than of their integrals, smoothing techniques are necessary (Andersen et al. 1993).

An important feature of the nonparametric approach is its elegant generalization by Aalen & Johansen (1978) to estimating transition probabilities. The basic tool is the (matrix) product integral. Let $\mathbf{I}$ be the identity matrix and $\mathbf{G}$ a matrix-valued function. The corresponding product integral is defined as

$$\Pi_0^t\left(\mathbf{I} + \mathbf{G}(\mathrm{d}s)\right) = \lim_{\max|t_v - t_{v-1}|\to 0} \prod \left(\mathbf{I} + \mathbf{G}\left(t_v\right) - \mathbf{G}\left(t_{v-1}\right)\right),$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[0, t]$. In particular, if G is continuous and scalar,

$$\Pi_0^t\left(1 + G(\mathrm{d}s)\right) = e^{G(t)-G(0)},$$

and if G is a scalar step function,

$$\Pi_0^t\,(1+G(\mathrm{d}s)) = \prod_{k=1}^{K}\left(1+\Delta G\left(t_{(k)}\right)\right),$$

where $t_{(0)}=0$ and $0<t_{(1)}<\cdots<t_{(k)}\le t$ are the jump times of G and

$$\Delta G\left(t_{(k)}\right) = G\left(t_{(k)}\right) - G\left(t_{(k-1)}\right).$$

Define $\alpha_{hh}(t) = -\sum_{j\neq h}\alpha_{hj}(t)$ and the intensity matrix function $\mathbf{A}(t)=(\alpha_{hj}(t))$; then the matrix $\mathbf{P}(s,t)=(P_{hj}(s,t))$ of transition probabilities

$$P_{hj}(s,t) = \mathrm{Prob}\left(X_i(t)=j \,\middle|\, X_i(s)=h\right)$$

is given by

$$\mathbf{P}(s,t) = \Pi_s^t\,(\mathbf{I}+\mathbf{A}(\mathrm{d}u)).$$

The Aalen–Johansen estimator of $\mathbf{P}(s,t)$ is obtained by plugging the matrix of Nelson–Aalen estimators $(\hat{A}_{hj}(t))$ into the formula

$$\hat{\mathbf{P}}(s,t) = \Pi_s^t\left(\mathbf{I}+\hat{\mathbf{A}}\,(\mathrm{d}u)\right).$$

For the simple two-state model for survival data, $\hat{P}_{00}(0,t)$ reduces to the classical Kaplan–Meier estimator $\hat{S}(t)=\prod_{T_i\le t}(1-\mathrm{d}N_{01}(T_i)/Y_0(T_i))$ of the survival function $S(t)$ (Kaplan & Meier 1958). As documented in detail by Andersen et al. (1993), there is a well-developed theory, again based on stochastic integrals and martingales, about the asymptotic properties of the Aalen–Johansen estimator.

5.3. Markov Regression Models

Most regression models for multistate processes focus on modifying the transition intensities. For an individual i with time-fixed covariates $Z_i \simeq (Z_{im})$, the proportional intensity model introduced by Cox (1972) for survival data postulates the decomposition

$$\alpha_{hj}^i(t) = \alpha_{hj0}(t)\exp(\beta_{hj}' Z_i) \tag{4}$$

of the $h\to j$ transition intensity into a freely varying baseline $h\to j$ transition intensity $\alpha_{hj0}(t)$, assumed common for all individuals, and a factor independent of time t, describing the effect of a covariate Z_{im} by factors of proportionality $\exp(\beta_{hjm})$. Choosing the baseline intensity piecewise constant leads to Poisson regression models. In both cases, inference may be based on the likelihood given by Equation 2, which for the Cox model leads to the so-called Cox's partial likelihood (Cox 1975, Andersen et al. 1993). The choice between Cox and Poisson models is frequently a matter of convenience, though the latter may be advantageous in large studies where a sufficiency reduction of data into tables of event counts and person-years within groups of (categorical) covariates is feasible (Clayton & Hills 1993). In contrast, application of the Cox model requires one data record per individual for each transition, leading to a considerable computational burden in large studies.

Another regression model for survival data that readily extends to multistate models is Aalen's nonparametric additive model (Aalen 1980, 1989; Andersen et al. 1993),

$$\alpha_{hji}(t) = \alpha_{hj0}(t) + \beta_{hj}'(t) Z_i.$$

In this model, both the baseline transition intensities $\alpha_{hj0}(t)$ and the regression functions $\beta_{hjm}(t)$ are left unspecified, and nonparametric estimates may be obtained using a generalized least-squares procedure. Aalen et al. (2001, 2008) and Martinussen & Scheike (2006) reviewed and generalized this model and its use in event history analysis.

6. OTHER TARGETS OF INFERENCE: NON-MARKOV MODELS

The exposition of the multistate models in the previous sections has implicitly assumed that the mathematical building blocks, the transition intensities $\alpha_{hj}(t)$, are also the natural target of the statistical inference. The proportional intensity models in the spirit of Cox (1972) or the additive intensity models pioneered by Aalen (1980) both lead to regression coefficients interpretable as modifiers of these intensities. At times, formulating regression models for targets other than the transition intensities has been useful. Below, I list some examples of alternative regression models as well as some special functionals of the underlying processes that have motivated special statistical developments.

6.1. Mean Residual Life and Backward Recurrence Time

As a simple example in survival analysis, it may be relevant to study the mean residual life

$$e(x) = E(X - x \,|\, X > x) = (S(x))^{-1} \int_x^\infty S(u)\mathrm{d}u.$$

Oakes & Dasu (1990) defined the proportional mean residual life model using

$$e_Z(x) = \exp(\gamma Z)\, e_0(x)$$

for covariates Z. In this model, the regression coefficient γ is directly interpretable in the mean residual life context, but it has no simple relation to the familiar regression coefficients from the Cox proportional hazards model (for a recent survey of nonparametric estimation of mean residual life from censored data, see McLain & Ghosh 2011).

Another example from simple survival analysis regards estimation from current duration (backward recurrence) data (for an application to time to pregnancy, see Keiding et al. 2002, 2011, 2012; for an application to last-episode data in sociology, see Yamaguchi 2003). In the example of time to pregnancy, a possible design consists in asking (e.g., in a telephone survey) a sample of women of fertile age "Are you currently trying to become pregnant?" and, if the answer is affirmative, "For how long have you tried?". Yamaguchi's (2003) example is about residential mobility: From the distribution of elapsed duration in the current residence, we want to derive the duration distribution and relate it to covariates.

The statistical problem is to estimate the distribution of a random variable X with density $f(x)$ and survival function $S(x) = \int_x^\infty f(u)\mathrm{d}u$ from observations of the corresponding backward recurrence time (here conveniently termed as current duration) with density $g(y)$. Under stationarity, as with similar problems in renewal theory, $S(x) = g(x)/g(0)$. As discussed by Keiding et al. (2011, 2012) and Yamaguchi (2003), this structure makes accelerated failure time (AFT) regression models useful, because, if the current duration Y satisfies an AFT model with baseline density g_0 and baseline survival function $S_0(y) = \int_y^\infty g_0(u)\mathrm{d}u$, i.e.,

$$P(Y > y|z) = S_0(ye^{\beta z}),$$

then the survival function S_Z of X is given by

$$S_Z(x) = \frac{g_Z(x)}{g_Z(0)} = \frac{g_0(xe^{\beta Z})}{g_0(0)},$$

which is again an AFT model with the same β but a new baseline survival function $g_0(\cdot)/g_0(0)$.

6.2. Cumulative Incidence in Competing Risks Models

Returning to multistate models, the competing risks model has been particularly subject to controversy. Assume without essential loss of generality that there are two competing risks, then the model is specified by the transition intensities $\alpha_1(t)$ and $\alpha_2(t)$, for which hazard regression models can be easily posulated. However, in most applications, a central functional of interest is the cumulative incidence

$$CI_h(t) = \int_0^t \exp\left(-\int_0^u (\alpha_1(v) + \alpha_2(v))\,\mathrm{d}v\right)\alpha_h(u)\,\mathrm{d}u.$$

Although it is perfectly feasible to estimate $CI_h(t)$ by plugging in the estimated regression coefficients from the component models for $\alpha_1(t)$ and $\alpha_2(t)$, these estimates do not deliver a direct message about the dependence of $CI_h(t)$ on covariates.

This situation has generated a demand for models that more directly describe the dependence of this functional on covariates. The proportional subdistribution hazard model of Fine & Gray (1999) provided a starting point, defining

$$\tilde{\alpha}_h(t) = \frac{\partial}{\partial t}\left(-\log\left(1 - CI_h(t)\right)\right) \tag{5}$$

and postulating, in the Cox tradition, models

$$\tilde{\alpha}_h(t; Z) = \tilde{\alpha}_h(t; 0)\exp(\beta Z).$$

This model is widely used and has the important feature that the implied $CI_h(t, \beta Z)$ does vary monotonically with β. A drawback, however, is the somewhat convoluted interpretation of the subdistribution hazard function: As discussed in detail by Andersen & Keiding (2012), $\tilde{\alpha}_h(t)$ is the hazard of the improper random variable $\inf_t(X(t) = h)$;

$$\tilde{\alpha}_h(t)\mathrm{d}t = P\left(X\,(t + \mathrm{d}t) = h \,\middle|\, X(t) \neq h\right),$$

which translates into the hardly interpretable infinitesimal probability that individuals will die of cause h given that they either are still alive or have already died from another cause. Alternative proposals of direct binomial regression of $CI_h(t)$ (avoiding intensity modelling) were given by Scheike & Zhang (2007, 2008), Scheike et al. (2008), and Gerds et al. (2012) using inverse probability weighting. Yet another approach is based on jackknife-inspired pseudovalues (see, e.g., Klein & Andersen 2005, Andersen & Klein 2007, Graw et al. 2009).

6.3. Estimation of Stage Occupation Probabilities for Non-Markov Multistate Models

Pepe (1991) noted that stage occupation probabilities in multistate models may often be represented as differences between survival probabilities, with no assumption on the detailed probabilistic structure (such as finite-state Markov process). Accordingly, stage occupation probabilities may be estimated as differences of Kaplan–Meier estimators, thus Pepe (1991) derived variance

estimates and two-sample test statistics from this representation. Klein et al. (2000) compared these useful ideas with the Aalen–Johansen approach to estimate a current leukemia-free survival curve for stem cell–transplanted, chronic myeloid leukemia patients.

In an important series of papers, Datta & Satten (2001, 2002; Satten & Datta 2002) showed that the Aalen–Johansen estimator of stage occupation probabilities remains valid for censored observations from a general class of non-Markov multistate models under general censoring patterns. This work has been followed up by de Uña-Álvarez and colleagues (see, e.g., Meira-Machado et al. 2006, 2009; Meira-Machado 2011; Rodríguez-Girondo & deUña-Álvarez 2012), who developed estimators in the non-Markov illness-death process and the three-state progressive process (differing from the illness-death process by excluding deaths from the first stage) as well as nonparametric tests for Markovianity. Commenges et al. (2007) proposed a general Kullback–Leibler-type criterion for choosing between Markov and specific non-Markov models.

7. OBSERVATIONAL STUDIES OF DISEASE INCIDENCE AND MORTALITY

Many problems regarding observational structures for multistate processes are well exemplified by the illness-death process, which is a versatile tool for modeling many studies of disease incidence and mortality. Such studies illustrate the rich variety of observational patterns in event history analysis.

7.1. Interval Censoring

Assume first that we are studying the occurrence of a particular chronic disease such as dementia using the simple illness-death model described in Section 3.2. In the prospective PAQUID cohort study around Bordeaux, France (Commenges et al. 2004), healthy people (state 0) were followed from some age $t_0 > 65$ years, and at follow-up visits 1, 3, 5, and 8 years later, researchers assessed whether dementia had developed. For participants who died during follow-up, the exact date of death was available. This very common observational pattern implies that the transition $0 \to 1$ occurred at some time in the interval (t_1, t_2) for participants in state 0 at a visit at age t_1 and state 1 at a later visit at age t_2. For a diseased patient who dies, we know the precise age of death, that is, the precise age of the $1 \to 2$ transition. The particularly problematic case is when a participant who was healthy at the last visit (at age t_1, say) dies at age t_d before any further visit. We know that this participant entered state 2 at age t_d, but we do not know whether a transition $0 \to 1$ in the interval (t_1, t_d) had occurred. Therefore, we also lack information whether the transition into state 2 was from state 0 or state 1. In a series of papers, Frydman (e.g., 1992, 1995) and Frydman & Szarek (2009) generalized Turnbull's (1976) nonparametric self-consistency equation approach to illness-death processes, whereas Joly et al. (2002, 2009) postulated freely varying continuous intensity functions and approximated roughness-penalized maximum likelihood estimators using splines.

A variant of the above problem occurs when the date of transition to the absorbing state is also interval censored, as was the case in a study of survival of dental fillings in primary teeth (Joly et al. 2012; cf. a similar example from dental research by Frydman et al. 2013). Additionally, a monograph by Sun (2006) surveyed interval censoring, whereas Commenges (2002) and Lesaffre et al. (2005) provided other surveys with emphasis on applications to dental research (see also Goméz et al. 2009). Finally, using Bayesian inference, Komárek & Lesaffre (2008) recommended the AFT model for multivariate doubly interval-censored data.

7.2. Informative Observation Plans: Patient Self-Selection

In Section 7.1, it was implicitly assumed that the observation times delineating the interval censored observations are noninformative with respect to the disease process. An obvious violation of this assumption is patient-initiated visits to the doctor or dentist: Because patients are more likely to seek medical care when they need it, we cannot take the time of the visit as uninformative with respect to the disease process. This issue has been discussed surprisingly rarely in the biostatistical literature. An important early paper by Grüger et al. (1991) gave a formal definition of a noninformative observation scheme in the spirit of the then-recently developed concept of noninformative censoring in survival analysis. The authors noted the nontestability of noninformativeness from the observed data and gave careful practical advice on informative observation schemes, with patient self-selection as the most critical example. They illustrated the concepts using a four-state event history model describing serum alpha-fetoprotein as a marker for hepatocellular carcinoma.

In another example, Åhlström et al. (1999) studied an illness-death model for relapse and postulated that relapse could be detected either at a preplanned visit or at a spontaneous (i.e., patient-initiated) visit in response to symptoms. These authors derived a bivariate-phase-type distribution for the joint distribution of time X to relapse and time Y to symptoms and explained the censoring patterns generated by the observation scheme. They then highlighted the connection to the appearance of subclinical disease in statistical models for screening for chronic diseases. Liestøl & Andersen (2002) proposed an elaboration of the illness-death model by a crisis state with reversible transitions, from which patient-initiated visits would be particularly common, and Guihenneuc-Jouyaux et al. (2000) elaborated this idea as a hidden Markov process handled by Markov chain Monte Carlo techniques.

7.3. Sampling at a Cross Section

In practice, many observational studies happen with consecutive (staggered) entry, and it becomes necessary to keep track of the calendar time in which the study takes place, particularly when sampling of the process takes place at a particular point in time, i.e., at a cross section. A simple example assumes that all individuals are entered at the calendar time of their birth and followed through to possible disease and ultimate death. **Figure 4** is a Lexis diagram representation of this model, with a so-called cross section demonstrating the additional feature of sampling the population at calendar time t_0.

Keiding (1991, 2006; cf. Lund 2000) gave examples of problems in observational epidemiology that could all be viewed as special cases of this type of sampling, including current status data, retrospective incidence estimation, current duration data and prevalent cohort sampling. The illness-death process noted in Section 3.3 is still the basic framework. Current status data record the ages at t_0 of healthy and diseased individuals. Under the restrictive assumptions of no differential mortality between healthy and diseased and with no calendar time effects on birth, incidence, and mortality, it is possible to estimate the incidence rate $\alpha_{01}(a)$, in practice represented by the corresponding distribution function $1 - \exp(-\int_0^a \alpha_{01}(u)\mathrm{d}u)$.

Retrospective incidence estimation can sometimes be based on information about onset ages for each individual in the prevalent sample. Because individuals are included conditional on having survived until the sampling date t_0, information on mortality during the years before t_0 is required so that each individual may be weighted with the inverse probability of being sampled. Retrospective information (e.g., census data) on the population at risk is also necessary. Keiding et al. (1989) presented an early case study concerning historical incidence of diabetes, which was further developed by Ogata et al. (2000). Such calculations are still being developed (see, e.g., Alioum et al. 2005, Addona et al. 2009).

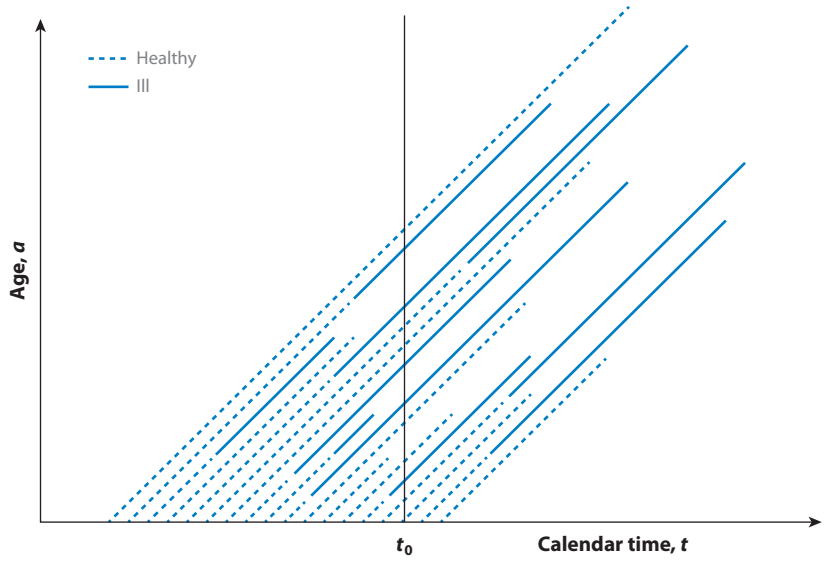

Figure 4

Lexis diagram of individuals born healthy (*dashed lines*), possibly becoming ill (*solid lines*), and dying. A cross section is taken at time t_0. Redrawn from figure 1 of Keiding (2006) in accordance with the *Statistics in Medicine* practice of allowing authors to reuse their own figures.

Current duration data (backward recurrence times) are used to estimate mortality from the age distribution of the living at the cross section, assuming stationarity. The basic idea has been known for a long time in demography: In a stationary population, the life table (survival function of the individuals) is proportional to the density function of the age distribution of the living. More recently, such data are termed open interval or last episode (cf. Yamaguchi 2003, who was inspired by migration analysis; see also Tomé et al. 2006). As briefly mentioned above, Keiding et al. (2002, 2012) and Scheike & Keiding (2006) surveyed application of this approach to time-to-pregnancy studies, and Slama et al. (2006, 2012) applied the technique to a French telephone survey on time to pregnancy.

Mortality estimation from follow-up of a prevalent cohort is an important biostatistical technique, which is usefully illustrated by the Lexis diagram in **Figure 4**. For a brief account, assume that the target is the mortality of the diseased or lethality $\alpha_{12}(t, a, d)$ (see **Figure 3**), which in general depends on calendar time t, age a, and disease duration d. A standard version of the problem based on the Lexis diagram in **Figure 4** is outlined in **Figure 5**. Let $V =$ age at disease onset, $Y =$ age at entry into the study (i.e., at time t_0), $X =$ age at death, and $T =$ time at death; then, $T_e = T - (X - V)$ is time at entry (t_0, above). A simple concrete calculation (Keiding 1992) shows that the intensity of a diseased individual dying at time t and age x, given disease onset at age $v < x$ and entry into the study at age $y > v$ equals the lethality $\alpha_{12}(t, x, x - v)$ for $x > y$. In other words, despite the length-biased recruitment of prevalent patients into the prevalent cohort, standard survival analysis with delayed entry applies. Note, however, that this exposition works conditionally on the realized prevalent sample. Because individuals with long disease durations are overrepresented in the prevalent sample (length bias), including the distribution of the sample in the analysis may yield further information (see Bergeron et al. 2008, Cook & Bergeron 2011).

Wang et al. (1993; cf. Ripley & Solomon 1995) discussed the use of Cox regression models for prevalent cohort data. For these models, the choice of time origin for the time variable entering

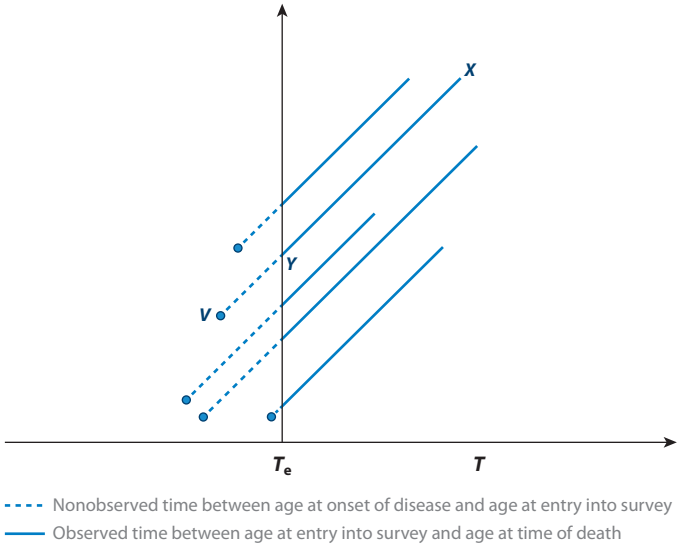

Figure 5

Lexis diagram illustrating follow-up of diseased individuals. $V =$ age at onset of disease (assumed known), dashed lines indicate nonobserved time between age at onset of disease and age at entry into the survey, $Y =$ age at entry into survey, solid lines indicate observed time between age at entry into survey and age at time of death, $X =$ age at death, $T_e =$ time at entry, and $T =$ time at death. Redrawn from figure 16 in Keiding (2006) in accordance with the *Statistics in Medicine* practice of allowing authors to reuse their own figures.

into the underlying intensity is important, and the use of time-dependent covariates requires special care (Keiding & Knuiman 1990). Using a special version of Cox regression modelling of the relative mortality, Andersen et al. (1985) presented a case study on survival of a prevalent cohort of diabetics.The above setup regarding statistical inference of survival data based on sampling on the Lexis diagram has recently been generalized by Guilloux (2007) and Brunel et al. (2008).

8. LOCAL INDEPENDENCE

The four-state model in **Figure 6** describes how occurrence of one event B in a life history may change the occurrence of another event A. Schweder (1970) introduced this model as a four-state so-called composable Markov process. Taking the temporal order of the two events into account makes it possible to develop an asymmetric concept of dependence: Assume $\alpha_{0B} = \alpha_{A,AB}$ but $\alpha_{0A} \neq \alpha_{B,AB}$, then A may be considered as locally dependent on B, but B is locally independent of A. Schweder's (1970) paper included references only to classical texts on probability and lacks explicit advice about possible statistical implementation of the ideas.

In their nonparametric analysis of a case study of local independence, Aalen et al. (1980) implemented ideas from the counting process approach to event history analysis that Aalen (1975, 1978) had pioneered. The chronic skin disease pustulosis palmoplantaris is much more common among women than among men, and its first occurrence is often when women are in their fifties. A dermatologist had collected a cross-sectional sample of retrospectively observed ages at first occurrence as well as of (right-censored) age at (natural or induced) menopause. Various elaborations of Aalen's (1975, 1978) new procedures were necessary to handle the retrospective design, which is interpretable in the framework of sampling on the Lexis diagram

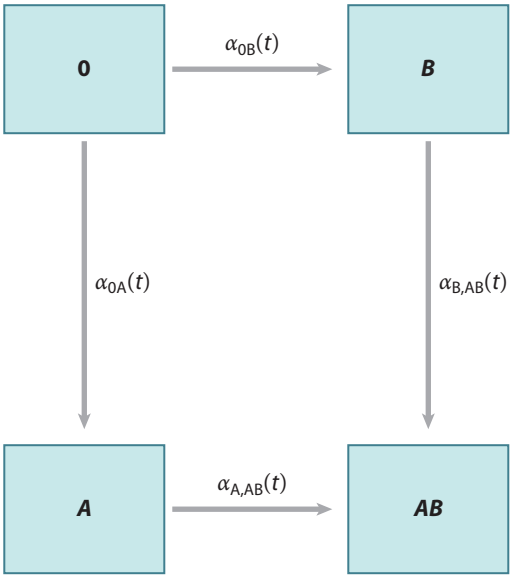

Figure 6

Schweder's (1970) composable Markov process.

(see Keiding 2006). Nevertheless, Aalen et al. (1980) became one of the first publications to indicate the newfound power of event history analysis. The study indicated that menopause may be considered a possible risk factor for incidence of this disease, and this finding was confirmed in another, independently collected data set (see Keiding 2006, 2013). Schweder's (1970) basic idea turned out to correspond with what is now called Granger causality (Granger 1969, Aalen 1987, Florens & Fougere 1996, Aalen et al. 2008). Although this approach to local independence was first carried through in a medical application, it attracted particular interest in the social sciences, where around 1980 the importance of event history analysis approaches to studying interferences between social processes was being recognized (e.g., Tuma et al. 1979, Hoem & Funck Jensen 1982).

In demography, event history analysis of composable Markov processes was notably followed up by D. Courgeau and his colleagues. Courgeau & Lelièvre (1992) summarized their experience to date, focusing on the interaction between marriage and leaving the farming community. Exploring event history analysis for social scientists, Blossfeld & Rohwer (1995) took the composable Markov processes as their point of departure for a general discussion of parallel and interdependent processes, where an important additional issue is the distinction between the individual (micro) and group (macro) levels (an issue discussed below). Among the several examples concerning the new roles of women were the interaction between employment and marriage and that between marriage and birth of first child.

8.1. Further Development of the Local Dependence Concept

The local dependence concept due to Schweder (1970) and discussed above has been further developed. Aalen (1987) generalized the concept beyond the original Markov process framework: Consider two stochastic processes $Y_i(t), i = 1, 2$ with histories (technically, filtrations)

$(\mathcal{F}_t^i)$, $i = 1, 2$, where

$$\mathcal{F}_t^i = \sigma\left(Y_i(s)\middle|\, 0 \le s \le t\right) \text{ and } \mathcal{F}_t = \mathcal{F}_t^1 \vee \mathcal{F}_t^2$$

and compensators $\Lambda_i(t)$ with respect to $(\mathcal{F}_t^i)$ so that

$$M_i(t) = Y_i(t) - \Lambda_i(t)$$

is a martingale with respect to $(\mathcal{F}_t^i)$, $i = 1, 2$. Assume that $M_1(t)$ and $M_2(t)$ are orthogonal (this is actually a no-unmeasured-confounder assumption). We then have the following:

Definition 1: Y_1 is locally independent of Y_2 ($Y_2 \not\to Y_1$), if $\Lambda_1(t)$ is measurable with respect to $(\mathcal{F}_t^1)$ for all t.

However, Aalen's (1987) definition still considers only two processes at a time. A multivariate generalization was later obtained by Didelez (2000, 2007, 2008). I here outline the multivariate definition of Didelez (2000, 2008): Consider a k-variate process $(Y(t)) = (Y_1(t), \ldots, Y_k(t))$, and define subprocesses $Y_A(t) = (Y_i(t), i \in A)$ for $A \subset \{1, \ldots, k\}$. Assume histories given by $\mathcal{F}_t^i = \sigma(Y_i(s)|0 \le s \le t)$, and define $\mathcal{F}_t^A = \vee_{i\in A}\mathcal{F}_t^i$. For all subsets $A, B \subset \{1, \ldots, k\}$, define (vector) compensators Λ_A, Λ_B, and assume that the martingales $Y_A - \Lambda_A$ and $Y_B - \Lambda_B$ are orthogonal.

Definition 2: Y_B is locally independent of Y_A given Y_C if all $\mathcal{F}_t^{A\cup B\cup C}$-compensators $\Lambda_i, i \in B$ are measurable with respect to $\mathcal{F}_t^{B\cup C}$. Write $Y_A \not\to Y_B\,|Y_C$ or $A \not\to B\,|C$. Otherwise, Y_B is locally dependent of Y_A given Y_C.

Didelez then constructed a theory of graphical models to describe the local independencies. Let $V = \{1, \ldots, k\}$ be a local independence graph—a directed (not necessarily acyclic) graph defined by the pairwise dynamic Markov property:

$$\text{no edge from } j \text{ to } k \Leftrightarrow Y_j \not\to Y_k \,\big|Y_{V\setminus\{j,k\}}\,.$$

Define also the local dynamic Markov property

$$\forall i \in V : V \setminus \text{closure}(i) \not\to \{i\} \,\big|\text{parents}(i)$$

and the global dynamic Markov property: For subsets $A, B, C \subset V$, C δ-separates A from B in the directed graph $A \not\to B|C$. (δ-separation is a generalization of the well-known concept of d-separation to directed graphs.) Accordingly, Didelez proved the following:

Theorem 1: Under regularity conditions, the three Markov properties are equivalent.

The relation of local independence to the recent development in statistical causality based on hypothetical interventions (cf. Pearl 2009) has been discussed by Eichler & Didelez (2010), Gégout-Petit & Commenges (2010), and Aalen et al. (2012).

9. TIME-DEPENDENT COVARIATES AND TIME-DEPENDENT CONFOUNDING

In his famous paper introducing the multiplicative hazard regression model (see Section 5.3), Cox (1972) noted that the estimation algorithm would work equally well if the covariate z depended

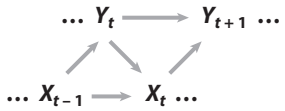

Figure 7

Time-dependent confounding.

on time t. Such an assumption obviously destroyed the attractive decomposition

$$\alpha^i_{hj}(t) = \alpha_{hj0}(t) \ \exp\ \left(\beta'_{hj} Z_i\right)$$

of the hazard into a factor $\alpha_{hj0}(t)$ depending only on time but common to all values of the covariates and another factor $\exp(\beta'_{hj} Z_i)$ depending only on covariates, not on time. Cox's main purpose was not to propose a wide-ranging generalization of his new model, but rather to suggest a simple goodness-of-fit test against a well-defined alternative of nonproportional hazards. Nevertheless, the option of letting covariates depend on time was quickly implemented in the many statistical packages that facilitated the widespread use of Cox's model, and this option has been commonly used, not always with due attention to the intricate causal structure sometimes implied.

Already in the published discussion of Cox's (1972) paper, Kalbfleisch and Prentice had sensed that there may be difficulties with a general inclusion of time-dependent covariates, and in their important monographs (Kalbfleisch & Prentice 1980, 2002), they pointed out that considerable care is needed to obtain meaningful analyses with time-dependent covariates. Few difficulties arise as long as the covariates are either external [fixed; defined before the study (for example, age of the individual under study)] or ancillary (generated by a stochastic mechanism unrelated to the process under study). The basic problem concerns internal covariates, corresponding to what economists call endogeneity: when the development of the response process before t influences the value of the covariate at t. The ensuing dilemma is usefully phrased in basic terms from observational epidemiology: Let X_t and Y_t be covariate and responses at time t, then we have the causal pattern in **Figure 7**. Because Y_t is a confounder for the association between X_t and Y_{t+i}, we should control for (i.e., condition on) Y_t, but because Y_t is also intermediate between X_{t-1} and Y_{t+i}, we should not dilute the effect between X_{t-1} and Y_{t+i} by conditioning on Y_t. Kalbfleisch and Prentice did not propose a way out of this dilemma, which was also gaining increased attention in the emerging interest within the social sciences in event history analysis (see, e.g., Courgeau & Lelièvre 1992; also see Blossfeld & Rohwer 1995, who essentially proposed modifications of the composable Markov processes approach).

An important example of time-dependent confounding is the healthy worker survival effect, admirably explained by Ogle (1885, p. xxiii):

> The weaker individuals, and those whose health is failing them, are thus being constantly drafted out of each industrial occupation, and especially out of those which require much vigour; and the consequence is that the death-rates in these latter occupations are unfairly lowered, as compared with the death-rates in occupations of an easier character, and still more as compared with the death-rates among those persons who are returned as having no occupation at all. (. . .)
>
> Another very serious flaw in these death-rates, when taken as measures of the relative healthiness of different industries, is due to the fact that these several industries do not start on equal terms as regards the vitality of those who follow them. A weakling will hardly adopt the trade of a Blacksmith, a Miner, or a Railway Navvy, but will preferentially take to some lighter occupation such as that of a Tailor, a Weaver, or a Shopman.

Robins (1986) clarified that a difficulty in obtaining a relevant measure of the mortality of workers in particularly exposed jobs from longitudinal studies may be explained by time-dependent confounding: The variable "employment status at time t" is an independent risk factor for death (because, as Ogle pointed out a century earlier, the weak are in particular risk of leaving the workforce, while, preferentially, the strong join the workforce in the first place), and it certainly influences later exposure history, as those who left are no longer exposed. Robins et al. (1992) further explained that the Cox model with time-dependent covariates cannot estimate what Robins (1986) had termed the causal effect of exposure, which may be loosely described here as the effect of an hypothetical intervention in which study participants were assigned to exposure rather than nonexposure in a possibly sequential randomized trial. It is important to realize that this objection to the use of the Cox model also rules out the use of other hazard-based event history models such as the local dependence models discussed in the previous section: Indeed, a new approach was necessary, one that Robins and coworkers have carried out since 1986 (for recent surveys of time-dependent confounding, see Robins & Hernán 2009 and the very accessible tutorial by Daniel et al. 2013). These new methods have been gaining more ground in epidemiology and biostatistics than in the social sciences (for a recent concrete example in demography, however, see Gerster et al. 2013, who identified a feedback effect of fertility behavior on Danish women's higher education, implying that the apparent effect from standard analyses of reduced fertility for women with academic training did not correspond to the causal effect). New methods for handling time-dependent confounding continue to be developed, and it is interesting that an important part is played by the additive intensity models pioneered by Aalen (1980).

10. THE ROLE OF RANDOM VARIATION

So far, I have presented event history analysis in a conventional statistical framework in which the statistical model essentially generalizes the binomial distribution of the number of deaths X in some fixed time interval of n independent, identically distributed individuals, all with the same death probability p. We assume, often without explicit concern, that meaningful statistical analyses may be performed using the variance $np(1-p)$ given by this binomial distribution. Early standard error calculations in mortality analysis may be considered precursors to this practice (for some details, see Keiding & Clayton 2013).

From an applied point of view, there is an elementary but important problem with this general trust in the simple binomial variance and its analogues: In most event history models used in practice, the statistical model is essentially a generalized linear model in the sense of McCullagh & Nelder (1989), which specifies the variance as a function of the mean, as in the binomial and Poisson distributions. In contrast to what is generally assumed when working with normally distributed data, there is no free parameter to account for the noise. Accordingly, the statistical inference (usually based on the likelihood function) estimates the noise in the data, not from the empirical noise, but from the empirical means. Because regression models are often underspecified in practice, the noise may be seriously underestimated, making significance tests quite anticonservative. Two main suggestions in the methodological literature remedy this problem. One is based on a postulation of an overdispersion parameter to capture at least some of the unaccounted excess variation (e.g., Breslow 1984, 1990; McCullagh & Nelder 1989, section 4.5; for more recent developments in this direction, also see Lin 2007, Molenberghs et al. 2007).

A more general approach is based on the sandwich estimator of Huber (1967) and White (1982), which provides a consistent estimator of the variance of a maximum likelihood estimator, even when the likelihood is misspecified (for surveys of this estimator, see Hardin & Hilbe 2001, Hardin

2003). In contrast to the maximum likelihood estimator of the variance, the sandwich estimator is explicitly based on the empirical variation in the data. As pointed out, e.g., by Kauermann & Carroll (2001), the sandwich estimator can be more variable than the maximum likelihood estimator of the variance. Although the sandwich estimator is making its way into major statistical software packages, it is not yet a standard theme in trend-setting textbooks, at least not in biostatistics and epidemiology.

On the more principal level is the hypothesis of identically distributed individuals. Here it is useful to focus on the example of demography. Much classical demography was concerned with the large databases of official statistics where births, deaths, marriages, etc., were classified according to age, sex, and coarse geographical stratification. Individual random variation was not always considered important (some exceptions are mentioned above). A breakthrough came with the formalization that included random variation between the individual mortalities in the multiplicative frailty model of Vaupel et al. (1979). Here the mortality of an individual i at time t, age a, and sex s is assumed to have the multiplicative form

$$\mu_i(t, a, s) = \mu(t, a, s)Z_i,$$

where the frailties Z_i in the simplest case are assumed independent, identically distributed according to some distribution (e.g., the gamma distribution is a popular choice). The individual frailties are assumed to catch at least some of the possible variation in mortality across the population. Even this simple model has generated important qualitative insights: For example, because the frail die first each individual mortality $\mu_i(t, a, s)$ increases more quickly with age a than does the average mortality of the survivors in the population. There are many further uses of frailty models as possible explanations for real-life phenomena (for a good collection, see Aalen et al. 2008).

My focus here is on the statistical uses of frailty models. An early obvious idea was to combine the semiparametric multiplicative intensity model of Cox (1972) with the gamma-distributed individual frailty factor of Vaupel et al. (1979), in the simplest case to yield a survival model where, for given frailty Z_i, individual i with covariates x_i has death intensity

$$Z_i \lambda_0 (t) \exp (\beta x_i).$$

Andersen et al. (1993, ch. IX) reviewed the early literature and gave a detailed discussion of the delicate likelihood construction and the basic EM algorithm fundamental to most of the semiparametric inference.

However, researchers soon realized that the semiparametric frailty model quickly runs into serious identifiability problems. Thus, analogous to classical results on true and spurious contagion, Lancaster & Nickell (1980) noted that, if there are no individual covariates in the above model, the frailty distribution governing Z_i is unidentifiable. Elbers & Ridder (1982) further clarified that, if there are covariates that are not equal for all individuals, then the frailty distribution is identifiable. Other issues arise in frailty modeling where standard reasoning from regression analysis with normally distributed data breaks down. As in the above-mentioned problem of estimating residual variation, we need to calibrate our intuition to the absence of a free noise parameter.

For example, an important aspect in much current survival analysis based on the Cox proportional hazards model is to assess proportionality of hazards. As mentioned above, Vaupel et al. (1979) pointed out that, even if individual hazards are proportional, heterogeneity governed by frailty distributions such as the gamma distribution generate converging hazards for the observed population. This may encourage the use of frailty models to handle the problem of converging hazards as a case of unobserved heterogeneity, but Hougaard et al. (1994) and Keiding et al. (1997) showed that there are serious problems with using frailty models in complete parallel to standard practice in ordinary linear models where we can throw variation from unobserved covariates into

the noise term—because no such independent noise term exists in the frailty model. These authors also pointed out that AFT models may be preferable in this situation because such are often ordinary linear models of log (time to death) and a separate variance parameter to handle the noise is easily available (for further discussion of this issue, see Lesaffre et al. 2005). The important monograph by Hougaard (2000) focuses primarily on frailty models for multivariate survival analysis, where information on the random variation is more obviously available as within-cluster variation. Xu et al. (2010) recently postulated a frailty model for the illness-death process, including full statistical modeling of the common situation that disease occurrence and mortality of the healthy are correlated, a central issue in the context of the present survey.

To return to demographic applications, the multilevel nature of many determinants in demographic analyses have been pointed out in many contributions by D. Courgeau and É. Lelièvre. In recent work, these authors proposed a combination of the description of individual random variation in event history analysis with various levels of covariates, using the Cox regression model to obtain a regression analysis framework. Courgeau (2010b, ch. 8) illustrated this approach through an analysis of departure from the parental home with individual fixed effects such as cohort, nationality of parents, number of siblings, father's occupation, own employment career, and random effects at the department (county) and period (i.e., calendar year) levels. Courgeau (2010a) also gave a historical survey of the use of the concept of dispersion in demography, going back to the earliest life tables in the seventeenth century: Event history analysis plays a central part in modeling dispersion in demography in the late-twentieth century. Courgeau also mentioned the arrival of Bayesian event history analysis and quoted the monograph by Ibrahim et al. (2001) in particular.

11. CONCLUDING REMARKS

Event history analysis is a well-established set of tools for analyzing empirical studies in epidemiology, demography, sociology, and economics. The central methodology is mature and well developed, both mathematically and algorithmically. A significant part of the statistical theory was unified and further developed during the years 1975–1990 using counting processes and martingales. That approach cannot, however, solve everything. For example, Andersen et al. (1993, ch. X) has already outlined situations with several timescales (really, several time origins), where the pretty martingale structures disappear and more brute-force techniques, perhaps from empirical processes, become necessary.

The current methodological activity in the area is more geared to specifics, and I have mentioned a selection of these: estimation of special functionals, special sampling patterns, and robustness to model assumptions. More functionals are constantly being covered: Andersen (2013) recently gave a cumulative incidence-based approach to cause-specific measures of life lost, compared with the standard demographic approaches by Andersen et al. (2013). I have also commented on the role of random variation: On a very practical side, it would be useful if our day-to-day noise estimates were more realistic; for applications, perhaps particularly, in the social sciences (and social medicine), more general multilevel models would be useful. Bayesian methods have not been central to development so far, and whether a fully Bayesian theory of event history models would appeal to users remains to be seen.

Let me conclude by returning to the interface between event history analysis with its local dependence/Granger causality approach to causal inference, on the one hand, and the counterfactual approach based on hypothetical interventions, on the other. As I mention above, new ideas were clearly needed to crack the problem of time-dependent confounding, and they did not come from within event history analysis. However, many tools from event history analysis are proving

very useful in the current detailed development of causal inference. Diggle et al. (2007) wrote an important paper on general drop-out patterns for longitudinal studies, with discrete-time martingales as a central tool, thereby providing a connection to the stochastic process basis of event history analysis developed during 1975–1990. Diggle et al. (2007) also provided a connection to the fascinating problem of "truncation by death"; the generic example is the follow-up studies of quality of life of the elderly (for a survey, see Kurland et al. 2009). This problem activates tools from event history analysis as well as from recent developments in causal inference. My expectation is that important new results in event history analysis will come mainly from detailed and careful analyses of concrete practical problems, just as usually happens in statistical methodology.

DISCLOSURE STATEMENT

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

LITERATURE CITED

Aalen OO. 1975. *Statistical inference for a family of counting processes*. PhD Thesis, Univ. Calif., Berkeley

Aalen OO. 1978. Nonparametric inference for a family of counting processes. *Ann. Stat.* 6:701–26

Aalen OO. 1980. A model for non-parametric regression analysis of counting processes. In *Springer Lecture Notes in Statistics*, Vol. 2: *Mathematical Statistics and Probability Theory*, ed. W Klonecki, A Kozek, J Rosiniski, pp. 1–25. New York: Springer

Aalen OO. 1987. Dynamic modelling and causality. *Scand. Actuar. J.* 1987(3–4):177–90

Aalen OO. 1989. A linear regression model for the analysis of life times. *Stat. Med.* 8:907–25

Aalen OO, Andersen PK, Borgan Ø, Gill R, Keiding N. 2009. History of applications of martingales in survival analysis. *Electr. J. Hist. Prob. Stat.* 5:1–28

Aalen OO, Borgan Ø, Fekjær H. 2001. Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* 57:993–1001

Aalen OO, Borgan Ø, Gjessing HK. 2008. *Survival and Event History Analysis. A Process Point of View*. New York: Springer

Aalen OO, Borgan Ø, Keiding N, Thormann J. 1980. Interaction between life history events: nonparametric analysis of prospective and retrospective data in the presence of censoring. *Scand. J. Stat.* 7:161–71

Aalen OO, Johansen S. 1978. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Stat.* 5:141–50

Aalen OO, Røysland K, Gran JM. 2012. Causality, mediation and time: a dynamic viewpoint. *J. R. Stat. Soc. A* 175:831–61

Addona V, Asgharian M, Wolfson DB. 2009. On the incidence-prevalence relation and length-biased sampling. *Can. J. Stat.* 37:206–18

Åhlström L, Olsson M, Nerman O. 1999. A parametric estimation procedure for relapse time distributions. *Lifetime Data Anal.* 5:113–32

Alioum A, Commenges D, Theibaut R, Dabis F. 2005. A multistate approach for estimating the incidence of human immunodeficiency virus by using data from a prevalent cohort study. *J. R. Stat. Soc. C* 54:739–52

Allison PD. 1984. *Event History Analysis: Regression for Longitudinal Event Data.* Beverly Hills/London: Sage

Andersen PK. 2013. Decomposition of number of years lost according to causes of death. *Stat. Med.* In press. doi: 10.1002/sim.5903

Andersen PK, Borch-Johnsen K, Deckert T, Green A, Hougaard P, et al. 1985. A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics* 41:921–32

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

Andersen PK, Canudas-Romo V, Keiding N. 2013. Cause-specific measures of life years lost. *Demogr. Res.* In press

Andersen PK, Keiding N. 2002. Multi-state models for event history analysis. *Stat. Methods Med. Res.* 11:91–115
Andersen PK, Keiding N, eds. 2006. *Survival and Event History Analysis*. Chichester: Wiley. 533 pp.
Andersen PK, Keiding N. 2012. Interpretability and importance of functionals in competing risks and multi-state models. *Stat. Med.* 31:1074–88
Andersen PK, Klein JP. 2007. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scand. J. Stat.* 34:3–16
Andersen PK, Perme MP. 2013. Event history analysis. In *Handbook of Survival Analysis*, ed. JP Klein, HC van Houwelingen, JG Ibrahim, TH Scheike, pp. 417–39. London: Chapman & Hall
Armitage P, Colton T, eds. 2005. *Encyclopedia of Biostatistics*. Chichester, UK: Wiley. 2nd ed.
Bergeron P-J, Asgharian M, Wolfson DB. 2008. Covariate bias induced by length-biased sampling of failure times. *J. Am. Stat. Assoc.* 103:737–42
Blossfeld H, Rohwer G. 1995. *Techniques of Event History Modeling.* Mahwah, NJ: Lawrence Erlbaum
Breslow N. 1990. Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *J. Am. Stat. Assoc.* 85:565–71
Breslow NE. 1984. Extra-Poisson variation in log-linear models. *J. R. Stat. Soc. C* 33:38–44
Brunel E, Comte F, Guilloux A. 2008. Estimation strategies for censored lifetimes with a Lexis-diagram type model. *Scand. J. Stat.* 35:557–76
Chiang CL. 1968. *Introduction to Stochastic Processes in Biostatistics*. New York: Wiley & Sons
Clayton D, Hills M. 1993. *Statistical Models in Epidemiology*. Oxford, UK: Oxford Univ. Press
Collett D. 2003. *Modelling Survival Data in Medical Research.* Boca Raton, FL: Chapman & Hall/CRC. 2nd ed.
Commenges D. 1999. Multi-state models in epidemiology. *Lifetime Data Anal.* 5:315–27
Commenges D. 2002. Inference for multi-state models from interval-censored data. *Stat. Methods Med. Res.* 11:167–82
Commenges D, Gégout-Petit A. 2007. Likelihood for generally coarsened observations from multistate or counting process models. *Scand. J. Stat.* 34:432–50
Commenges D, Gégout-Petit A, Liquet B. 2007. Choice between semi-parametric estimators of Markov and non-Markov multi-state models from coarsened observations. *Scand. J. Stat.* 34:33–52
Commenges D, Joly P, Letenneur L, Dartigues JF. 2004. Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Stat. Med.* 23:199–210
Cook RJ, Bergeron PJ. 2011. Information in the sample covariate distribution in prevalent cohorts. *Stat. Med.* 30:1397–1409
Courgeau D. 2010a. Dispersion of measurements in demography: a historical view. *Electr. J. Hist. Prob. Stat.* 6:1–19
Courgeau D. 2010b. *Multilevel Synthesis. From the Group to the Individual*. Dordrecht, Neth.: Springer
Courgeau D, Lelièvre E. 1992. *Event History Analysis in Demography*. Oxford, UK: Clarendon
Cox DR. 1972. Regression models and life tables (with discussion). *J. R. Stat. Soc. B* 34:187–220
Cox DR. 1975. Partial likelihood. *Biometrika* 62:269–76
Cox DR, Oakes D. 1984. *Analysis of Survival Data*. London: Chapman & Hall
Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. 2013. Methods for dealing with time-dependent confounding. *Stat. Med.* 32:1584–618
Datta S, Satten GA. 2001. Validity of the Aalen Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for non-Markov models. *Stat. Prob. Lett.* 55:403–11
Datta S, Satten GA. 2002. Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring. *Biometrics* 58:792–802
Didelez V. 2000. *Graphical Models for Event History Data Based on Local Independence*. Berlin: Logos
Didelez V. 2007. Graphical models for composable finite Markov processes. *Scand. J. Stat.* 34:169–85
Didelez V. 2008. Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. B* 70:245–64
Diggle P, Farewell D, Henderson R. 2007. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal (with discussion). *Appl. Stat.* 56:499–550
Eichler M, Didelez V. 2010. On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* 16:3–32

Elbers C, Ridder G. 1982. True and spurious duration dependence: the identifiability of the proportional hazard model. *Rev. Econ. Stud.* 49:403–9

Fine JP, Gray RJ. 1999. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* 94:496–509

Fix E, Neyman J. 1951. A simple stochastic model of recovery, relapse, death and loss of patients. *Hum. Biol.* 23:205–41

Fleming TR, Harrington DP. 1991. *Counting Processes and Survival Analysis.* New York: Wiley

Florens J-P, Fougere D. 1996. Noncausality in continuous time. *Econometrica* 64:1195–212

Frydman H. 1992. A non-parametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *J. R. Stat. Soc. B* 54:853–66

Frydman H. 1995. Nonparametric estimation of a Markov "illness-death" process from interval-censored observations, with applications to diabetes survival data. *Biometrika* 82:773–89

Frydman H, Gerds T, Grøn R, Keiding N. 2013. Nonparametric estimation in an "illness-death" model when all transition times are interval censored. *Biom. J.* 55:823–43

Frydman H, Szarek M. 2009. Nonparametric estimation in a Markov "illness-death" process from interval censored observations with missing intermediate transition status. *Biometrics* 65:143–51

Gégout-Petit A, Commenges D. 2010. A general definition of influence between stochastic processes. *Lifetime Data Anal.* 16:33–44

Gerds TA, Scheike TH, Andersen PK. 2012. Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Stat. Med.* 31:3921–30

Gerster M, Ejrnæs M, Keiding N. 2013. The causal effect of educational attainment on completed fertility for a cohort of Danish women—Does feedback play a role? *Stat. Biosci.* In press. doi: 10.1007/s12561-013-9102-0

Gómez G, Calle ML, Oller R, Langohr K. 2009. Tutorial on methods for interval-censored data and their implementation in R. *Stat. Model.* 9:259–97

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–38

Graw F, Gerds T, Schumacher M. 2009. On pseudo-values for regression analysis in multi-state models. *Lifetime Data Anal.* 15:241–55

Grüger J, Kay R, Schumacher M. 1991. The validity of inferences based on incomplete observations in disease state models. *Biometrics* 47:595–605

Guihenneuc-Jouyaux C, Richardson S, Longini IM Jr. 2000. Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. *Biometrics* 56:733–41

Guilloux A. 2007. Nonparametric estimation for censored lifetimes suffering from an unknown selection bias. *Math. Methods Stat.* 16:202–16

Hardin JW. 2003. The sandwich estimate of variance. *Adv. Econ.* 17:45–73

Hardin JW, Hilbe JM. 2001. *Generalized Linear Models and Extensions*. College Station, TX: Stata

Hoem JM, Jensen UF. 1982. Multistate life table methodology: a probabilist critique. In *Multidimensional Mathematical Demography*, ed. KC Land, A Rogers, pp. 155–264. New York: Academic

Hoem JM. 1976. The statistical theory of demographic rates. A review of current developments (with discussion). *Scand. J. Stat.* 3:169–85

Hougaard P. 1999. Multi-state models: a review. *Lifetime Data Anal.* 5:239–64

Hougaard P. 2000. *Analysis of Multivariate Survival Data*. New York: Springer

Hougaard P, Myglegaard P, Borch-Johnsen K. 1994. Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics* 50:1178–88

Huber PJ. 1967. The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Berkeley Symp. Math. Stat. Probab., 5th*, Vol. 1, pp. 221–33. Berkeley, CA: Univ. Calif. Press

Ibrahim JG, Chen M-H, Sinha D. 2001. *Bayesian Survival Analysis*. New York: Springer

Jacobsen M. 1982. *Statistical Analysis of Counting Processes.* Lecture Notes in Statistics, Vol. 12. New York: Springer

Joly P, Commenges D, Helmer C, Letenneur L. 2002. A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* 3:433–43

Joly P, Durand C, Helmer C, Commenges D. 2009. Estimating life expectancy of demented and institutionalized subjects from interval-censored observations of a multi-state model. *Stat. Model.* 9:345–60
Joly P, Gerds TA, Qvist V, Commenges D, Keiding N. 2012. Estimating survival of dental fillings based on interval censored data and multi-state models. *Stat. Med.* 31:1139–49
Kalbfleisch JD, Prentice RL. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley
Kalbfleisch JD, Prentice RL. 2002. *The Statistical Analysis of Failure Time Data*. New York: Wiley. 2nd ed.
Kaplan EL, Meier P. 1958. Non-parametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–81
Kauermann G, Carroll RJ. 2001. The sandwich variance estimator: efficiency properties and coverage probability of confidence intervals. *J. Am. Stat. Assoc.* 96:1386–97
Keiding N. 1991. Age-specific incidence and prevalence: a statistical perspective (with discussion). *J. R. Stat. Soc. A* 154:371–412
Keiding N. 1992. Independent delayed entry (with discussion). In *Survival Analysis: State of the Art*. ed. JP Klein, PK Goel, pp. 309–26. Dordrecht: Kluwer
Keiding N. 2006. Event history analysis and the cross-section. *Stat. Med.* 25:2343–64
Keiding N. 2013. Event history analysis: local dependence and cross-sectional sampling. In *The Demography of Europe*, ed. G Neyer, G Andersson, H Kulu, L Bernardi, C Bühler, pp. 207–20. Dordrecht: Springer
Keiding N, Andersen PK, Klein JP. 1997. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat. Med.* 16:215–24
Keiding N, Clayton D. 2013. Standardization and control for confounding in observational studies: a historical perspective. *Stat. Sci.* In press
Keiding N, Fine JP, Hansen OH, Slama R. 2011. Accelerated failure time regression for backward recurrence time and current durations. *Stat. Probab. Lett.* 81:724–29
Keiding N, Hansen OHH, Sørensen DN, Slama R. 2012. The current duration approach to estimating time to pregnancy (with discussion). *Scand. J. Stat.* 39:185–213
Keiding N, Holst C, Green A. 1989. Retrospective estimation of diabetes incidence from information in a current prevalent population and historical mortality. *Am. J. Epidemiol.* 130:588–600
Keiding N, Knuiman MW. 1990. Letter to the editor on "Survival analysis in natural history studies of disease" by A. Cnaan & L. Ryan. *Stat. Med.* 9:1221–22
Keiding N, Kvist K, Hartvig H, Tvede M, Juul S. 2002. Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* 3:565–78
Klein JP, Andersen PK. 2005. Regression modelling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 61:223–29
Klein JP, Keiding N, Shu Y, Szydlo RM, Goldman JM. 2000. Summary curves for patients transplanted for chronic myeloid leukemia salvaged by a donor lymphocyte infusion: the current leukemia-free survival cure. *Br. J. Haematol.* 109:148–52
Klein JP, Moeschberger ML. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer
Komárek A, Lesaffre E. 2008. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *J. Am. Stat. Assoc.* 103:523–33
Kurland BF, Johnson LL, Egleston BL, Diehr PH. 2009. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Stat. Sci.* 24:211–22
Lancaster T, Nickell SJ. 1980. The analysis of re-employment probabilities for the unemployed. *J. R. Stat. Soc. A* 143:141–65
Lawless JF. 2003. *Statistical Models and Methods for Lifetime Data.* New York: Wiley. 2nd ed.
Lesaffre E, Komarek A, Declerck D. 2005. An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Stat. Methods Med. Res.* 14:539–52
Liestøl K, Andersen PK. 2002. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Stat. Med.* 21:3701–14
Lin X. 2007. Estimation using penalized quasilikelihood and quasi-pseudo-likelihood in Poisson mixed models. *Lifetime Data Anal.* 13:533–44
Lund J. 2000. Sampling bias in population studies: how to use the Lexis diagram. *Scand. J. Stat.* 27:589–604

Martinussen T, Scheike TH. 2006. *Dynamic Regression Models for Survival Data*. New York: Springer
McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Boca Raton, FL: Chapman & Hall/CRC. 2nd ed.
McLain AC, Ghosh SK. 2011. Nonparametric estimation of the conditional mean residual life function with censored data. *Lifetime Data Anal.* 17:514–32
Meira-Machado L. 2011. Inference for non-Markov multi-state models: an overview. *REVSTAT* 9:83–98
Meira-Machado L, Ũna-Álvarez J, Cadarso-Suáres C. 2006. Nonparametric estimation of transition probabilities in a non-Markov illness death model. *Lifetime Data Anal.* 12:325–44
Meira-Machado L, Ũna-Álvarez J, Cadarso-Suáres C, Andersen PK. 2009. Multi-state models for the analysis of time-to-event data. *Stat. Methods Med. Res.* 18:195–22
Mills M. 2011. *Introducing Survival and Event History Analysis*. Los Angeles: Sage
Molenberghs G, Verbeke G, Demétrio CGB. 2007. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Anal.* 13:513–31
Oakes D, Dasu T. 1990. A note on residual life. *Biometrika* 77:409–10
Ogata Y, Katsura K, Keiding N, Holst C, Green A. 2000. Empirical Bayes age-period-cohort analysis of retrospective incidence data. *Scand. J. Stat.* 27:415–32
Ogle W. 1885. Letter to the Registrar-General on the mortality in the registration districts of England and Wales during the ten years 1871–80. In *Supplement to the Forty-Fifth Annual Report of the Registrar-General of Births, Deaths, and Marriages in England*, pp. iii–lxiv. London: Eyre & Spottiswoode
Pearl J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.
Pepe MS. 1991. Inference for events with dependent risks in multiple endpoint studies. *J. Am. Stat. Assoc.* 86:770–78
Putter H, Fiocco M, Geskus RB. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Stat. Med.* 26:2389–430
Ripley BD, Solomon PJ. 1995. Correspondence: statistical methods for prevalent cohort data. *Biometrics* 51:373–75
Robins JM, Blevins D, Ritter G, Wulfsohn M. 1992. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology* 3:319–36
Robins JM, Hernán MA. 2009. Estimation of the causal effects of time-varying exposures. In *Longitudinal Data Analysis*, ed. G Fitzmaurice, M Davidian, G Verbeke, G Molenberghs, pp. 553–99. New York: Chapman & Hall/CRC
Robins J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Math. Model.* 7:1393–512
Rodríguez-Girondo M, Ũna-Álvarez J. 2012. A nonparametric test for Markovianity in the illness-death model. *Stat. Med.* 31:4416–27
Satten GA, Datta S. 2002. Marginal estimation for multi-state models: waiting time distributions and competing risks analyses. *Stat. Med.* 21:3–19
Scheike T, Keiding N. 2006. Design and analysis of time to pregnancy. *Stat. Methods Med. Res.* 15:127–40
Scheike, TH, Zhang MJ. 2007. Direct modelling of regression effects for transition probabilities in multistate models. *Scand. J. Stat.* 34:17–32
Scheike, TH, Zhang M-J. 2008. Flexible competing risks regression modelling and goodness-of-fit. *Lifetime Data Anal.* 14:464–83
Scheike, TH, Zhang M-J, Gerds TA. 2008. Predicting cumulative incidence probability by direct binomial regression. *Biometrika* 95:205–20
Schweder T. 1970. Composable Markov processes. *J. Appl. Probab.* 7:400–10
Slama R, Ducot B, Carstensen L, Lorente C, de La Rochebrochard E, et al. 2006. Feasibility of the current duration approach to study human fecundity. *Epidemiology* 17:440–49
Slama R, Hansen OKH, Ducot B, Bohet A, Sorensen D, et al. 2012. Estimation of the frequency of involuntary infertility on a nation-wide basis. *Hum. Reprod.* 27:1489–98
Sun J. 2006. *The Statistical Analysis of Interval-Censored Failure Time Data*. New York: Springer
Sverdrup E. 1965. Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skand. Akt. Tidsskr.* 48:184–211
Therneau TM, Grambsch PM. 2000. *Modelling Survival Data: Extending the Cox Model*. New York: Springer

Tomé RA, Giráldez MSO, de Uña-Álvarez J. 2006. Estimación de la duración residencial a partir del period de ocupación declarado por los hogares españoles. *Hacienda Públ. Esp.* 176:87–105
Trussell J, Hankinson R, Tilton J, eds. 1992. *Demographic Applications of Event History Analysis*. Oxford, UK: Clarendon
Tuma NV, Hannan MT, Groeneveld LP. 1979. Dynamic analysis of event histories. *Am. J. Sociol.* 84:820–54
Turnbull BW. 1976. Empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Stat. Soc. B* 38:290–95
Vaupel JW, Manton KG, Stallard E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16:439–54
Wang M-C, Brookmeyer R, Jewell NP. 1993. Statistical models for prevalent cohort data. *Biometrics* 49:1–11
White H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25
Xu J, Kalbfleisch JD, Tai B. 2010. Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics* 66:716–25
Yamaguchi K. 1991. *Event History Analysis.* New York: Sage
Yamaguchi K. 2003. Accelerated failure-time mover-stayer regression models for the analysis of last episode data. *Sociol. Methodol.* 33:81–110

Annual Review of Statistics and Its Application

Volume 1, 2014

Contents