PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Statistical Inference in the Lexis Diagram

Niels Keiding

Phil. Trans. R. Soc. Lond. A 1990 **332**, 487-509
doi: 10.1098/rsta.1990.0128

Statistical inference in the Lexis diagram

By Niels Keiding

Statistical Research Unit, Faculty of Medicine, University of Copenhagen, Blegdamsvej 3, Copenhagen, Denmark, DK-2200

The Lexis diagram is a (time, age) coordinate system, representing individual lives by line segments of unit slope, joining (time, age) of birth and death. The main theme of this paper is non-parametric continuous-time statistical analysis on the Lexis diagram, where I indicate some possible approaches within modern survival analysis. I also introduce the history of the diagram, point processes on the diagram, and the classical statistical approach based on piecewise constant intensities. The Lexis diagram is also useful for describing morbidity, and the methodology is illustrated by two Danish studies of diabetes incidence.

1. Introduction

The statistical description of occurrence of events (such as death or disease incidence) in calendar time and age was facilitated by the graphical representation very carefully discussed by Lexis (1875) and since then termed the Lexis diagram. This is simply a coordinate system with calendar time (henceforth denoted 'time') as abscissa and age as ordinate. Individuals are represented by line segments of slope 1 joining (time, age) at birth and at death, see figure 1. Deaths are represented by points in this diagram and the deaths of a certain population could be seen as a realization of a bivariate point process on the Lexis diagram. The purpose of this presentation is to discuss the interpretation of a continuously varying (time, age)-dependent mortality rate (death intensity) as the intensity of such a point process and, in particular, the statistical estimation of this intensity.

The paper is organized as follows: §2 reviews Lexis's original introduction and discussion of the diagram. We also briefly survey recent contributions to deterministic population mathematics (the 'Lexis surface') including useful differential equations. In §3 the basic point process on the Lexis diagram is introduced and a number of counting processes and associated martingales are defined on it.

For statistical inference, a common procedure has been to assume the death intensity piecewise constant; special generalized linear models called Poisson regression models are then available, as surveyed in §4. A classical topic in this framework is the decomposition of the (time, age)-dependent intensities into products of contributions depending only on time and only on age (and perhaps a third factor depending only on 'cohort', that is, birth time). The main topic of the paper is non-parametric estimation in continuous time.

In §5 I indicate how various recent developments in non-parametric continuous-time survival analysis might be adapted for this purpose, and I discuss particularly

Phil. Trans. R. Soc. Lond. A (1990) **332**, 487–509
Printed in Great Britain

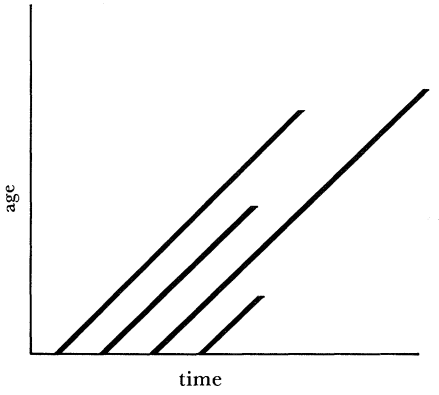


Figure 1. A Lexis diagram.

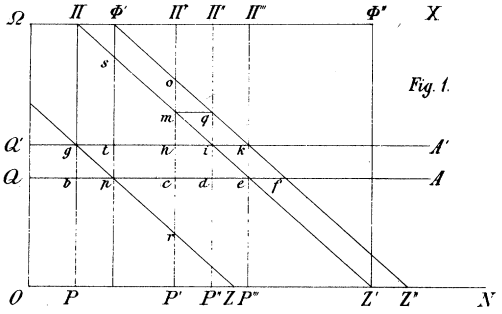
how far multiplicative decompositions of the type discussed in §4 are feasible in this new framework.

So far the exposition has been in terms of mortality, but it is obvious that morbidity may be studied by the same means. In §6 I illustrate some of the methods discussed in the previous sections by two Danish data-sets on diabetes incidence. One set comprises all Danish males born in 1950–64, including diabetics with onset before age 20. The other set of data is methodologically somewhat more complicated. All surviving diabetics on 1 July 1973 in Fyn county, Denmark, were recorded and their (time, age) of onset were registered. For those with onset before age 31 years additional mortality data made a retrospective estimation of diabetes incidence feasible.

The exposition considers no other heterogeneity of the population than that described by time and age, although the regression models in §§4 and 5 could easily be amended to accommodate other (known) covariates. From the elaborate practical experience with the piecewise constant intensity models it is well known that residual heterogeneity may lead to overdispersion relative to the simple 'Poisson' models surveyed in §4. Efforts of incorporating such effects in models with continuously varying intensities are much more preliminary, and in particular little practical experience exists.

2. The Lexis diagram: history and some deterministic theory

Lexis (1875, Fig. 1) suggested the graphical representation in figure 2. Here the abscissa is time of birth ('cohort' as we would say today) and the ordinate is age. Individuals are represented by vertical lines, and deaths happening at the same (calendar) time are on a line with slope -1 (an 'isochronic' line). The original Lexis diagram was thus a (cohort, age)-diagram rather than the (time, age)-diagram used today and exemplified in figure 1. Lexis pointed out that perhaps the right-angled basic triangles were less useful than equilateral triangles that would be obtained by inclining the ordinate axis to get a 60° angle to the abscissa, cf. figure 3, which is Lexis's Fig. 2. In the equilateral triangles, one year is represented by the same length whether it is travelled in the cohort or birth time direction for fixed age (horizontally), in the age direction for fixed cohort (angle 60° to abscissa) or in the age direction for



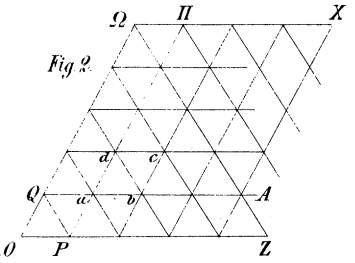
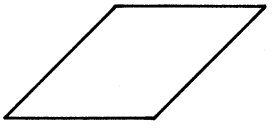


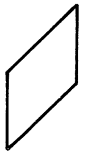
Figure 2. Lexis's diagram (Lexis 1875, fig. 1).
Figure 3. Lexis's equilateral diagram (Lexis 1875, fig. 2).

fixed (calendar) time (angle 120° to abscissa). In Lexis's diagram the latter direction has one year represented as $\sqrt{2}$ times that of the other two; in the modern Lexis diagram the age for fixed cohort direction is the longer.

Lexis's main use of the diagram was in providing a very detailed and careful discussion of the various categories of living and dead, delineated by rectangles or parallellograms in the diagram. The *three principal sets of dead* were defined as follows. Dead individuals in a first principal set were born in a given period and died between given age limits; for example, the rectangle *bchg* in figure 2. In modern Lexis diagrams this corresponds to the parallellogram



Dead individuals in a second principal set were born in a given period and died in a given period; for example, the parallellogram *ekom*. In modern Lexis diagrams this corresponds to the parallellogram



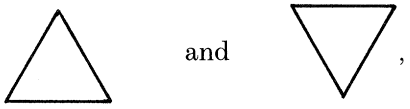
Finally, dead individuals in a third principal set died in a given period, between given age limits, example: the parallellogram *peig*. In modern Lexis diagrams this corresponds to rectangles. Of course all principal sets are unions of certain triangles: in Lexis's diagram

 and  ,

in modern Lexis diagrams

 and 

(and in the equilateral diagram



note again the aesthetic superiority) which are called *lower* and *upper* elementary sets (or elementary triangles). Lexis went on to discuss, equally carefully, the possibilities of obtaining exact counts in the various categories from censuses, and which approximations are necessary in some cases. (The Lexis diagram is still essential in this capacity in demography, in textbooks as well as in daily practice.) For populations with more than one change of state (Lexis studied the example entering and dissolving marriage) Lexis introduced a three-dimensional diagram yielding a 'stereometric' representation. The latter was introduced 'primarily of theoretical interest' and is still rarely met (although I know that Dr P. Cooper uses it for teaching the able–disabled death transitions to actuarial students in Southampton).

Although Lexis concluded his text with a characteristically precise and careful chapter on the application of probability to population change, he did not present a proper stochastic theory for the Lexis diagram.

The formalization of population mathematics as well as careful and specific advice for collection of vital statistics was a topic of research for several of Lexis's contemporary Germany statisticians. Lexis himself mentioned primarily Knapp (1868), who gave a stringent continuous-time theory using differential calculus, and introduced the three principal sets.

Zeuner (1869) gave a particularly elegant exposition including a series of graphs of a cohort–age plane with a continuous curved surface in the third dimension, representing the density of the living. Thus, for example, cohort and period survival curves are obtained as intersections of this surface with the relevant vertical planes. This surface was recently discussed by Arthur & Vaupel (1984). By using the modern convention of a (time, age) diagram, let the non-negative differentiable function $n(t,a)$ describe population density at (time, age) $= (t,a)$. Distinguish between the three directions of change: time separately, age separately, and time and age in tandem. That is, define the *age-specific growth rate*

$$r(t,a) = [\partial n(t,a)/\partial t]/n(t,a),$$

the *age intensity*

$$\nu(t,a) = -[\partial n(t,a)/\partial a]/n(t,a)$$

and

$$\mu(t,a) = (-[\partial n(t+x,a+x)/\partial x]/n(t,a))_{x=0},$$

where in a closed population, μ is simply the mortality rate. More generally, μ gives the relative rate of change in the density of the population in the cohort direction.

It is now an elementary mathematical fact that $\mu(t,a) = \nu(t,a) - r(t,a)$ and it is readily seen that this is equivalent to the so called Von Foerster equation of population dynamics

$$\partial n(t,a)/\partial t + \partial n(t,a)/\partial a = -\mu(t,a)\, n(t,a)$$

(McKendrick 1926, example 9; Von Foerster 1959). Arthur & Vaupel went on to derive not only classical equations from deterministic stable population theory, but

also general relations in population dynamics outside of the restrictive stable population framework.

3. Point processes on the Lexis diagram

In the usual Lexis diagram (figure 1) individual lives are represented as line segments between (time at birth, 0) and (time, age) of death. Brillinger (1986) assumed that the births happen according to some (possibly non-homogeneous) Poisson process with intensity function $\beta(t)$, and that the life lengths are independent random variables, independent of the birth process, with distribution specified by the death intensity $\mu(t, a)$ given by

$$P\{a \leqslant X < a+h \,|\, X \geqslant a\} \approx \mu(t, a)\, h$$

for $t = \sigma + a$ and h small, where X is the life length of a person born at time σ. Brillinger showed that under these assumptions the bivariate point process of deaths in the Lexis diagram is Poisson with intensity function

$$\lambda(t, a) = \beta(t-a)\, \mu(t, a) \exp\left\{-\int_0^a \mu(t-a+y, y)\, \mathrm{d}y\right\}.$$

Of course these assumptions are not all strictly reasonable for human populations: in particular, no account is taken of the dependence between individuals introduced by the human reproduction process. But Brillinger (1986) argued convincingly that they provide a suitable framework for a sampling theory on the Lexis diagram, in Brillinger's exposition concentrated on classical demographic rates. It should be emphasized that the Poisson distribution in Brillinger's context derives from the postulated poissonian birth process. Thus there is no connection either to the 'rare disease' assumption or to the Poisson likelihood mentioned in §4 below.

Keiding (1990) extended Brillinger's model also to allow (irreversible) morbidity. This provided a framework adapting modern continuous-time survival analysis to the estimation of incidence and mortality from the cross-sectional information on the age distribution of healthy and diseased, possibly supplemented with retrospective information on age at onset, for the survivors in the cross-sectional sample. We shall discuss a concrete calculation of this type in §6. It was mentioned in §2 that Lexis (1875, ch. IV) himself introduced the additional structure of an irreversible transition (in his case marriage rather than morbidity) and derived the logically corresponding three-dimensional Lexis diagram.

While Brillinger's assumption that a person born at t–a and aged a has (continuous) death intensity $\mu(t, a)$ and that individual life lengths are independent is canonical for almost any stochastic model on the Lexis diagram, the assumptions on the birth process are more debatable, and indeed one might often prefer to condition on the actually realized births.

A more general point process description of the Lexis diagram was given by Capasso (1988), motivated by Kendall's (1949) age-dependent birth-and-death process. By using the Lexis diagram formalism, in contrast to Capasso, we may paraphrase his approach as follows. The population is described by the random measure U on the Lexis diagram in the sense that

$$\int_A U(t, \mathrm{d}a)$$

denotes the number of individuals alive at time t with age $\in A$. (That is, the number of intersections with lines in the Lexis diagram of a vertical line segment at t covering ages $\in A$.) Similarly,

$$\int_T U(\mathrm{d}t, a)$$

denotes the total number of individuals of exact age a at times $\in T$. In particular, $B(\mathrm{d}t) = U(\mathrm{d}t, 0)$ is what could be termed the 'birth measure'. (Capasso considered in effect a time-left-truncated Lexis diagram and therefore also worried about individuals who had not been followed from birth; we shall not discuss this added complication.)

Capasso used this structure to derive the classical Lotka renewal equation from a stochastic model; his main interest was, however, to obtain martingale dynamics for statistical inference, as we shall discuss in §5 below. Earlier continuous-time stochastic population models generalizing the age-dependent birth and death processes were discussed by Keiding & Hoem (1976) (cf. Jagers 1975, ch. 8), who also discussed the renewal equations.

4. Piecewise constant intensity models

Assume that the death intensity $\mu(t, a)$ is constant ($= \mu$) over some region Ω in the Lexis diagram. For each individual i let T_i be the time lived in Ω, geometrically $(\sqrt{2})^{-1}$ times the length of the intersection of Ω and the lifeline of i. If i died in Ω, the likelihood contribution for μ is $\mu \exp(-\mu T_i)$, if i did not die in Ω (perhaps still being alive), it is $\exp(-\mu T_i)$. In general, taking D_i as the indicator that i die in Ω, the likelihood based on observation of one individual i is $\mu^{D_i} \exp(-\mu T_i)$ and that based on independent individuals, all with death intensity μ,

$$L(\mu) = \mu^D \exp(-\mu T),$$

where $D = \Sigma D_i$ is the number of deaths in Ω and $T = \Sigma T_i$ the total time lived in Ω. The maximum likelihood estimator is $\hat{\mu} = D/T$ with large-sample variance approximation estimated by $\widehat{\mathrm{var}}(\hat{\mu}) = \{I(\mu)\}^{-1} = D/T^2$ where $I(\mu) = -\mathrm{d}^2 \log L(\mu)/\mathrm{d}\mu^2 = D/\mu^2$ is the observed information.

This analysis holds whether or not the individuals are assumed to be born in Brillinger's Poisson birth process described in §3, as long as they are independent. In Brillinger's case D is Poisson distributed, otherwise not necessarily so.

The likelihood L is the same as would have been obtained by assuming D Poisson distributed with mean μT and the model based on L and its generalizations are therefore often called Poisson (regression) models. However, in our context T is a random variable and the conditional distribution of D given T is not Poisson, not even in Brillinger's case.

In practice, aggregate data from official vital statistics are often used in which case T, the total time lived in Ω, will not be directly observable. For the three classical 'principal sets' in the Lexis diagram, cf. §2, Hoem (1969) gave specific approximation formulae for T based on various assumptions on observational patterns much in the style of Lexis as mentioned in §2 above. Consider as an example the mortality in 1990 of persons born in 1920, and disregard migration. This is Lexis's (Knapp's) *second* principal set of dead, corresponding to the parallellogram of figure 4. If the total time lived (T) is not directly observed, perhaps the number M of persons attaining age 70 in 1990 is available (that is, the number of lifelines crossing AB in figure 4). A

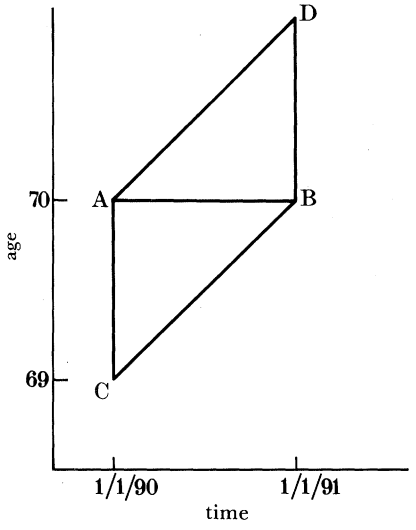


Figure 4. Approximation of the total time lived. If the total time lived in the parallellogram ACBD is not available, an approximation may be obtained by the number of life lines crossing AB times 1 year. If even this is not available, an approximation would be the average of the number of life lines crossing CA and BD respectively.

reasonable approximation would then be $T = M \cdot 1$ person years. If M is not available, perhaps the number $L_{69}(90)$ of 69 year olds at the start of 1990 and the number $L_{70}(91)$ of 70 year olds at the start of 1991 are available (number of lifelines crossing CA and BD respectively). Approximate T by $\{L_{69}(90)+L_{70}(91)\}/2(\times 1)$.

It is obvious that the choice of principal set over which to assume the death intensity constant is arbitrary in the first place. This is one of the complications behind the widespread use of the Poisson regression models for describing disease incidence, particularly cancer incidence.

Continuing to use the language of mortality assume now that a tessellation of the Lexis diagram in terms of some principal set is given and that $\mu_{t,a}$ represents the constant death intensity around (time, age) $= (t, a)$, which for this purpose varies across some suitable (not necessarily rectangular) lattice. To obtain further insight into the variation of $\mu_{t,a}$ across the lattice a multiplicative age–period model

$$\mu_{t,a} = \alpha_t \beta_a$$

can be useful. Or perhaps an age–cohort model

$$\mu_{t,a} = \gamma_a \delta_c, \quad c = t-a$$

is more suitable. The obvious general possibility of an age–period–cohort model,

$$\mu_{t,a} = \epsilon_t \phi_a \kappa_c,$$

has been discussed at some length, particularly in cancer epidemiology, see Osmond & Gardner (1982), who also surveyed earlier work, and Clayton & Schifflers (1987). All of these models preserve the Poisson-type likelihood derived at the beginning of this section and are therefore direct special cases of generalized linear models in the sense of McCullagh & Nelder (1989).

However the age–period–cohort model has a non-trivial identification problem. Assume for simplicity that a rectangular lattice of parameters $\mu_{t,a}$, $t = 1, \ldots T$,

$a = 1, \ldots, A$ is given. Then $c = t - a$ varies across the $A + T - 1$ values $-A + 1, \ldots, T - 1$, and the total number of parameters ϵ_t, ϕ_a, κ_c is $2A + 2T - 1$. But the dimension of the model is only $2A + 2T - 4$, because of the three independent constraints that derive from

$$\epsilon_t \phi_a \kappa_{t-a} = (\epsilon_t \xi \eta^t)(\phi_a \xi^{-1} \eta^{-a} \zeta)(\kappa_{t-a} \eta^{-(t-a)} \zeta^{-1}).$$

This mathematical fact considerably complicates interpretation of the parameter estimates, and many approaches have been suggested. One of them is to go back to the elementary triangles in the Lexis diagram; the events are usually known by exact day of (time, age) already and the necessary approximations of time lived (if necessary) are not much worse than before. If it is maintained that $\mu_{t,a}$ is still constant across the two elementary triangles into which the relevant principal set has been split, then the age–period–cohort model does become fully identifiable. However, Clayton & Schifflers (1987) argued very carefully, and Carstensen (1990) provided additional mathematical documentation, that there are several objections to this approach, the most important being that the average age, time and birth time of individuals (and of cases) differ between the two triangles.

In my view the piecewise constant intensity models have been stretched to their limit here. It is time to develop statistical methods for the Lexis diagram based on continuous time; this is the topic of §5.

5. Non-parametric continuous-time statistical analysis

The recent development of non-parametric survival analysis in continuous time has been much influenced by the approach via counting processes and martingales pioneered by O. O. Aalen, see Jacobsen (1982) and Andersen & Borgan (1985) for reviews. I briefly review this approach below.

In the simplest situation Aalen considered a univariate counting process $N(t)$ with intensity process $\alpha(t)\, Y(t)$, that is,

$$N(t) - \int_0^t \alpha(s)\, Y(s)\, \mathrm{d}s$$

is a martingale. The predictable process $Y(t)$ usually denotes a 'number at risk' at time t. Heuristically we ignore the martingale 'noise' to obtain an estimating equation

$$\mathrm{d}N(t) = \alpha(t)\, Y(t)\, \mathrm{d}t,$$

yielding the estimator

$$\hat{\alpha}(t)\, \mathrm{d}t = \mathrm{d}N(t)/Y(t).$$

More formally, use the Nelson–Aalen estimator,

$$\hat{\beta}(t) = \int_0^t \frac{I\{Y(s) > 0\}}{Y(s)} \mathrm{d}N(s),$$

as estimator of

$$\beta(t) = \int_0^t \alpha(s)\, \mathrm{d}s.$$

Properties of this estimator are easily available from the fact that $\hat{\beta} - \beta^*$ is a martingale, where

$$\beta^*(t) = \int_0^t I\{Y(s) > 0\}\, \alpha(s)\, \mathrm{d}s.$$

Asymptotically, one will usually have $Y(t) > 0$ for all t, and hence $\beta^*(t) = \beta(t)$.

A basic trick behind this non-parametric estimator is that the use of the cumulative intensity allows a discrete measure as estimator of an absolutely continuous one; the consistency and asymptotic normality results guarantee the coherence, and plots of $\hat{\beta}(t)$ are also directly interpretable. However, to arrive at an estimator of $\alpha(t)$ itself, some form of smoothing is required, such as the kernel estimator (Ramlau-Hansen 1983)

$$\hat{\alpha}(t) = \frac{1}{b}\int K\left(\frac{t-s}{b}\right) \mathrm{d}\hat{\beta}(s),$$

where $K(x)$ is a kernel (probability density symmetric about 0) and b the bandwidth.

In our situation we are concerned with drawing inference about an intensity depending on the two time parameters (calendar) time and age. One might therefore want to know what happens to Aalen's approach in two time dimensions.

From an absolutely continuous bivariate distribution function F with density f Pons (1986) considered what she termed the 'two-dimensional hazard function' $\alpha(s,t) = f(s,t)/\{1-F(s,t)\}$; the cumulative hazard is then

$$\beta(s,t) = \int_0^s \int_0^t \alpha(u,v)\,\mathrm{d}v\,\mathrm{d}u.$$

Since
$$I\{S \leqslant s, T \leqslant t\} - \int_0^s \int_0^t I\{S \geqslant u, T \geqslant v\}\,\alpha(u,v)\,\mathrm{d}v\,\mathrm{d}u$$

is a weak martingale, important parts of Aalen's one-dimensional theory may be retained in two dimensions. For n independent replications the bivariate counting process

$$N(s,t) = \sum_{i=1}^{n} I\{S_i \leqslant s, T_i \leqslant t\}$$

has compensator with respect to the product filtration of the self-exciting filtrations generated by $S_1, \ldots, S_n$ and $T_1, \ldots, T_n$,

$$\int_0^s \int_0^t \alpha(u,v)\,Y(u,v)\,\mathrm{d}v\,\mathrm{d}u,$$

where $Y(s,t) = \Sigma I\{S_i \geqslant s, T_i \geqslant t\}$. Hence

$$\hat{\beta}(s,t) = \int_0^s \int_0^t \frac{I\{Y(u,v) > 0\}}{Y(u,v)} \mathrm{d}N(u,v)$$

may be taken as estimator of $\beta(s,t)$, again studying $\hat{\beta} - \beta^*$ with

$$\beta^*(s,t) = \int_0^s \int_0^t I\{Y(u,v) > 0\}\,\alpha(u,v)\,\mathrm{d}v\,\mathrm{d}u$$

and utilizing that $\beta^* = \hat{\beta}$ for large populations. (It is important to notice the interpretation of $\mathrm{d}N(s,t)$: this jumps one when, for some i, $S_i = s$ and $T_i = t$. When the formalism is generalized to censored variables, this implies that $\hat{\beta}$ is based only on the doubly uncensored pairs (S_i, T_i).)

Pons proved asymptotic unbiasedness and asymptotic normality of $\hat{\beta}$ and derived a test for independence of S and T based on $\hat{\beta}$.

Dabrowska (1988) pointed out that Pons's approach uses only some of the information (as explained above) and also that the bivariate distribution function F

cannot be recovered from α: indeed the marginal distributions $F(s,\infty)$ and $F(\infty,t)$ are also required. Dabrowska obtained for bivariate censored data a non-parametric estimator of the bivariate distribution function $F(s,t)$ by combining Pons's estimator and the standard Kaplan–Meier estimators of the marginal distributions; she proved consistency of this estimator.

(a) *Capasso's proposal for Aalen theory on the Lexis diagram*

Pons's estimator solves the heuristic infinitesimal estimating equation

$$\alpha(s,t)\,\mathrm{d}s\,\mathrm{d}t = \mathrm{d}N(s,t)/Y(s,t),$$

which directly generalizes Aalen's idea. For the Lexis diagram, let $N(t,a)$ denote the number of deaths by time t to individuals born at time $t-a$. Then

$$N(t,a) - \int_0^t \mu(s,a-t+s)\,Y(s,a-t+s)\,\mathrm{d}s$$

is a (local) martingale with respect to the filtration spanned by the whole population process before time t, and $Y(t,a)$ is the number of individuals born at time $t-a$ and alive (in more general settings, and uncensored) at time t.

Capasso (1988) attempted to mimic Aalen's approach directly by using the estimating equation

$$\mathrm{d}N(t,a) = \mu(t,a)\,Y(t,a)\,\mathrm{d}t$$

to obtain an estimator

$$\hat{\mu}(t,a) = \mathrm{d}N(t,a)/Y(t,a)$$

as usual interpreted in a cumulative fashion:

$$\hat{\beta}(t) = \int_0^\infty \frac{I\{Y(t,a)>0\}}{Y(t,a)} N(t,\mathrm{d}a),$$

as estimating

$$\beta(t) = \int_0^\infty \mu(t,a)\,\mathrm{d}a.$$

In the original model with continuous intensity $\mu(t,a)$, and assuming also a birth process with continuous intensity, this, however, degenerates: $Y(t,a)$ will always be 0 or 1 and $\hat{\beta}(t)$ will just count the number of deaths at exact time t, no matter how large the population. Capasso assumed discrete initial distributions to obtain asymptotic results. (Note that this may be said to violate the basic absolute continuity assumption that Lexis (1875, p. 26) formulated very specifically. 'As we may, for the distinction between individuals, only use the three timescales, and since one cannot speak about a set of *simultaneously born*, we have to base ourselves either on a set of *simultaneously living* or a set of *living at a certain age*.' As I emphasized in §3, the latter two concepts are directly covered by the absolute continuous point process, the former is not.) Alternatively one might interpret Capasso's idea in the context of piecewise constant intensities, in the spirit of §4 above.

One way of understanding the difficulties in establishing an Aalen theory in the Lexis diagram is that although the diagram is two-dimensional, all movements are in the same direction (slope 1) and in the fully non-parametric model the diagram disintegrates into a continuum of life lines of slope 1 with freely varying intensities across lines. The cumulation trick from Aalen's estimator (generalizing ordinary empirical distribution functions and Kaplan & Meier's (1958) non-parametric empirical distribution function from censored data) does not help us here. We have to directly assume smoothness across life lines, as we shall do in the next section.

(*b*) *Non-parametric regression analysis in the Lexis diagram*

McKeague & Utikal (1990) followed the important paper by Beran (1981) in studying what in the present context amounts to

$$N_M(t,a) = \int_{a-w}^{a+w} N(t,u)\,\mathrm{d}u, \quad Y_M(t,a) = \int_{a-w}^{a+w} Y(t,u)\,\mathrm{d}u,$$

so that $N_M(t,a)$ counts the number of deaths before time t to those born in $[t-a-w, t-a+w]$ and $Y_M(t,a)$ counts the number born in that interval and still at risk at time $t-$. Still keeping a fixed, define the Nelson–Aalen type estimator

$$\hat{M}(t,a) = \int_0^t \frac{I\{Y_M(s,a)>0\}}{Y_M(s,a)} N_M(\mathrm{d}s,a).$$

From here kernel smoothing yields

$$\hat{\mu}(t,a) = \frac{1}{b}\int K\left(\frac{t-s}{b}\right)\hat{M}(\mathrm{d}s,a)$$

and, finally, smoothing in the age direction yields a smooth estimator

$$\tilde{\mu}(t,a) = \frac{1}{\tilde{b}}\int \tilde{K}\left(\frac{a-u}{\tilde{b}}\right)\hat{\mu}(t,u)\,\mathrm{d}u$$

of μ. McKeague & Utikal went on to derive consistency and asymptotic normality for $\hat{M}$, $\hat{\mu}$ and $\tilde{\mu}$. As usual this was done by assuming that some parameter n (which we can take as some population size) increases to infinity, while $w\to 0$, $b\to 0$, $\tilde{b}\to 0$ at subtly balanced rates. Of particular conceptual importance are the conditions of asymptotic stability of $Y_M(t,a)/(nw)$ which expresses the 'density' of the cohort of individuals born at time $t-a$ (cf. Zeuner 1869; Arthur & Vaupel 1984). The asymptotic results yield approximative standard errors, in particular $\hat{\mu}(t,a)$ has approximative standard error $\sigma(t,a)/(nw^2)^{\frac{1}{2}}$, where an estimator of

$$\sigma^2(t,a) = h(t,a)\int K^2(u)\,\mathrm{d}u$$

is obtained from

$$\hat{h}(t,a) = \frac{nw}{b}\int K\left(\frac{t-s}{b}\right)\frac{N_M(\mathrm{d}s,a)}{Y_M(\mathrm{d}s,a)^2}.$$

McKeague & Utikal obtained their final smoothed estimator $\tilde{\alpha}(t,a)$ via *three* smoothings. First the original counting process $N(t,a)$ and at risk process $Y(t,a)$ are smoothed in the age direction, using essentially a uniform kernel with bandwidth w. The resulting Nelson–Aalen type (discrete) estimator $\mathrm{d}\hat{M}(t,a)$ is then smoothed in the time direction to obtain $\hat{\mu}(t,a)$, which is finally smoothed in the age direction.

At least in the present Lexis diagram context it would seem more direct to do just one bivariate smoothing, as follows. Let $K_Y(t,a)$ and $K_N(t,a)$ be bivariate kernel functions (probability densities) and b_Y and b_N and bandwidths and define

$$Y_K(t,a) = \frac{1}{b_Y^2}\iint K_Y\left(\frac{t-s}{b_Y},\frac{a-u}{b_Y}\right)Y(s,u)\,\mathrm{d}u\,\mathrm{d}s,$$

$$\hat{\mu}_K(t,a) = \frac{1}{b_N^2}\iint K_N\left(\frac{t-s}{b_N},\frac{a-u}{b_N}\right)\frac{I\{Y_K(s,u)>0\}}{Y_K(s,u)}N(\mathrm{d}s,\mathrm{d}u).$$

I conjecture that suitable regularity conditions could be worked out to ensure that $\hat{\mu}_K$ is consistent and asymptotically normal. In the application of §6*a* Y is so large that it seems safe to disregard its random variation in a first approximation to an estimate of $\mathrm{var}\{\hat{\mu}_K(t,a)\}$; indeed we use

$$\hat{\sigma}^2_K(t,a) = \frac{1}{b_N^2}\iint K_N^2\left(\frac{t-s}{b_N}, \frac{a-u}{b_N}\right)\frac{I\{Y_K(s,u)>0\}}{Y_K(s,u)}N(\mathrm{d}s,\mathrm{d}u).$$

(*c*) *Age–period models in continuous time*

As discussed in §4 above for piecewise constant intensity models, there has been considerable interest in multiplicative decompositions of death intensities $\mu(t,a)$ or disease incidence $\alpha(t,a)$ into products of factors depending only on time and age (age–period models), time of birth and age (age–cohort models) or time, time of birth and age (age–period–cohort models). In this section I shall indicate some possibilities for obtaining similar decompositions in the models with continuously varying intensity. We concentrate on the age–period models as an example.

The multiplicative decomposition

$$\mu(t,a) = \xi(t)\,\lambda(a)$$

was studied by McKeague & Utikal (1988). Their study was based on the double cumulative intensity

$$\mathscr{M}(t,a) = \int_0^t\int_0^a \mu(s,u)\,\mathrm{d}u\,\mathrm{d}s,$$

with estimator

$$\hat{\mathscr{M}}(t,a) = \int_0^a \hat{M}(t,u)\,\mathrm{d}u = \int_0^a\int_0^t \frac{I\{Y_M(s,a)>0\}}{Y_M(s,a)}N_M(\mathrm{d}s,a).$$

The marginal cumulative intensities

$$\Lambda(a) = \int_0^a \lambda(u)\,\mathrm{d}u, \quad \Xi(t) = \int_0^t \xi(s)\,\mathrm{d}s$$

may be estimated by

$$\hat{\Lambda}(a) = \hat{\mathscr{M}}(\infty,a)/\hat{\mathscr{M}}(\infty,\infty), \quad \hat{\Xi}(t) = \hat{M}(t,\infty)$$

(resolving the indeterminacy in the parametrization by imposing the constraint $\Lambda(\infty)=1$). McKeague & Utikal proved asymptotic results for these estimators, in particular for the residual process $\hat{\mathscr{M}} - \hat{\Lambda}\hat{\Xi}$ that may be used for testing the multiplicativity hypothesis.

In my view the suggestion, essentially of using the marginal cumulative intensities as estimators, will in practice put some rather restrictive implicit orthogonality conditions on the design. As we shall see in the examples in §6, these conditions are sometimes far from fulfilled.

Other suggestions for fitting the 'non-parametric' multiplicative intensity model

$$\mu(t,a) = \lambda(a)\,\xi(t)$$

are based on likelihood methods generalizing those of the Cox (1972) 'semi-parametric' regression model. Before discussing these, we want to dwell a little on the Cox model itself.

(*d*) *The Cox model in the Lexis diagram: Sellke & Siegmund's approximate martingale construction*

In clinical trials with staggered entry, patients are put on trial as they are diagnosed and followed until death (or recurrence of disease or some other endpoint) or censoring. The time variable of interest is usually duration on trial rather than calendar time, but information is accumulated in calendar time, for (greatly) varying durations since patient entry. A remarkable study of martingale issues connected to the use of the Cox model in this situation is due to Sellke & Siegmund (1983) whose paper is much easier to read when one sketches the obvious (time, duration) Lexis diagram along the way.

To use the Cox model for obtaining an age–period model one has to choose one time variable as basic (to be modelled in the non-parametric part of the model), the other will then enter as ('time-dependent') covariate to be modelled in the parametric regression part. Usually one would assume the age variation to be rather more dramatic than the secular trend in calendar time, and we shall assume accordingly that

$$\mu(t, a) = \lambda(a) \exp(\beta \cdot \boldsymbol{z}(t)),$$

where $\boldsymbol{z}(t)$ is a vector of (known) functions of time t, such as t, t^2, $\log_{10} t$, etc. Note that $\boldsymbol{z}(t)$ is 'age-dependent' in the sense of the Cox model. Let $\boldsymbol{z}_i(a)$ be $\boldsymbol{z}$ evaluated at the time at which individual i is of age a.

The statistical analysis of the regression part of the Cox model is based on the partial likelihood

$$L_{\mathrm{p}} = \prod_{a_i} [\exp(\beta \cdot \boldsymbol{z}_i(a_i)) / \sum_{j \in R_i} \exp(\beta \cdot \boldsymbol{z}_j(a_i))],$$

where the 'risk set' R_i contains those individuals alive and uncensored at age a_i and where the product is over the distinct ages of death a_i.

Sellke & Siegmund phrased their discussion in terms of a scalar covariate function z, hence a scalar regression parameter β. The basic quantity is the score process (the logarithmic derivative of L_{p}) $\dot{l}(t, a, \beta)$, which is a martingale in a for each fixed t with respect to the filtration $\{\mathscr{F}_{t,a}, a > 0\}$ where $\mathscr{F}_{t,a}$ is generated by events observed by time t and of age less than or equal to a. In order to study the joint behaviour of the process at different times (which is at least necessary for the asymptotic theory) it is also necessary to study $\dot{l}(t, a, \beta)$ as function of t; it is not a martingale with respect to $\mathscr{F}_{t,a}$, but Sellke & Siegmund constructed an approximating martingale and went on to prove asymptotic properties of the maximum partial likelihood estimator $\hat{\beta}$ for large t.

(*e*) *Non-parametric regression modifying the Cox model*

We shall briefly survey some recent suggestions for modifying the Cox model from 'semi-parametric'

$$\mu(t, a) = \lambda(a) \exp(\beta \cdot z(t)),$$

to non-parametric, making fewer assumptions on the factor describing the dependence on time

$$\mu(t, a) = \lambda(a)\, \xi(t) = \lambda(a) \exp \eta(t).$$

Thus Thomas (1983) developed a Grenander type isotonic estimator of $\xi(t)$, assuming it to be increasing. Thomas's context was cancer incidence at dose t, where the monotonicity constraint will usually be more natural than for the time effect in the Lexis diagram.

Hastie & Tibshirani (1986) fitted their 'generalized additive models' using one of two alternative data analytic techniques, called local scoring and local likelihood, respectively. Both methods are based on Cox's partial likelihood which in the age-dependent generalization necessary here takes the form

$$L_{\mathrm{p}} = \prod_{a_i} [\exp[\eta\{t_i(a_i)\}] / \sum_{j \in R_i} \exp[\eta\{t_j(a_i)\}]],$$

where $t_i(a)$ is the time at which individual i is of age a and R_i as usual denotes the individuals at risk at the age a_i where individual i was observed to die. Let $l = \log_{10} L_{\mathrm{p}}$ as before, and let $C_i = \{k : i \in R_k\}$ (the risk sets containing individual i), $\delta_i = 1$ if person i dies (in contrast to becoming censored), then

$$\frac{\partial l}{\partial \eta\{t_i(a_i)\}} = \delta_i - \exp[\eta\{t_i(a_i)\}] \sum_{k \in C_i} 1 / \sum_{j \in R_k} \exp[\eta\{t_j(a_i)\}],$$

$$\frac{\partial^2 l}{\partial \eta\{t_i(a_i)\}^2} = -\exp[\eta\{t_i(a_i)\}] \sum_{k \in C_i} 1 \Big/ \sum_{j \in R_k} \exp[\eta\{t_j(a_i)\}]$$

$$+ \exp[2\eta\{t_i(a_i)\}] \sum_{k \in C_i} 1 / [\sum_{j \in R_k} \exp \eta\{t_j(a_i)\}]^2.$$

The local scoring method now uses some 'scatterplot smoother' $\mathscr{S}$ in a Newton–Raphson type iteration:

$$\eta_1(a) = \mathscr{S}\left\{ \eta(a) - \frac{\mathrm{d}l/\mathrm{d}\eta}{\mathscr{S}(\mathrm{d}^2 l/\mathrm{d}\eta^2)} \right\}.$$

Hastie & Tibshirani have developed the GAIM software (available in various versions) to perform these calculations as well as the local likelihood technique (Tibshirani 1984). In this algorithm $\log_{10} \eta(z)$ is assumed to depend linearly on z locally. Accordingly, for each i a local partial likelihood is formed by including in the product only the nearest neighbouring z_j. The ith local partial likelihood yields a regression parameter estimate $\hat{\beta}_i$, and $\eta(t)$ is estimated by some interpolation procedure, given these local slopes of $\log_{10} \eta(t)$.

O'Sullivan (1988) proposed a *penalized partial likelihood* approach as an alternative to Hastie & Tibshirani's data analytic procedures, i.e. assume η twice differentiable and maximize

$$\sum_{a_i} \{\eta\{t_i(a_i)\} - \log \sum_{j \in R_i} \exp[\eta\{t_j(a_i)\}]\} - \kappa \int \{\ddot{\eta}(a)\}^2 \, \mathrm{d}a,$$

where $\ddot{\eta}$ denotes second derivative and the penalty parameter $\kappa > 0$ regulates the smoothness of $\hat{\eta}$. O'Sullivan noted that any solution to this minimization problem is a cubic spline with knots at $\{t_i(a_i)\}$. In a further contribution O'Sullivan (1989) proved asymptotic convergence results for the penalized partial likelihood approach, although to be applicable to the present context, some generalization to time-dependent covariates is needed, and the difficulties pointed out by Sellke & Siegmund would seem to be at least as large here.

(*f*) *Ogata & Katsura's empirical Bayes penalized likelihood approach*

As an alternative to the kernel smoothing methods originating with Beran (1981) and for the present purpose reviewed in §5*b* above one might mention an algorithm

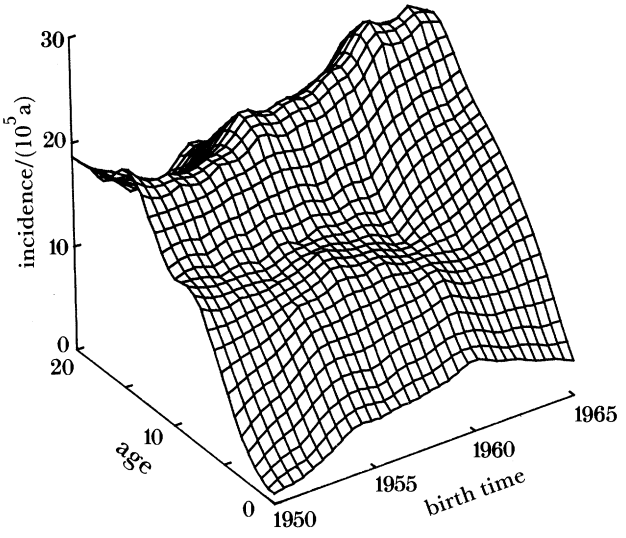


Figure 5. Estimated yearly (time, age)-specific diabetes incidence rates for Danish males born 1950–64, ages 0–19 yrs.

due to Ogata & Katsura (1988) based on penalized likelihood, choosing the penalty parameter by empirical Bayes.

Ogata & Katsura's context was that of a marked point process in the plane (their main motivation being the geographical location and severity of earthquakes). In our context the marks would usually be disregarded (although see the example in §6*b*) and the (inhomogeneous) Poisson process likelihood is

$$\prod_{i=1}^{n} \mu(t_i, a_i) \exp\left\{ -\int_{\Omega} \mu(t, a)\, \mathrm{d}t\, \mathrm{d}a \right\},$$

where (t_i, a_i) are the points in the study region Ω in the Lexis diagram where deaths are observed. The intensity is expressed by the cubic B-spline bases F_i and G_j with equally spaced knots,

$$\log_{10} \mu(t, a) = \sum_{i=1}^{I+3} \sum_{j=1}^{J+3} \gamma_{ij} F_i(t)\, G_j(a),$$

and the task is to estimate $\{\gamma_{ij}\}$. Ogata & Katsura imposed two roughness penalties $\Phi_1(\mu)$ and $\Phi_2(\mu)$ and maximized the penalized log likelihood

$$\log L - \eta_1 \Phi_1(\mu) - \eta_2 \Phi_2(\mu)$$

with respect to $\{\gamma_{ij}\}$. The roughness parameters η_1 and η_2 were chosen by interpreting the penalty terms in the penalized log likelihood as deriving from a prior distribution with (hyper-)parameters η_1 and η_2. This distribution is assumed multivariate normal, allowing a (numerically involved) maximum-likelihood estimation also of η_1 and η_2.

6. Examples

In this final section two sets of data concerning the (time, age)-variation of diabetes incidence will be briefly presented: a prospective (historical) study and a retrospective study, and some preliminary nonparametric estimation calculations reported.

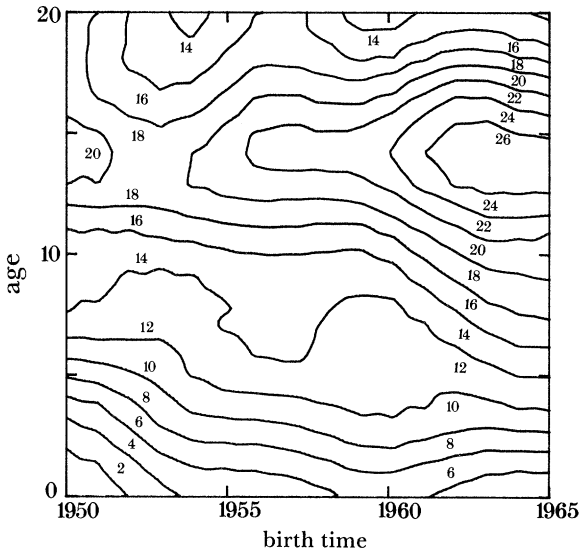


Figure 6. Estimated yearly (time, age)-specific diabetes incidence rates per 10^5 person years for Danish males born 1950–64, ages 0–19 yrs.

(*a*) *Prospective data based on the Danish national service conscript registry*

From the eight Danish male birth cohorts born 1949–56, Green *et al.* (1980) identified those who had diabetes mellitus (DM) before the age of 20, by going through the files of the Danish national service conscript authorities. Since DM leads to unconditional rejection from military duty, these files are expected to have high reliability on DM. Indeed, when supplemented with information on cause of death from the computerized national registry of deaths ascertainment was evaluated to be at least 95%. Green *et al.* (1990) later collected data from a further eight cohorts in a similar fashion. Below we shall use the data from the 15 cohorts 1950–64.

The basic estimation was a bivariate smoothing such as discussed in §5*b* as

$$\hat{\mu}_K(t,a) = \frac{1}{b_N}\int\int K_N\left(\frac{t-s}{b_N}, \frac{a-u}{b_N}\right)\frac{I\{Y_K(s,u)>0\}}{Y_K(s,u)}N(\mathrm{d}s,\mathrm{d}u).$$

We used a bivariate Epanechnikov kernel

$$K(x,y) = 2\pi^{-1}(1-x^2-y^2), \quad x^2+y^2<1$$

and bandwidth $b_N = 3$ years; estimates closer than three years to the boundary were obtained by reflection. The expression Y_K for the size of the risk set was not obtained by bivariate smoothing as suggested in §5*b*; instead the average number of boys alive and unaffected by DM in an elementary 1×1 year square of the Lexis diagram was derived from the very accurate figures of time lived (corrected for emigration, cf. Andersen & Green (1985)) as derived by Green *et al.* (1990) for their Poisson regression analysis.

Figures 5 and 6 show the estimated (time, age)-specific incidence rates. Since the data are really collected in a (cohort, age)-rectangle, they are presented in a (cohort, age)-Lexis diagram (similar to Lexis's original diagram). We see an increase with age until puberty and then a decrease, also an increase with cohort, particularly for the teenagers. Figures 7–9 detail these tendencies by following the trends along line

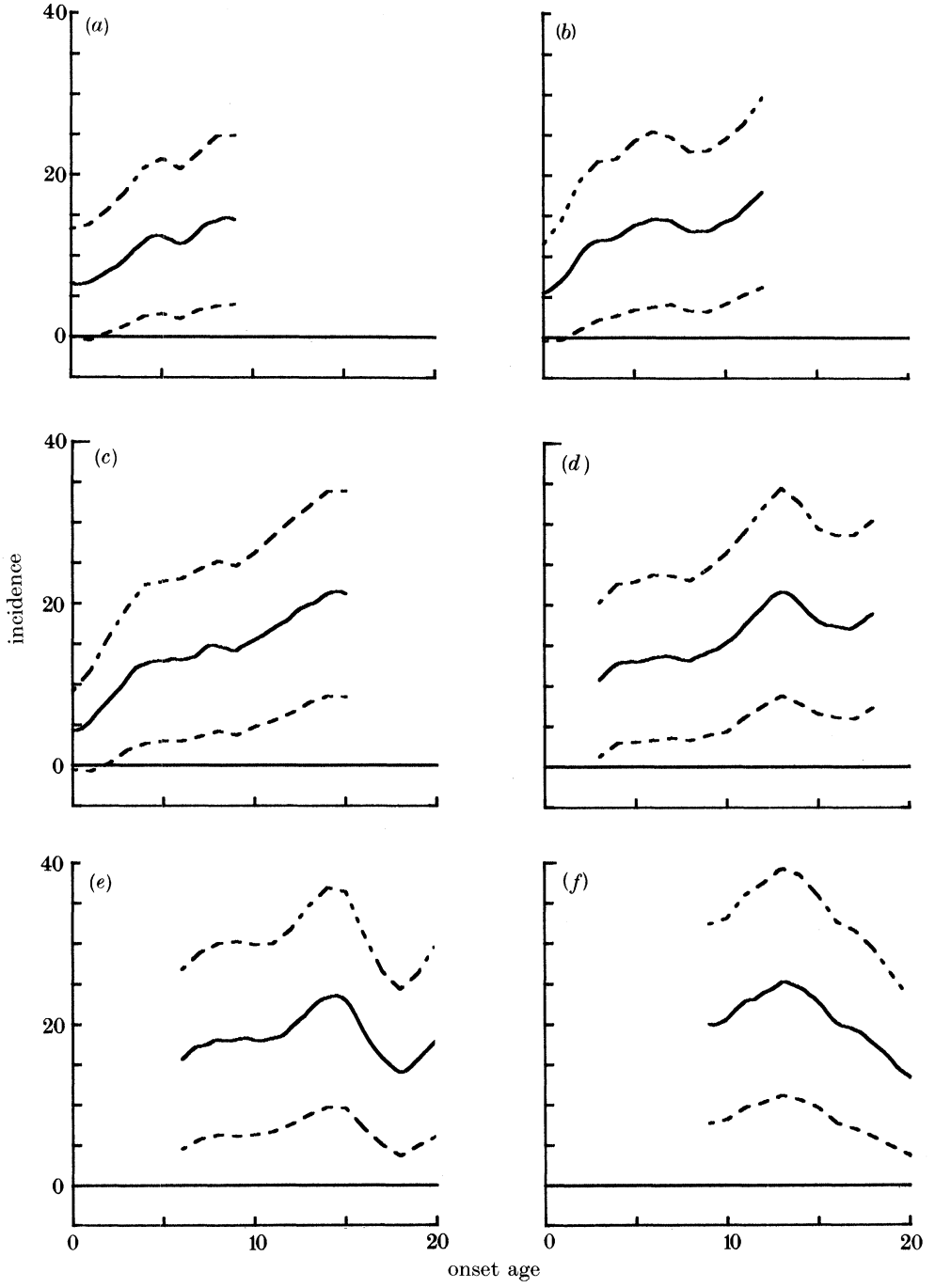


Figure 7. Estimated yearly diabetes incidence rates per 10^5 person years for Danish males plotted as a function of onset age for fixed onset year, with pointwise approximate 95% confidence limits. (*a*) 1959, (*b*) 1962, (*c*) 1965, (*d*) 1968, (*e*) 1971, (*f*) 1974.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

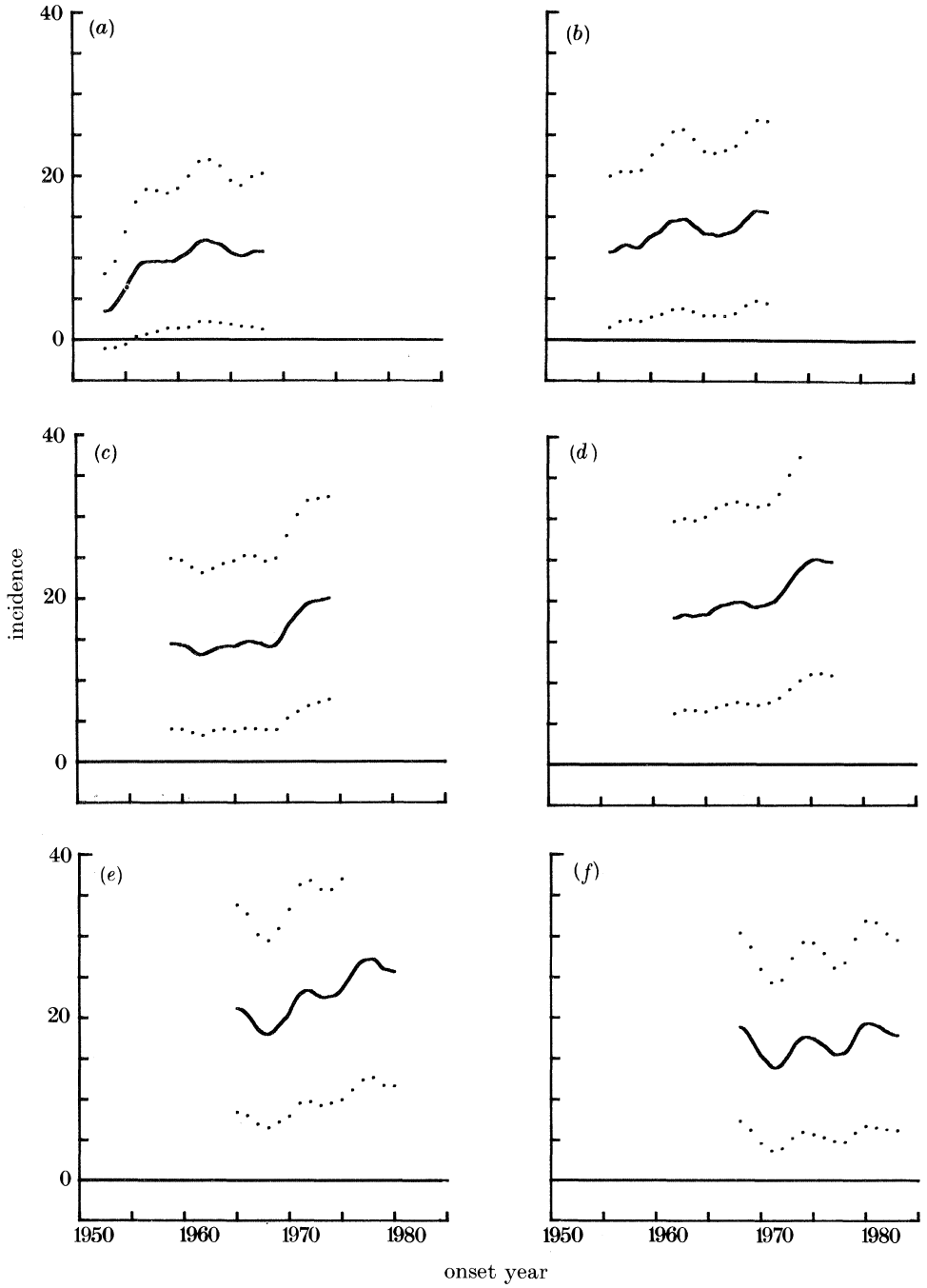


Figure 8. Estimated yearly diabetes incidence rates per 10^5 person years for Danish males plotted as a function of onset year for fixed onset age, with pointwise approximate 95% confidence limits. Age = 3 (*a*), 6 (*b*), 9 (*c*), 12 (*d*) 15 (*e*), 18 (*f*).

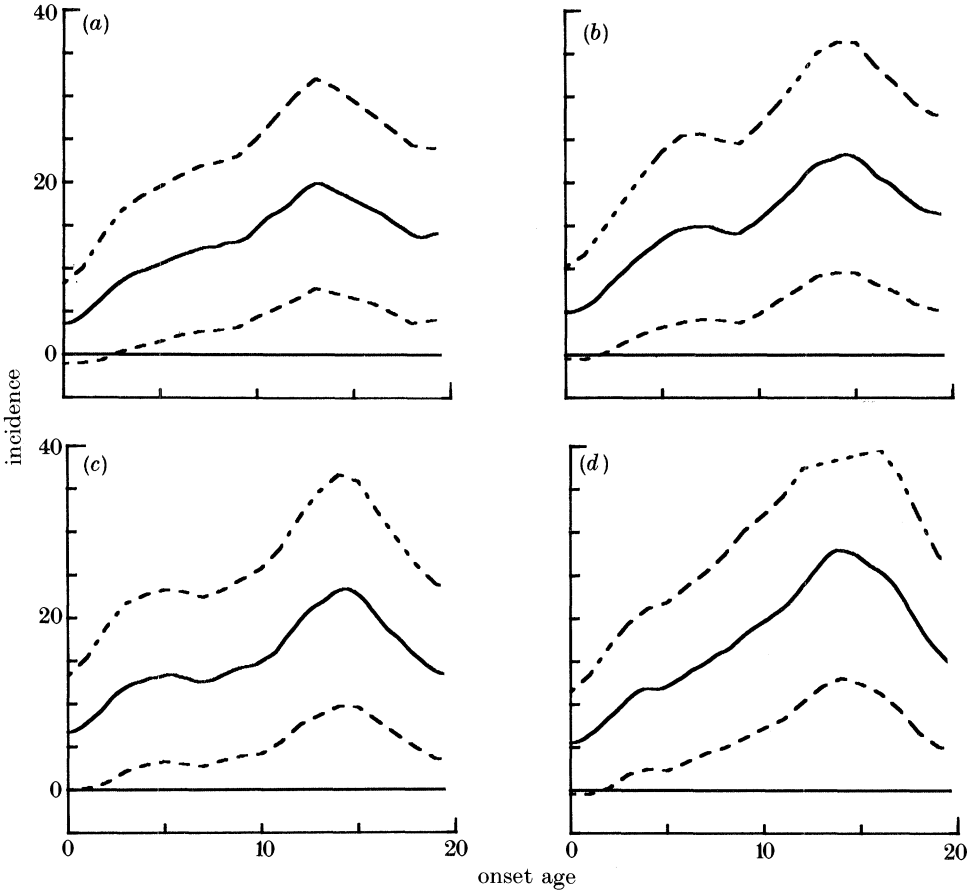


Figure 9. Estimated yearly diabetes incidence rates per 10^5 person years for Danish males plotted as function of onset age (and hence onset year) for fixed time of birth ('cohort'), with pointwise approximate 95% confidence limits. (*a*) 1953 cohort, (*b*) 1956 cohort, (*c*) 1959 cohort, (*d*) 1962 cohort.

segments with slope -1 (Lexis's isochronical lines, age effects for fixed calendar time), parallel to the cohort axis (time effects for fixed age) and parallel to the age axis (age effects for fixed cohort). In the modern Lexis diagram the three figures correspond to line segments that are vertical, horizontal and with slope 1 respectively. 95% approximate pointwise confidence limits are derived from $\hat{\sigma}^2_K(t,a)$ as postulated in §5*b*.

Green *et al.* (1990) obtained a statistically acceptable fit of a classical age–period multiplicative piecewise constant intensity model, with clearly significant different age–effects having a maximum in the 12–15 year age group and a clearly significant increase over time that seemed to stabilize during the last approximately twelve years of observation 1973–84.

Non-parametric continuous-time multiplicative intensity analyses along the lines of §§5*c* and 5*e* are in preparation. Note that data are available in a parallelogram that would make the use of McKeague & Utikal's (1988) suggestion unfeasible for an age–period model.

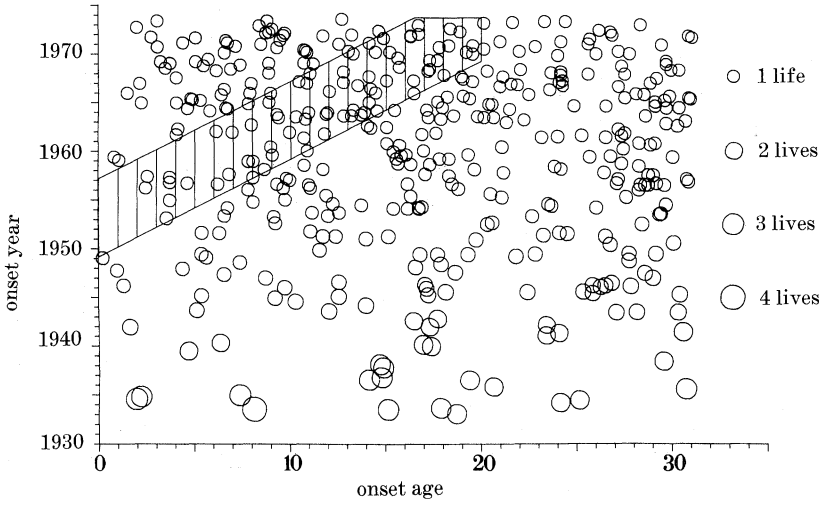


Figure 10. Time-age-specific onset of disease for male diabetic patients alive in Fyn County on July 1, 1973. The area of the circular disks is inversely proportional to the survival probability until that date. The hatched area covers the direct registration of incident cases based on conscription records (Keiding *et al.* 1989, fig. 3).

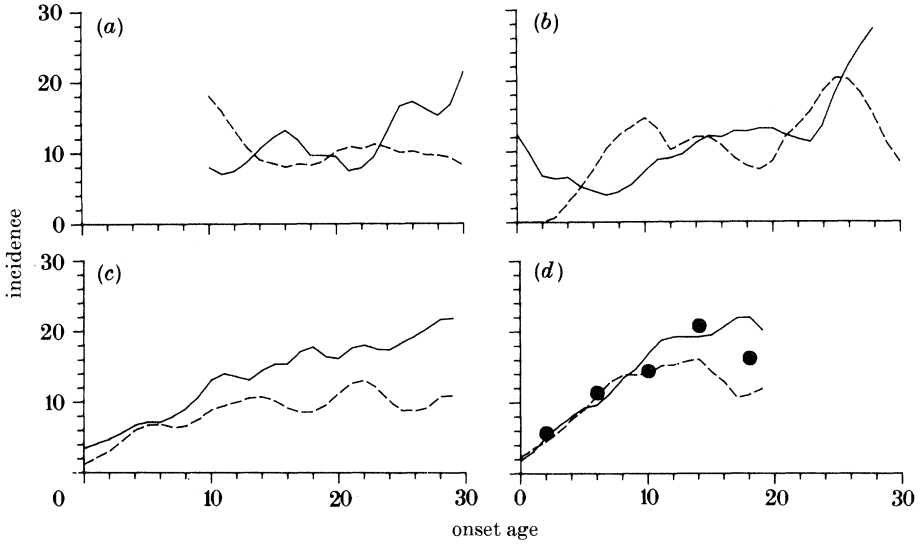


Figure 11. Estimated yearly diabetes incidence rates per 10^5 person years in Fyn County plotted as function of onset age (and hence onset year) for a fixed cohort. ——, males; – – –, females. Direct estimates for males are based on follow-up for the whole of Denmark (Green & Andersen 1983) marked by filled rings (Keiding *et al.* 1989, fig. 11). (*a*) 1923 cohort, (*b*) 1933 cohort, (*c*) 1943 cohort, (*d*) 1953 cohort.

(*b*) *Retrospective data based on prescriptions in the Danish National Health Service*

Based on prescriptions, Green *et al.* (1981) traced 1499 insulin-dependent diabetic patients in Fyn County (population approximately 450000) in Denmark, as of July 1, 1973. The case registration rate was close to 100%. Keiding *et al.* (1989) studied the 710 (410 males and 300 females) who had their disease diagnosed before or at 30

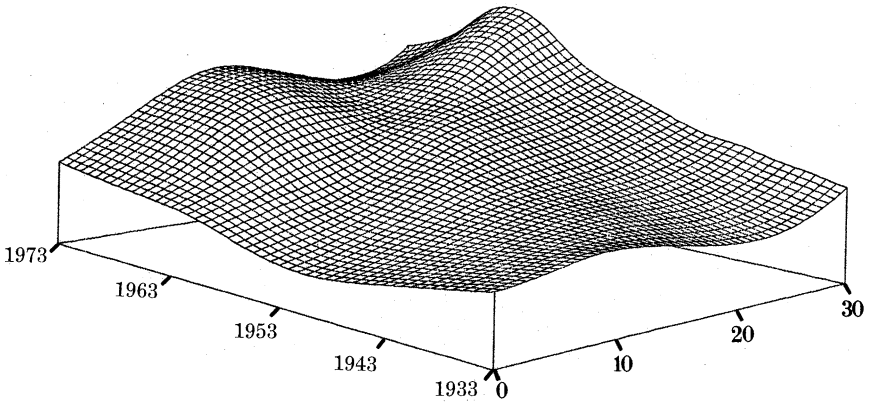


Figure 12. Estimated diabetes incidence rates in Fyn County for males 0–30 yrs old, during 1933–73, calculated by Professor Y. Ogata using penalized likelihood.

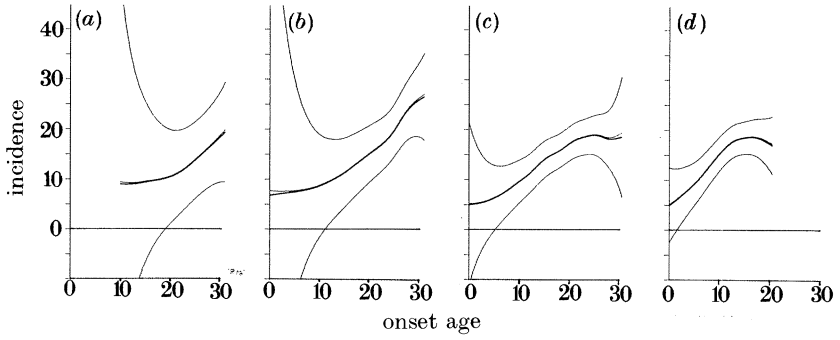


Figure 13. Estimated diabetes incidence rates per 10^5 person years in Fyn County plotted as function of onset age (and hence onset year) for fixed cohort. Males; ——, median; ——, mean. One standard error limits are indicated. Calculated by Professor Y. Ogata using penalized likelihood, see figure 11 above. (*a*) 1923 cohort, (*b*) 1933 cohort, (*c*) 1943 cohort, (*d*) 1953 cohort.

years and after 1 January 1933. For each of these patients the time of disease onset (as certified by a doctor) was ascertained from patient records, so that a retrospective (time, age) pattern of diabetes incidence is available over the (Jan. 1933, June 1973) × (0, 30 years)-rectangle in the Lexis diagram (see figure 10). Of course, estimation of diabetes incidence along the lines of the previous section requires correction for the retrospective nature of the data, indeed each case should be weighted by $1/p_s(t,a)$ where $p_s(t,a)$ is the probability for a diabetic patient of sex s with onset at time t and age a of surviving until 1 July 1973 (see the areas of the disks in figure 10). Estimates of $p_s(t,a)$ were derived from the data of Ramalau-Hansen *et al.* (1987), and the estimation itself was then done as explained in the previous section. Figure 11 (analogous to figure 9 above) shows the age trends for four selected cohorts. Y. Ogata (personal communication) reanalysed the data using the Ogata & Katsura (1988) empirical Bayes penalized likelihood approach briefly mentioned in §5*f*. Ogata did two independent smoothings: one (by penalized gaussian log likelihood) of the weights $\hat{p}(t,a)^{-1}$, and another (by spatial non-homogeneous Poisson) of the incidence data. The final surface was obtained by multiplication and is shown in figure 12 for males. The trends for the same four selected cohorts as in figure 1 are shown in figure 13, this time incorporating 'standard error limits', that is, estimate ± 1 estimated standard error. As was to be expected, the latter are rather

wide, particularly in the early part of the period. And they do not include the estimation uncertainty of the $\hat{p}(t, a)$.

The main qualitative findings of this exercise were a clear increasing secular trend in diabetes incidence over the studied period, particularly for males, and an increase in age of diabetes incidence with a tendency to a local maximum around puberty. There was some time (time, age)-overlap between these retrospective data and the prospective data (from the whole of Denmark) reported in the previous section, and for these (time, age)-combinations the agreement was surprisingly good (figure 11).

I am grateful to Martin Jacobsen for several important conversations on point processes on the Lexis diagram, to Anders Green and Per Kragh Andersen for permission to quote from their upcoming paper (Green *et al.* 1990), to Y. Ogata (Tokyo) for doing the reanalysis of the retrospective Fyn data and for permission to quote from it, to Vincenzo Capasso, Bendix Carstensen and Ian McKeague for supplying preprints of their results, and to Rob Tibshirani for information and instructions regarding the GAIM programs.

References

Andersen, P. K. & Borgan, Ø. 1985 Counting process models for life history data: a review (with discussion). *Scand. J. Statist.* **12**, 97–158.

Andersen, P. K. & Green, A. 1985 Evaluation of estimation bias in an illness-death-emigration model. *Scand. J. Statist.* **12**, 63–68.

Arthur, W. B. & Vaupel, J. W. 1984 Some general relationships in population dynamics. *Population Index* **50**, 214–226.

Beran, R. 1981 Nonparametric regression with randomly censored survival data. *Tech. Rep.* University of California, Berkeley.

Brillinger, D. R. 1986 The natural variability of vital rates and associated statistics (with discussion). *Biometrics* **42**, 693–734.

Capasso, V. 1988 A counting process approach for stochastic age-dependent population dynamics. In *Biomathematics and related computational problems* (ed. L. M. Ricciardi), pp. 255–269. Dordrecht: Kluwer Academic Publishers.

Cartensen, B. 1990 Analysis of rates classified by age, period and cohort: the problem of 3 directions in 2 dimensions. *Tech. Rep. Danish Cancer Registry.*

Clayton, D. & Schifflers, E. 1987 Models for temporal variation in cancer rates. II: age–period–cohort models. *Stat. Med.* **6**, 469–481.

Cox, D. R. 1972 Regression models and life tables (with discussion). *Jl R. stat. Soc.* B **34**, 187–220.

Dabrowska, D. M. 1988 Kaplan–Meier estimate on the plane. *Ann. Statist.* **16**, 1475–1489.

Green, A., Hauge, M., Holm, N. V. & Rasch, L. L. 1980 Epidemiological studies of diabetes mellitus in Denmark I. A case finding method based on the national service conscript registry. *Diabetologia* **19**, 355–358.

Green, A. & Andersen, P. K. 1983 Epidemiological studies of diabetes mellitus in Denmark: 3. Clinical characteristics and incidence of diabetes among males aged 0 to 19 years. *Diabetologia* **25**, 226–230.

Green, A., Andersen, P. K., Svendsen, A. J. & Mortensen, K. 1990 Further characterization of the increasing incidence of Type 1 (insulin-dependent) diabetes in young Danish males. *Techn. Rep.* University of Odense.

Hastie, T. & Tibshirani, R. 1986 Generalized additive models. *Stat. Sci.* **1**, 297–318.

Hoem, J. M. 1969 Fertility rates and reproduction rates in a probabilistic setting. *Biometrie-Praximétrie* **10**, 38–66. (Correction note *ibid.* **11**, 20 (1970).)

Jacobsen, M. 1982 Statistical analysis of counting processes. *Lect. Notes Stat.*, vol. **12**. Berlin: Springer-Verlag.

Jagers, P. 1975 *Branching processes with biological applications.* Wiley, New York.

Kaplan, E. L. & Meier, P. 1958 Non-parametric estimation from incomplete observations. *J. Am. statist. Assoc.* **53**, 457–481.

Keiding, N. 1990 Estimation of age specific incidence from cross-sectional data. *Res. Rep. 90/1,* Statistical Research Unit, University of Copenhagen.

Keiding, N. & Hoem, J. M. 1976 Stochastic stable population theory with continuous time. I. *Scand. Actuarial J.*, 150–175.

Keiding, N., Holst, C. & Green, A. 1989 Retrospective estimation of diabetes incidence from information in a prevalent population and historical mortality. *Am. J. Epidemiol.* **130**, 588–600.

Kendall, D. G. 1949 Stochastic processes and population growth. *Jl R. stat. Soc.* B **11**, 230–264.

Knapp, G. F. 1868 *Über die Ermittlung der Sterblichkeit aus den Aufzeichnungen der Bevölkerungs-Statistik.* Leipzig: Hinrichs.

Lexis, W. 1875 *Einleitung in die Theorie der Bevölkerungsstatistik.* Strassburg: Trübner. (Pages 5–7 translated to English by N. Keyfitz and printed, with figure 1, in *Mathematical Demography* (ed. D. Smith & N. Keyfitz). Berlin: Springer (1977).)

McCullagh, P. & Nelder, J. A. 1989 *Generalized linear models*, 2nd edn. London: Chapman & Hall.

McKeague, I. W. & Utikal, K. J. 1988 Identifying nonlinear covariate effects in semi-martingale regression models. *Tech. Rep.* The Florida State University.

McKeague, I. W. & Utikal, K. J. 1990 Inference for a nonlinear counting process regression model. *Ann. Statist.* (In the press.)

McKendrick, A. G. 1926 Applications of mathematics to medical problems. *Proc. Edinb. math. Soc.* **44**, 98–130.

Ogata, Y. & Katsura, K. 1988 Likelihood analysis of spatial inhomogeneity for marked point patterns. *Ann. Inst. Stat. Math.* **40**, 29–39.

Osmond, C. & Gardner, M. J. 1982 Age, period and cohort models applied to cancer mortality rates. *Stat. Med.* **1**, 245–259.

O'Sullivan, F. 1988 Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. stat. Comput.* **9**, 531–542.

O'Sullivan, F. 1989 Nonparametric estimation in the Cox proportional hazards model. *Tech. Rep.* University of California, Berkeley.

Pons, O. 1986 A test of independence between two censored survival times. *Scand. J. Statist.* **13**, 173–185.

Ramlau-Hansen, H. 1983 Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453–466.

Ramlau-Hansen, H., Jespersen, N. C. B., Andersen, P. K., Borch-Johnsen, K. & Deckert, T. 1987 Life insurance for insulin-dependent diabetics. *Scand. Actuarial J.* 19–36.

Sellke, T. & Siegmund, D. 1983 Sequential analysis of the proportional hazards model. *Biometrika* **70**, 315–326.

Thomas, D. C. 1983 Nonparametric estimation and tests of fit for dose-response relations. *Biometrics* **39**, 263–268.

Tibshirani, R. 1984 Local likelihood estimation. *Tech. Rep.* Department of Statistics, Stanford University.

Von Foerster, H. 1959 Some remarks on changing populations. In *The kinetics of cellular proliferation* (ed. F. Stohlman, Jr.). New York: Greene and Stratton.

Zeuner, G. 1869 *Abhandlungen aus der mathematischen Statistik.* Leipzig: Felix.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES