*Statistical Science*
2001, Vol. 16, No. 1, 1–22

# Understanding the Shape of the Hazard Rate: A Process Point of View

**Odd O. Aalen and Håkon K. Gjessing**

*Abstract.* Survival analysis as used in the medical context is focused on the concepts of survival function and hazard rate, the latter of these being the basis both for the Cox regression model and of the counting process approach. In spite of apparent simplicity, hazard rate is really an elusive concept, especially when one tries to interpret its shape considered as a function of time. It is then helpful to consider the hazard rate from a different point of view than what is common, and we will here consider survival times modeled as first passage times in stochastic processes. The concept of quasistationary distribution, which is a well-defined entity for various Markov processes, will turn out to be useful.

We study these matters for a number of Markov processes, including the following: finite Markov chains; birth–death processes; Wiener processes with and without randomization of parameters; and general diffusion processes. An example of regression of survival data with a mixed inverse Gaussian distribution is presented.

The idea of viewing survival times as first passage times has been much studied by Whitmore and others in the context of Wiener processes and inverse Gaussian distributions. These ideas have been in the background compared to more popular appoaches to survival data, at least within the field of biostatistics, but deserve more attention.

*Key words and phrases:* First passage time, hazard rate, survival analysis, quasistationary distribution, Wiener process, Markov chain.

## 1. INTRODUCTION

In survival analysis one studies the time to occurrence of some event. This event may be death, or it may be the diagnosis of some disease such as cancer. On a more positive note, the event could be the birth of a child or graduation from school. Whatever the setting, one wishes to analyze the probability distribution of the time to the event by means of survival curves and hazard rates. What is, however, usually disregarded in the standard approach to survival analysis is that the event is the end point of some process. Apart from pure accidents the events do not happen out of the blue, but there is a development preceding each event.

The reason for ignoring this feature is, of course, that the underlying process leading to the event is largely unknown. This, however, does not imply that it should be ignored. Considering the underlying process, even in a speculative way, may improve our understanding of the hazard rate and give alternative regression models. Here we will study first passage time models for stochastic processes moving in a transient space until ending up in an absorbing state. This is of relevance for studying both hazard rates and regression models.

Hazard rates play a fundamental role in survival analysis. Although useful, they may not be easy to understand. Why, for instance, do hazard rates sometimes increase, sometimes decrease and sometimes first increase and then decrease; see Figure 1 for a picture of typical hazard shapes occurring in practice. Rates of divorce, for instance, as measured from time of marriage, first increase and then decrease somewhat; see Figure 2. This has resulted

*Odd O. Aalen is Professor, Section of Medical Statistics, University of Oslo, P.O. Box 1122 Blindern, N-0317 Oslo, Norway (e-mail: o.o.aalen@basalmed.uio.no). Håkon K. Gjessing is Associate Professor, Section of Medical Statistics, University of Oslo, P.O. Box 1122 Blindern, N-0317 Oslo, Norway.*

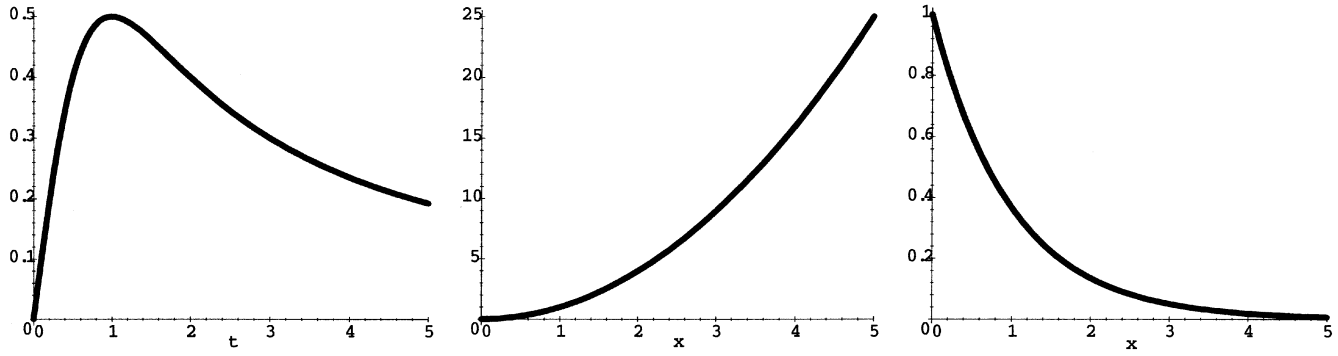


FIG. 1. *Typical shapes of hazard rates.*

in speculation about a crisis in the typical marriage after a few years leading either to divorce or to consolidation of the marriage. Within the context of frailty theory a different explanation has been given. It has been pointed out that the hazard rate is not merely a measure of the development of risk in a single individual, but is also influenced by selection effects among individuals. The most frail individuals will tend to fail first, leaving the more robust ones. This will imply that a population hazard may decrease even though the individual rates are increasing.

The process point of view assumed here gives yet another explanation of these phenomena. The shape of the hazard rate is created in a balance between two forces: the attraction of the absorbing state and the general diffusion within the transient space. It turns out that the various common shapes of hazard rates, as illustrated in Figure 1, occur naturally depending on how the starting distribution on the transient state space corresponds to what is termed a *quasistationary distribution*. Simplifying quite a bit, one could say that the shape of the hazard rate depends on the distance between the starting point, or starting distribution, and the state of absorption. A great distance leads to an increasing hazard rate; an intermediate distance leads to a hazard rate that is first increasing and then declining; a small distance leads to an (essentially) decreasing hazard rate. Furthermore, the hazard rate will typically (when the underlying process is reversible, i.e., can move back and forth) converge to a constant value. (Note that we do not use reversibility in the more precise sense of time reversibility as defined in Keilson, 1979). Hence a constant hazard should be expected as a limit in certain cases. As an example, the hazard rates in Figure 2 seem to exhibit an approximately constant level after the initial rise (it should be noted though that after 25 years of marriage divorce rates tend to start declining).

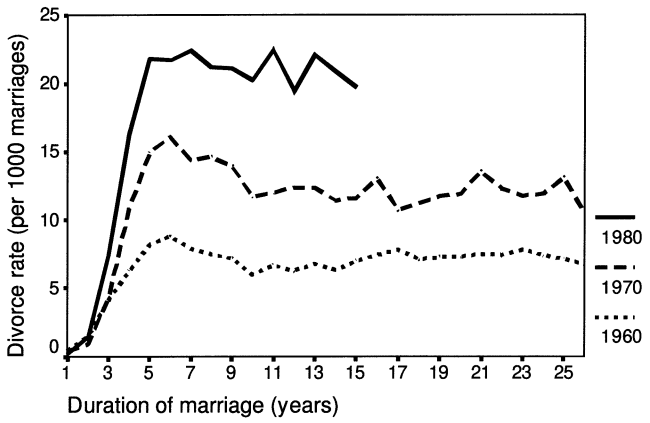


FIG. 2. *Rates of divorce among Norwegian couples married in* 1960 (*lower curve*), 1970 (*middle curve*) *and* 1980 (*upper curve*). (*Based on data from Statistics Norway.*)

The issue of an underlying stochastic process is also relevant to the interpretation of covariates. One may ask what covariates really influence. In Cox's and several other regression models it is assumed that covariates influence the hazard of an event. Generally, however, it would be more correct to think of covariates as influencing the underlying process. If I drive a car and fall asleep, then this immediately influences the risk (hazard) of an accident, so the time-dependent covariate "being asleep" certainly has a direct effect on the hazard. If, on the other hand, the blood pressure rises somewhat over time and becomes permanently increased, then this is something that does not directly and in an immediate sense influence the hazard of, say, myocardial infarction. Rather, the increased blood pressure promotes an underlying disease process and may only in the long term influence the hazard of disease.

In fact, often the covariate is not something that effectively influences even the underlying process, but is rather a measure of how far this process has advanced. This is the case of many covariates that

are used in the analysis of clinical studies. Prominent examples might be bilirubin in liver disease, CD4 counts in HIV infection and various staging measures in cancer. When measured repeatedly over time, covariates of this kind are often called markers; see, for example, Nielsen and Linton (1995) and Jewell and Kalbfleisch (1996).

Hence, it would be more fair to the real biological meaning of covariates if one considered them in relation to an underlying stochastic process.

The same holds for the, now popular, concept of frailty (a review of frailty theory may be found, e.g., in Aalen, 1994). As opposed to covariates that capture the measured differences between individuals, frailty denotes the unmeasured individual heterogeneity. In the standard frailty model it is assumed that the frailty variable acts multiplicatively on the individual hazard rate. This means that frailty is purported to influence directly the hazard of the event. Again it would be better to think of frailty as something which influences an underlying process. This is especially important when frailty is perceived as changing over time. As mentioned above for covariates, the biological meaning of frailty may vary from case to case. Sometimes, frailty may be best perceived as a measure of how advanced the underlying process is.

Of course, the point of view promoted here is not new. Much previous work has been done along these lines. The accelerated failure time model, where covariates are supposed to influence the acceleration factor, is a prominent example. Here one models one particular aspect of the underlying process leading to failure, namely its speed. For an interesting discussion of how this model adapts much better to a frailty adjustment than the proportional hazards model, see Keiding, Andersen and Klein (1997). Another important class of models is the first passage time models based on Wiener processes (see, e.g., Whitmore, 1986a). The inverse Gaussian distribution comes up naturally in this context, and no doubt this is an underused class of distributions in survival analysis; in fact, it is our distinct impression that more attention should be payed, also in biostatistics, to the work of Whitmore and others.

A further example of the use of a stochastic process point of view are the phase type models, that is, first passage time models in finite Markov chains. An important practical example of this is the modeling of progression of HIV disease by Longini and co-authors (see, e.g., Longini et al., 1989). An extended Markov chain model is studied by Aalen et al. (1997).

In reliability theory considerable effort has gone into proving results concerning the shape of the hazard rate, e.g., deciding conditions under which it is increasing, or increasing on the average. This has also been related to shock models, wear processes and first passage time distributions. Some useful references are Patel (1983), Griffiths (1988) and Shaked and Shanthikumar (1991). In reliability there has also been considerable interest in "bathtub"-shaped failure rates (i.e., first decreasing and then increasing); see, e.g., Lynn and Singpurwalla (1997). We will not consider that kind of shape here, but they may certainly be modeled within our framework; see Aalen (1995).

This paper focuses on the shape of hazard functions derived from first passage time distributions. We do not give general theorems concerning the shape of the hazard, but look at examples, special cases and illustrations. The quasistationary distribution, which is an important concept in probability theory, turns out to be useful in understanding the shape of the hazard. Examples of quasistationary distributions are exhibited for several models, both discrete and continuous. The hazard ratios for some models also are studied. A randomized version of the Wiener process is studied in connection with a practical example of analyzing survival data with covariates. We also consider briefly general diffusion processes on the positive real line with absorption in 0. Such a diffusion will have a distribution on the positive half-line which, when normalized to mass 1, converges to the quasistationary distribution. The hazard rate of the time to absorption is proportional to the derivative of the density of the normalized distribution at 0, a relationship that may give useful intuitive insight.

## 2. FIRST PASSAGE TIME MODELS

The basic approach in this paper is to model failure times as times of first passage for suitable stochastic processes. A discrete-space Markov chain will be considered, as well as continuous-space diffusion processes with the Wiener process being the most prominent example.

### 2.1 Phase Type Models

Consider a time-continuous Markov chain on a finite state space with constant intensities of transition. The state space of the Markov chain is assumed to consist of a set of transient states as well as one absorbing state. The process starts out according to a probability distribution on the transient state space, and the failure time is the time until absorption. The distribution of time to absorption is called a phase type distribution, and these have received quite a lot of attention in the literature (see, e.g., O'Cinneide, 1990, or the review in

Aalen, 1995). We will consider such processes below as well as an example with infinite state space.

### 2.2 Wiener Processes and Diffusion Processes

Consider a Wiener process which moves freely until it is absorbed when it reaches a certain point. If the process starts out at a specific point, then time to absorption follows an inverse Gaussian distribution. There are a number of papers on the application of Wiener processes in survival analysis; see, e.g., Lancaster (1982), Whitmore (1986a, b, 1995), Whitmore, Crowder and Lawless (1998), Eaton and Whitmore (1977) and Doksum and Høyland (1992). We study Wiener processes as well as more general diffusion processes below.

### 2.3 Different Types of Models

When studying stochastic processes for biological or social phenomena, a major distinction exists between reversible and irreversible processes. Some diseases, such as nontreatable cancer, might be irreversible, while others are clearly reversible in the sense that the patient is cured. Many diseases, such as rheumatism or migraine, might be chronic, but still go back and forth between good and bad periods, that is, are at least partially reversible. As an example of a social phenomenon, consider the incidence of divorce: it is clear that most marriages will have good and bad periods, undergoing reversible processes of deterioration and improvement, before some may end in divorce.

By reversibility we simply mean that the transient states of the process make up a single class. This irreducibility of the transient state space is important when considering quasistationary distributions as is done below. Typically such distributions are well defined for Markov processes with absorption when the transient states constitute a single class.

Another distinction, which is related to that above, is useful when studying hazard rates: by a progressive model we mean that with probability 1 the process starts out in an "extreme" state which represents the natural starting point in the state space and then moves toward the end state. Even though there may be temporary backward movements and some reversibility, there is still a clear direction in the development. An example, when studying the development of a disease, is that the individual starts out as completely disease free, but that the disease has a progression once it is established. A nonprogressive model, on the other hand, has no clear direction in the development. The process is not supposed to start in an extreme point, and it does not necessarily have any force of movement toward the absorbing state. Again, marriage can be used as an example: the marriage might improve or it might deteriorate after the wedding, but there is no law that says that it has to move in one direction.

The distinction between progressive and nonprogressive models is somewhat tentative. Discussion and examples for phase type models are given by Aalen (1995).

Typically, progressive models will tend to have increasing hazard rates. A complication may be frailty (i.e., unobserved differences in transition rates between individuals) which may result in the hazard rate being "bent down" and eventually decreasing.

For nonprogressive models, the most typical shape, also in the absence of frailty, will be a hazard rate that first increases and then decreases. This will be further demonstrated below.

## 3. QUASISTATIONARY DISTRIBUTIONS

To understand the shapes of the hazard rates that may result from first passage time distributions, an essential element is *the quasistationary distribution*. It turns out that often there exists a distribution on the transient state space such that if the process starts according to this distribution it will have a constant hazard of transition. This means that although probability mass is continuously being drained from the transient space, nevertheless the remaining probability distribution on this space converges to a limiting distribution.

The *shape of the hazard rate* is in many cases determined by how the starting distribution of the process relates to the quasistationary one, whether it is closer to or further apart from the absorbing state. Several examples of quasistationary distributions are given below.

### 3.1 Phase Type Models

A sufficient condition for a quasistationary distribution to exist is that the transient states make up a single class (i.e., the process is completely reversible prior to absorption); see, e.g., Aalen (1995) for examples and references. Let $A$ denote the intensity matrix restricted to the transient states. Then the quasistationary distribution is defined by the unique strictly positive left eigenvector of $A$ (normalized to sum 1). If the process is started on the transient state space according to this quasistationary distribution, the hazard rate of time to absorption will be constant and equal to minus the eigenvalue of the above eigenvector.

If the process starts out according to some other distribution, then the distribution on the state space will converge to the quasistationary one.

### 3.2 Finite Birth–Death Process (Reflected Random Walk) with Absorbing State

Consider the birth–death process on states 0 to 5, with absorption in state 0, as shown in Figure 3. The transition intensities are $\alpha$ for moving away from the absorbing state and $\beta$ for moving toward it. The quasistationary distribution is a normalized version of a left eigenvector of the transition intensity matrix restricted to the transient state space, that is, of the following matrix:

$$\begin{pmatrix} -\alpha-\beta & \alpha & 0 & 0 & 0 \\ \beta & -\alpha-\beta & \alpha & 0 & 0 \\ 0 & \beta & -\alpha-\beta & \alpha & 0 \\ 0 & 0 & \beta & -\alpha-\beta & \alpha \\ 0 & 0 & 0 & \beta & -\beta \end{pmatrix}.$$

More precisely, the quasistationary distribution is the left eigenvector corresponding to the dominant eigenvalue (i.e., the one closest to zero). Whenever the transient state space constitutes a single finite class (which is the case here) the quasistationary distribution is unique and strictly positive on all states. The absolute value of the dominant eigenvalue is the constant hazard rate of absorption corresponding to a quasistationary starting distribution.

For a numerical example, put $\alpha = 1.5$ and $\beta = 1.0$. The quasistationary distribution on states 1 to 5 is given as 0.037, 0.090, 0.167, 0.276, 0.430. The absolute value of the dominant eigenvalue is 0.037, which equals the constant hazard rate of absorption under quasistationarity.

Assume now that the process starts out in a single given state with probability 1. The hazard rate of the phase type distribution in this case depends on how the starting state is related to the quasistationary distribution. Three cases are considered here, namely those corresponding to starting out in state 1, state 3 or state 5. The first state is very close to the absorbing state, while the second state is close to the average of the quasistationary distribution. The third state (state 5) is clearly further removed from the absorbing state than the major part of the quasistationary distribution. In Figure 4 one sees how these cases correspond to hazard rates with distinctly different shapes. In particular, the hazard rate that first increases and then decreases arises naturally in the intermediate situation. The phenomenon observed here probably has a general validity, as different models below indicate.

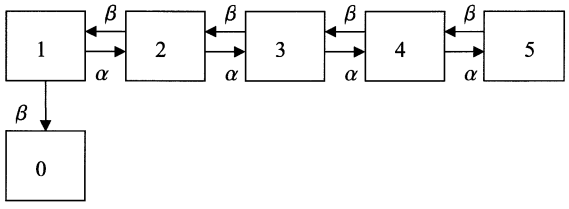


FIG. 3. *State space of a phase type model.*

When starting out in state 5, with probability 1, the model is of a progressive type and one sees that this yields an increasing hazard rate. Starting out in other states gives nonprogressive models and hazard rates that decrease from a certain time. A more detailed discussion is given by Aalen (1995).

### 3.3 Infinite Birth–Death Process (Infinite Random Walk)

Consider an infinite birth–death process with absorbing state in 0, that is, an extension of the above model where instead of five transient states there is an infinite number. Just as in the above example $\alpha$ and $\beta$ denote the rates of moving up and down the state space. The advantage of considering this extension is that simple explicit formulas may be given for interesting quantities. In particular, explicit formulas for quasistationary distributions are given by Cavender (1978) for the case when such distributions exist, namely when $\alpha < \beta$. There is in fact a whole set of them, but one distribution is a canonical (or minimal) choice in the sense that it represents the limiting case for a process starting out with probability 1 in a specific state. The

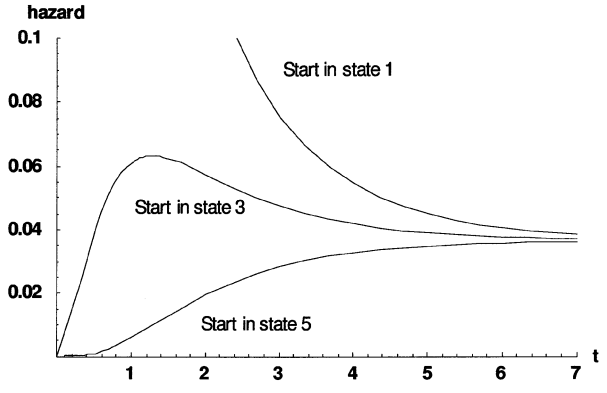


FIG. 4. *Hazard rates for time to absorption dependent on starting state in phase type model.* ($\alpha = 1.5$, $\beta = 1.0$)

formula for this canonical distribution is

$$q_1 = \frac{\alpha}{\beta} + 1 - 2\sqrt{\frac{\alpha}{\beta}},$$

$$q_n = n q_1 \left(\frac{\alpha}{\beta}\right)^{(n-1)/2}, \quad n = 2, 3, \ldots .$$

Note that this is a negative binomial distribution. The rate of absorption with this distribution will of course equal the probability $q_1$ of being in state 1 multiplied by the rate $\beta$ of passing down to 0. Hence, the constant hazard rate of absorption that arises under quasistationarity equals

$$\beta \times q_1 = \alpha + \beta - 2\sqrt{\alpha\beta}.$$

This is the limit of the hazard rate of the first passage time to 0 (starting out in a single state) when time goes to infinity. In the case we study here the probability density of this distribution may be found, for instance, from Gross and Harris, 1985, page 134, formula (2.109), and page 143. Starting out in state $n$ the density of time to absorption is given as

$$n\left(\frac{\beta}{\alpha}\right)^{n/2} t^{-1} \exp(-(\alpha+\beta)\,t)\, I_n(2\sqrt{\alpha\beta}\,t), \quad t > 0,$$

where $I_n(t)$ is the modified Bessel function of order $n$.

A general result of Keilson (1979) says that the hazard rate of the first passage time to a neighboring state in a general birth–death process is always decreasing. Hence, when starting out in the state which is closest to absorption, namely state 1, one will necessarily have a decreasing hazard rate. Otherwise, when starting out in a single state with probability 1, experience indicates that the hazard rate will increase first and then decrease, but for states far removed from the absorbing one, the hazard will be virtually only increasing.

The present theory may give an explanation of decreasing hazards arising in practice. It is well known that after myocardial infarction the death rate is initially very high and then falls sharply. Data on this have been analyzed from a frailty point of view by Hougaard (1986), that is, assuming that some individuals have a much higher risk of dying from the disease than others. The application of a conventional frailty model fits the data well. A quite different interpretation of the declining death rate can be given in the present framework. If the absorbing state 0 signifies death, then state 1 might correspond to a very critical illness, such as myocardial infarction, from which some individuals would die, while others would soon improve, that is, move upward in the state space. According to the present theory, individuals starting out in state 1 would experience a decreasing hazard rate of dying just as observed for the myocardial infarction patients. So it would be expected that patients in a very critical state would have a decreasing hazard rate.

### 3.4 Wiener Process with Absorption

Consider a Wiener process with drift $-\mu$ ($\mu > 0$) and variance coefficient $\sigma^2$. Assume the process starts out in some positive state and is absorbed whenever it hits zero. Quasistationary distributions for this case have been studied fairly recently (Martinez and San Martin, 1994), and it turns out that there is a whole family of such distributions (due to the infinite state space) in the case when $\mu > 0$ (i.e., the process has a drift toward zero). One of these distributions is "canonical" in the following sense: starting out in a single given state with probability 1, the distribution of "survivors" will converge to the canonical one. The other quasistationary distributions are more heavy-tailed and cannot be reached in this way.

The canonical one is given by $(\mu^2/\sigma^4)x \exp(-\mu\, x/\sigma^2)$, which is a gamma distribution. If the process is initiated with this distribution on the positive real line, then the hazard rate of absorption in zero is constant and given as $(\mu/\sigma)^2/2$. Putting $\mu = 1$ and $\sigma^2 = 1$ gives the distribution shown in Figure 5. It is of some interest here to note that the hazard rate *depends on the square of* $\mu$. Say one considers some rare event, with absorption corresponding to this event occurring. If the drift toward this event doubles, say, then the rate at which the event occurs is multiplied by 4.

When the process starts at a given point $c\,(>0)$ the distribution of time to absorption in 0 is an inverse Gaussian one (see, e.g., Chhikara and Folks, 1989), with density

$$f_{ig}(t, c, \mu, \sigma) = \frac{c}{\sigma\sqrt{2\pi}}\, t^{-3/2} \exp\left[-\frac{(c-\mu t)^2}{2\sigma^2 t}\right].$$

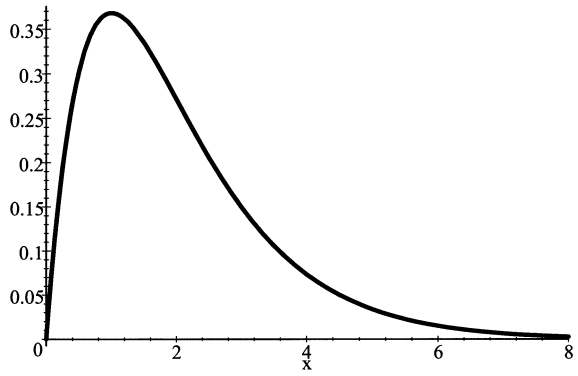


FIG. 5. *Quasistationary distribution for a Wiener process with absorption* (*parameters* $\mu/\sigma^2 = 1$).

The cumulative distribution function is

$$F_{ig}(t, c, \mu, \sigma) = 1 - \Phi\left(\frac{c - \mu t}{\sigma\sqrt{t}}\right) + \exp\left(\frac{2c\mu}{\sigma^2}\right)\Phi\left(\frac{-c - \mu t}{\sigma\sqrt{t}}\right),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution, and, accordingly, the hazard rate is

$$\lambda_{ig}(t, c, \mu, \sigma) = \frac{f_{ig}(t, c, \mu, \sigma)}{1 - F_{ig}(t, c, \mu, \sigma)}.$$

There are three parameters in the inverse Gaussian distribution, namely $c, \mu$ and $\sigma$, but the distribution only depends on these through the functions $c/\sigma$ and $\mu/\sigma$. Hence, from a statistical point of view, there are only two free parameters. This means, for instance, that we can put $\sigma = 1$ in a statistical analysis, without loss of generality. Note that this is only true when considering time to absorption and not when studying other aspects of the process as seen from (1).

The shape of the hazard rate of this distribution is similar to that observed for a phase type distribution above. If $c$ is close to zero compared to the quasistationary distribution one gets, essentially, a decreasing hazard rate; a value of $c$ far from zero gives essentially an increasing hazard rate, while an intermediate value of $c$ yields a hazard that first increases and then decreases. The wording "essentially" is used here because the continuous nature of the model and the noncompact state space yield hazard rates that will, strictly speaking, always increase to a maximum and then decrease (see, e.g., Seshadri, 1998, Proposition 5.1), but for $c$ small or large they can be seen as just decreasing or just increasing for most practical purposes. An illustration is shown in Figure 6 where values 0.2, 1 and 3 are chosen for c. It is important to note how these values relate to the quasistationary distribution: From Figure 5 one sees that these values are, respectively, placed at the beginning of the quasistationary distribution, close to the mode of the distribution, and in its tail.

Assume that the process starts out at the point $c$ at time 0. The probability density at time $t$ of the process to be at $x > 0$ may be found in Cox and Miller (1965, page 221):

$$g(x, t, c, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi t}}\left\{\exp\left[-\frac{(x - c + \mu t)^2}{2\sigma^2 t}\right] - \exp\left(\frac{2c\mu}{\sigma^2}\right)\exp\left[-\frac{(x + c + \mu t)^2}{2\sigma^2 t}\right]\right\}. \quad (1)$$

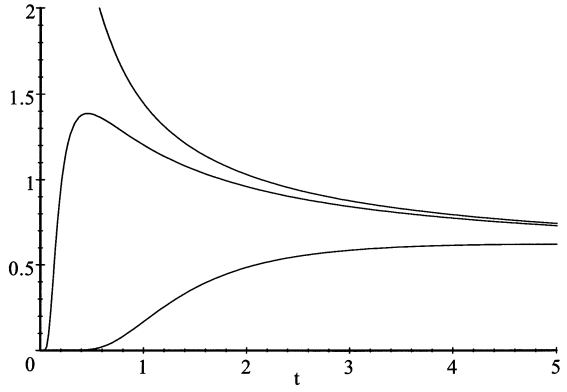


FIG. 6. *Hazard rates for time to absorption when process starts out in* $c = 0.2$ (*upper curve*), $c = 1$ (*middle curve*) *and* $c = 3$ (*lower curve*). *In all cases* $\mu = 1$ *and* $\sigma^2 = 1$.

This density is a linear combination of two normal distributions. The integral under the density will be less than 1 and decrease with time since there is an increasing probability of absorption at time 0.

The conditional distribution along the positive axis given that absorption has not taken place is given as

$$f(x, t, c, \mu, \sigma) = \frac{g(x, t, c, \mu, \sigma)}{1 - F_{ig}(t, c, \mu, \sigma)}.$$

The distribution illustrated in Figure 7 corresponds to $c = 1, \mu = 1$ and $\sigma = 1$. Here all individuals start out at $x = 1$, and the distributions are shown at times 0.05, 0.20, 1.00 and 10.00, together with the quasistationary distribution. One sees that the distribution of surviving individuals is rapidly spreading out from $x = 1$, and, in fact, approaching the quasistationary distribution.

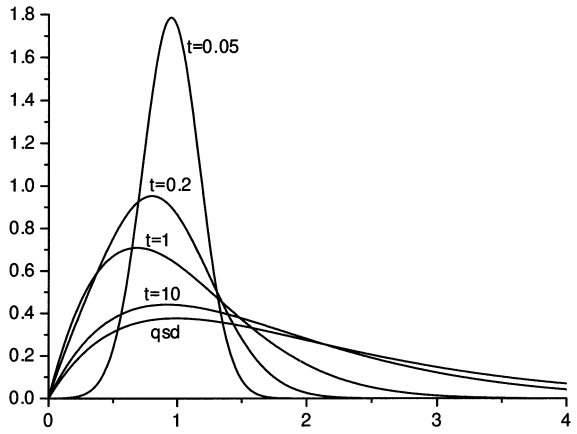


FIG. 7. *Normalized distribution of "survivors,"* $f(x, t, c, \mu, \sigma)$, *for* $c = 1$, $\mu = 1$, $\sigma = 1$ *at times* $t = 0.05, 0.2, 1, 10$. *The quasistationary distribution* (*qsd*) *is added for comparison.*

From results for diffusion processes given below [see (4)], it follows that the hazard rate of the time to absorption at time $t$ equals the derivative at $x = 0$ of $f(x, t, c, \mu, \sigma)$, multiplied by $\sigma^2/2$. In the example this corresponds to half the slope at zero of the curves in Figure 7. From the figure it is apparent that the slope, and hence the hazard, increases for small $t$ to reach a maximum and then declines to the value for the quasistationary distribution; see also the middle curve of Figure 6. This development of an increase followed by a decrease is quite reasonable: the mass will rapidly approach the absorbing state at zero resulting in an increasing slope, but later on the slope declines as mass keeps moving out into the extreme parts of the quasistationary distribution.

### 3.5 Wiener Process with Absorption Initiated According to Some Distribution

Instead of starting out in a single state, one may initiate the process according to some distribution on the positive real line. The results become particularly simple if the starting distribution is related to the quasistationary one. Assume for simplicity that $\sigma = 1$, that is, that the canonical quasistationary distribution equals $\mu^2 x \exp(-\mu x)$. Assume that the Wiener process with drift is initiated according to a distribution proportional to $x^k \exp(-\mu x)$ for some $k$. When $k < 1$ this will be a gamma distribution which is closer to the absorbing state 0 than the quasistationary one. If $k > 1$ the distribution is further removed from zero than the quasistationary one. By integrating the inverse Gaussian distribution with respect to the above distribution, one derives a new first passage distribution which is simply a gamma distribution proportional to

$$x^{(k-1)/2} \exp(-(\mu^2/2)x).$$

For $k < 1$ this is well known to have a decreasing hazard rate, while for $k > 1$ it is known to have an increasing hazard rate. Hence, this example demonstrates how the shape of the hazard of the first passage distribution is related to how the starting distribution of the process compares with the quasistationary one.

One may also start out with a distribution of the type $f(x, t, c, \mu, \sigma)$. One then gets the later part of a hazard rate (leaving out the beginning) and can hence achieve unimodal or decreasing hazard rates.

### 3.6 Wiener Process with Lower Reflecting Barrier and Upper Absorbing Barrier

When modeling a phenomenon by a Wiener process it is often natural to think that there should be a lower bound, so that there is not the infinite state space to be considered. Such an alternative is a Wiener process with a reflecting lower barrier and an absorbing upper barrier. There does not exist too much theory on this kind of process since it is difficult to handle. A very good paper is that by Schwarz (1992), who presents formulas for first passage time distributions. A simple formula for the quasistationary distribution may be derived from his work in the case of a process without drift, a reflecting barrier at 0 and an absorbing barrier at $L$. Then the distribution becomes

$$\frac{\pi}{2L} \cos\left(\frac{\pi x}{2L}\right)$$

and the corresponding hazard rate of absorption is

$$\frac{\pi^2 \sigma^2}{8L^2}.$$

It is interesting to note that the quasistationary distribution is independent of $\sigma$. If there is drift, then more complex formulas may be derived from the work of Schwarz.

### 3.7 Geometric Brownian Motion

Another natural way of restricting the process to positive values is to consider a geometric Brownian motion. If $X_t$ is a Wiener process with drift $-\mu$ and variance coefficient $\sigma^2$ starting in $c^* = \log(b) - \log(c) > 0$ and being absorbed in zero, then $Y_t = b \exp(-X_t)$ is a geometric Brownian motion (Øksendal, 1998) with positive drift, starting at $Y_0 = c$ and being absorbed when $Y_t$ reaches the level $b > c$. The process $Y_t$ is thus restricted to values in $(0, b)$. The hitting time distributions of $X_t$ and $Y_t$ are identical, and since $Y_t$ is a simple transformation of $X_t$ the distribution of $Y_t$ (conditional on not being absorbed) can be computed by a transformation of the inverse Gaussian distribution. The analysis of $Y_t$ is thus equivalent to that of $X_t$. Geometric Brownian motion is used as a model for unrestricted exponential growth. As a model for an underlying biological process $Y_t$ can spend a considerable amount of time close to zero, but when it reaches larger values it picks up speed and moves faster toward the barrier, since the drift is proportional to $Y_t$.

## 4. DIFFUSION PROCESS WITH ABSORPTION

We shall briefly view the situation from a more general point of view, by considering a Markovian diffusion process on the positive half-line with zero as the absorbing state. Let $\varphi_t(x)$ be the density on the state space, denoted by $x$, at time $t$ and let $\sigma^2(x)$ and $\mu(x)$ be the variance and drift diffusion coefficients, respectively. The evolution forward in time is

described by Kolmogorov's forward equation (Karlin and Taylor, 1981, page 220):

$$\frac{\partial}{\partial t}\varphi_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\big[\sigma^2(x)\varphi_t(x)\big] - \frac{\partial}{\partial x}[\mu(x)\varphi_t(x)]. \tag{2}$$

### 4.1 An Equation for the Quasistationary Distribution

Assume that the process is in a quasistationary state. Then one can write $\varphi_t(x) = e^{-\theta t}\psi(x)$, where $\theta$ is the constant hazard rate and $\psi(x)$ is the quasistationary distribution. Insertion into the above equation yields

$$-\theta\psi(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\big[\sigma^2(x)\psi(x)\big] - \frac{\partial}{\partial x}[\mu(x)\psi(x)].$$

The constant drift $\mu$ and variance coefficient $\sigma^2$ of the Wiener process give:

$$-\theta\psi(x) = \frac{1}{2}\sigma^2\psi''(x) - \mu\psi'(x).$$

The quasistationary distributions of the Wiener process with absorption can be found solving this equation with appropriate boundary conditions.

### 4.2 A Formula for the Hazard Rate

Consider the process prior to quasistationarity, and let $\theta_t$ denote the hazard rate of the time to absorption. Let $\psi_t(x)$ denote the normalized density (i.e., with integral 1) on the state space, that is,

$$\varphi_t(x) = \exp\left(-\int_0^t \theta_s\, ds\right)\psi_t(x). \tag{3}$$

Then the following result holds under suitable regularity assumptions:

$$\theta_t = \frac{\sigma^2(0)}{2}\psi_t'(0); \tag{4}$$

that is, the hazard rate is proportional to the slope of the normalized density at zero. The formula can easily be verified for the Wiener process with absorption studied above. Note that $\psi_t(x)$ can be considered the distribution of survivors in the context of survival analysis.

Formula (4) is related to known results on first passage time distributions [see, e.g., Goel and Richter-Dyn, 1974, Section 3.2, formula (24)]. We indicate briefly how the formula may be derived from the forward equation. Inserting (3) into (2) yields

$$\begin{aligned}-\theta_t\psi_t(x) + \frac{\partial}{\partial t}\psi_t(x) &= \frac{1}{2}\frac{\partial^2}{\partial x^2}\big[\sigma^2(x)\psi_t(x)\big] - \frac{\partial}{\partial x}[\mu(x)\psi_t(x)].\end{aligned}$$

Integrating on $x$ from 0 to $\infty$ gives

$$\begin{aligned}-\theta_t + \psi_t(\infty) - \psi_t(0) &= \frac{1}{2}\frac{\partial}{\partial x}\big[\sigma^2(x)\psi_t(x)\big]\big|_0^\infty - [\mu(x)\psi_t(x)]\big|_0^\infty \\ &= \frac{1}{2}\big[2\sigma(x)\sigma'(x)\psi_t(x) + \sigma^2(x)\psi_t'(x)\big]\big|_0^\infty \\ &\quad - [\mu(\infty)\psi_t(\infty) - \mu(0)\psi_t(0)]\end{aligned}$$

Since $\psi_t(x)$ is a proper probability density, one will have $\psi_t(\infty) = 0$ and $\psi_t'(\infty) = 0$. Furthermore, the absorption at 0 implies $\psi_t(0) = 0$. The above equation then reduces to (4).

## 5. COMPARISON OF HAZARD RATES

Much of survival analysis focuses on relative hazard rates, or hazard ratios, often assuming in fact that the hazard rates are proportional, an assumption which despite its popularity is problematic on theoretical grounds. Two hazard rates, for different values of $c$, the distance from the point of absorption, are shown in Figure 8. Computing the hazard ratio, that is, one hazard divided by the other, reveals a strongly decreasing function as shown in the same figure. This feature, which is typical in these comparisons, is the same phenomenon which is observed in frailty models where the relative hazards usually (but not always) decline (see, e.g., Aalen, 1994). In the present setting it means that if a high-risk group is defined as being closer to the point of absorption than the low-risk group, then comparing the hazards in these two groups would give a declining hazard ratio. In fact, the convergence toward a quasistationary distribution implies that the relative hazards decline toward the value 1. Since, as mentioned in the Introduction, covariates will often be indicators of how far the underlying process has advanced, comparisons of the kind given here are clearly of relevance to the question of regression models in survival analysis and is yet another indication that proportional hazard models would not always be expected to give a true picture.

It is also of interest to compare the hazard rates when the starting point $c$ is the same, while the drift is different. The result of this is shown in Figure 9, where a very different picture emerges from the previous one. In fact, the hazard rates seem more or less parallel after some time has passed. It is interesting to note the similarity with the divorce rates in Figure 2; this might give a rough indication that increasing divorce rates are to a large extent due to cohort effects, with the drift toward divorce increasing in the more recent cohorts.

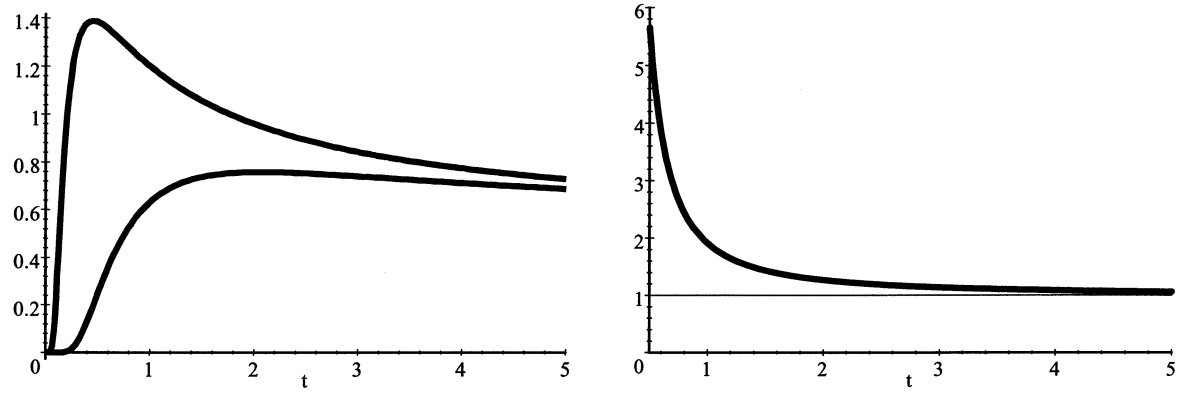


FIG. 8. (*Left panel*) *hazard rates for time to absorption when process starts out in* $c = 1$ (*upper curve*) *and* $c = 2$ (*lower curve*). (*Right panel*) *ratio of the two hazard rates. Parameters* $\mu = 1$ *and* $\sigma^2 = 1$.

Hence, when comparing hazard functions the result is very dependent on the causes of the differences between the groups. Difference in distance from point of absorption gives one type of result, while difference in the basic dynamics of the process, for example, drift and variance of a Wiener process, gives another type of result.

## 6. WIENER PROCESS WITH RANDOMIZED DRIFT

It is often natural to think that some degree of tracking is taking place, that is, that some individuals have a higher drift than others. This makes the processes into more flexible models, and we will consider this for the Wiener process with absorption. The results will be used in the regression model below. The model studied here is related to the frailty models, since there is individual heterogeneity.

Assume that the drift is randomized according to a normal distribution with expectation $-\mu$ and variance $\tau^2$ independent of the Wiener process. We will put the variance coefficient $\sigma^2$ of the Wiener process equal to 1. As noted above for the inverse Gaussian distribution, this entails no loss of generality when considering time to absorption.

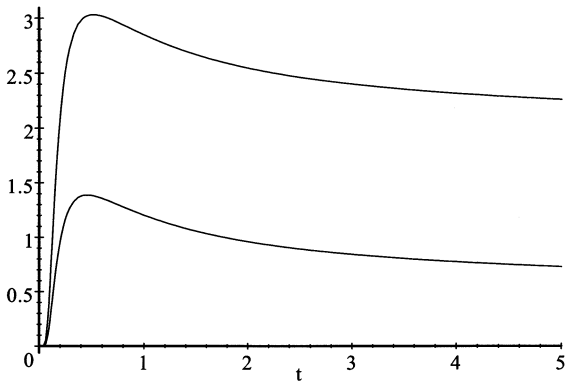


FIG. 9. *Hazard rates for time to absorption when process has drift parameters* $\mu = 2$ (*upper curve*) *and* $\mu = 1$ (*lower curve*). *In both cases* $c = 1$ *and* $\sigma^2 = 1$.

By integrating with respect to distribution of the drift parameter in (1) one gets the following density for transition from $c$ to $x$ over a time period of length $t$:

$$
\begin{aligned}
&b(x, c, \mu, t, \tau) \\
&\quad = \frac{1}{\sqrt{2\pi(t^2\tau^2 + t)}}\left\{\exp\left[-\frac{(x - c + \mu t)^2}{2(t^2\tau^2 + t)}\right]\right. \\
&\qquad - \exp\left(2c\mu + 2c^2\tau^2\right) \\
&\qquad \left.\times \exp\left[-\frac{(x + c + 2ct\tau^2 + \mu t)^2}{2(t^2\tau^2 + t)}\right]\right\}.
\end{aligned} \tag{5}
$$

(The computation of this distribution is basically a straightforward integration manipulating the quadratic forms in the exponents.) The probability of not being absorbed by time $t$ may be found, for example, in Aalen (1994):

$$
\begin{aligned}
B(t, c, \mu, \tau) &= \Phi\left(\frac{c - \mu t}{\sqrt{t^2\tau^2 + t}}\right) \\
&\quad - \exp\left(2c\mu + 2c^2\tau^2\right) \\
&\quad \times \Phi\left(\frac{-c - 2ct\tau^2 - \mu t}{\sqrt{t^2\tau^2 + t}}\right).
\end{aligned} \tag{6}
$$

The probability density of the position of the process at time $t$ conditioned on nonabsorption is found by dividing (5) by (6).

The probability density of time to absorption is the inverse Gaussian distribution with mixed drift parameter. Carrying out the integration yields the

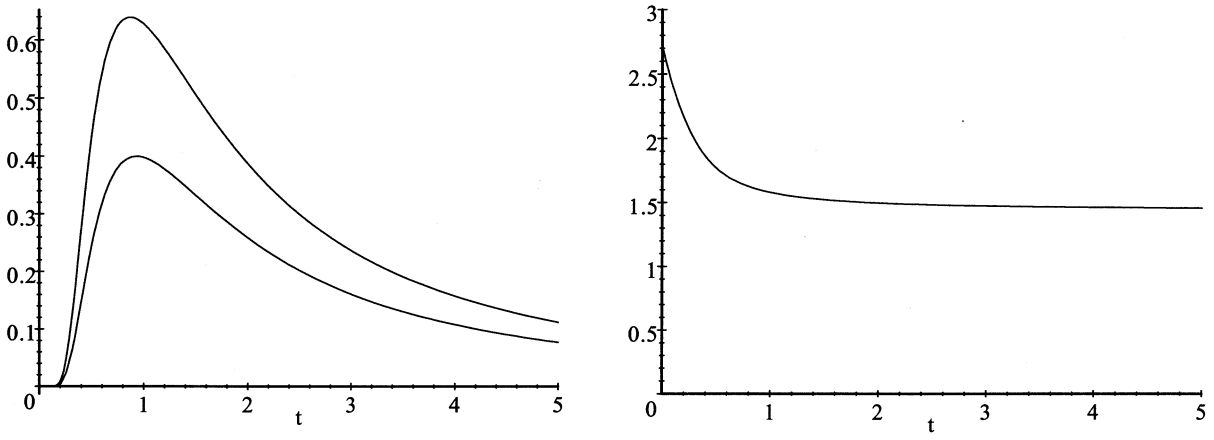


FIG. 10. *Wiener process with randomized drift*: (*left panel*) *hazard rates for time to absorption when process has mean drift parameters* $\mu = 1$ (*upper curve*) *and* $\mu = 0.5$ (*lower curve*); (*right panel*) *ratio of the two hazard rates. Parameters* $c = 2$ *and* $\tau^2 = 1$.

following density (Whitmore, 1986b; Aalen, 1994):

$$f_b(t, c, \mu, \tau) = \frac{c}{\sqrt{2\pi}} \frac{1}{t\sqrt{t^2\tau^2 + t}} \times \exp\left[-\frac{(c - \mu t)^2}{2(t^2\tau^2 + t)}\right]. \tag{7}$$

It should be noted that this is a defective survival distribution, because the crossing into the absorbing state may never take place. Defective survival distributions are important in practice, often under the name of "cure models" and may also be studied by means of frailty theory; see Aalen (1992). The hazard rate for the time to absorption is given by

$$\lambda_b(t, c, \mu, \tau) = \frac{f_b(t, c, \mu, \tau)}{B(t, c, \mu, \tau)}.$$

When $t$ increases the density and the hazard rate go to zero as $1/t^2$, and hence the hazard rates for different parameters will be asymptotically proportional. An example showing that approximate proportionality may be achieved very fast is given in Figure 10.

## 7. ANALYZING THE EFFECT OF COVARIATES

A test of the usefulness of the models presented here is whether they can in practice be used to analyze survival data with covariates. We will use the Wiener process with absorption and randomized drift, that is, a mixed inverse Gaussian distribution.

The model allows us to distinguish two different types of covariates, namely those which really only represent measures of how far the underlying process has advanced, and those which represent *causal* influences on the development. It is natural to model the first type as influencing the distance from absorption, that is, the parameter $c$, while the causal covariates influence the drift parameter $\mu$.

The distinction between two types of covariates is similar to the distinction between *internal* and *external* covariates (Kalbfleisch and Prentice, 1980), or *endogenous* and *exogenous* effects, that is commonly considered in economics.

Writing the likelihood for a set of, possibly censored, data is quite easy. The likelihood is a product of factors as follows: An individual which is censored at time $t$ has a likelihood contribution equal to $B(t, c, \mu, \tau)$. An individual which is noncensored and experiences an event at time $t$ has a likelihood contribution equal to $f_b(t, c, \mu, \tau)$. One or both of the parameters $\mu$ and $c$ are appropriate functions of covariates. The likelihood can be maximized by standard programs. The analysis presented here is carried out by the program GAUSS.

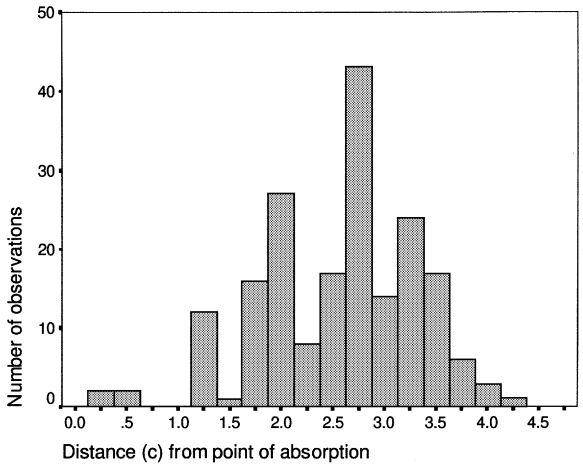


FIG. 11. *Regression model: histogram of estimated distances from point of absorption for all individuals.*

TABLE 1
*Estimated coefficients in the regression model; survival time is measured in units of* 100 *days*

| Parameter | Estimate | s.e. | Wald test |
|---|---|---|---|
| $\mu$ | 0.341 | 0.112 | 3.06 |
| $\tau^2$ | 0.061 | 0.080 | — |
| $a$ | 5.73 | 0.996 | 5.75 |
| Sex | 0.101 | 0.292 | 0.35 |
| Condition | −0.884 | 0.198 | −4.46 |
| T-stage | −0.548 | 0.201 | −2.72 |
| N-stage | −0.224 | 0.134 | −1.68 |

### 7.1 Example

Kalbfleisch and Prentice (1980, page 225) present a set of survival data concerning treatment of carcinoma of the oropharynx. These come from a clinical trial carried out by the Radiation Therapy Oncology Group in the United States. Survival time is measured in days from diagnosis and some of the patients are censored. The following covariates, measured at diagnosis, will be considered here: $Z_1$ = sex (1 = male, 2 = female), $Z_2$ = condition (1 = no disability, 2 = restricted work, 3 = requires assistance with self care, 4 = confined to bed), $Z_3$ = T-stage (an index of size and infiltration of tumor ranging from 1 to 4, with 1 indicating a small tumor and 4 a massive invasive tumor), $Z_4$ = N-stage (an index of lymph node metastasis ranging from 0 to 3, with 0 indicating no evidence of metastases and 3 indicating multiple positive nodes or fixed positive nodes). Data are given for 195 patients, 2 of whom (cases 136 and 159) are not included in the analysis due to missing values.

Apart from the covariate indicating sex, the covariates above are clearly of the character that they measure how advanced the disease is at time of diagnosis. In our terminology they measure how far some underlying disease process has advanced, and it seems natural that the modeling should take this into consideration. We apply here an inverse Gaussian distribution with randomized drift, as given by the density (7). The regression model is placed on the parameter $c$, which indicates how far the process is from the state of absorption. Specifically, we assume that this parameter depends on the covariates in a linear fashion:

$$c = a + b'z,$$

where $a$ and $b$ are coefficients and $z$ is the vector of covariates. Since $c$ must be positive, it might seem more natural to use a function which is bound to be positive instead of the linear one applied here. However, the linear one seems to fit the data best. For practical reasons survival time is measured in units of 100 days when carrying out the statistical analysis.

Estimating by means of maximum likelihood produces the results shown in Table 1. Figure 11 shows the estimated values of $c$ for all individuals. For a check of goodness of fit two groups of individuals have been considered, namely those who have all three covariates $Z_2$, $Z_3$ and $Z_4$ either *above* their respective means or *below* their respective means. Hence these constitute, respectively, a high-risk and a low-risk group. Within each group the average survival curve according to the estimated model is computed. The negative logarithm of this is then compared to the Nelson–Aalen estimate of the cumulative hazard function within each group. The results are shown in Figure 12.

Although this is a simple example it indicates that the idea of covariates as measuring how advanced an underlying process is, might be useful. The estimated cumulative hazards in Figure 12 are clearly nonproportional when comparing the high-risk and the low-risk groups. Rather, the low-risk group shows a *delay* in the hazard for the first year or so (having almost zero hazard for a while), before it gradually catches up with the other group. This delay is a natural consequence when regarding the low-risk group as having less advanced disease than the other group. It is our experience that survival curves for groups with lower risk often show this feature of delay compared to groups with higher risk, which also implies nonproportionality. Admittedly, regression in survival analysis is often a coarse business where one is not very interested in the details of how the hazard rates develop. On the other hand, there is no harm in bringing the insight one may have into the modeling.

## 8. CONCLUSION

By postulating some underlying process one may gain insight into the properties of survival models. One may also make practical models along these lines and apply them to the analysis of survival

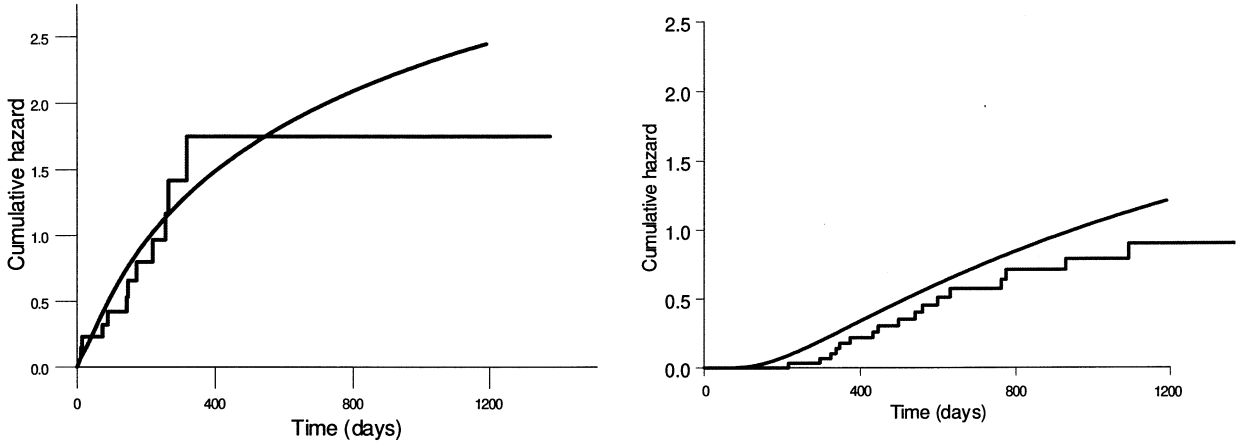


FIG. 12. *Cumulative hazards according to estimated model* (*smooth curve*) *compared with the Nelson–Aalen estimate* (*stepped curve*) *within a high-risk and a low-risk group.*

data. Probably models based on Wiener processes give the simplest formulas, but Markov processes on discrete state spaces may also be useful. A further development of the present models to incorporate time-dependent covariates might be of interest.

## REFERENCES


AALEN, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Ann. Appl. Probab.* **2** 951–972.

AALEN, O. O. (1994). Effects of frailty in survival analysis. *Statist. Methods Medical Res.* **3** 227–243.

AALEN, O. O. (1995). Phase type distributions in survival analysis. *Scand. J. Statist.* **22** 447–463.

AALEN, O. O., FAREWELL, V. T., DEANGELIS, D., DAY, N. E. and GILL, O. N. (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Statistics in Medicine* **16** 2191–2210.

CAVENDER, J. A. (1978). Quasi-stationary distributions of birth-and-death processes. *Adv. Appl. Probab.* **10** 570–586.

CHHIKARA, R. S. and FOLKS, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications.* Dekker, New York.

COX, D. R. and MILLER, H. D. (1965). *The Theory of Stochastic Processes.* Methuen, London.

DOKSUM, K. A. and HØYLAND, A. (1992). Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse gaussian distribution. *Technometrics* **34** 74–82.

EATON, W. W. and WHITMORE, G. A. (1977). Length of stay as a stochastic process: a general approach and application to hospitalization for schizophrenia. *J. Math. Sociol.* **5** 273–292.

GOEL, N. S. and RICHTER-DYN, N. (1974). *Stochastic Models in Biology.* Academic Press, New York.

GRIFFITHS, W. S. (1988). Shock models. In *Encyclopedia of Statistical Sciences* **8** 442–446. Wiley, New York.

GROSS, D. and HARRIS, C. M. (1985). *Fundamentals of Queuing Theory,* 2nd ed. Wiley, New York.

HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73** 387–396.

JEWELL, N. P. and KALBFLEISCH, J. D. (1996). Marker processes in survival analysis. *Lifetime Data Anal.* **2** 15–29.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

KARLIN, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.

KEIDING, N., ANDERSEN, P. K. and KLEIN, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16** 215–224.

KEILSON, J. (1979). *Markov Chain Models—Rarity and Exponentiality.* Springer, New York.

LANCASTER, D. J. (1982). *Stochastic Models for Social Processes,* 3rd ed. Wiley, New York.

LONGINI, I. M., CLARK, W. S., BYERS, R. H., WARD, J. W., DARROW, W. W., LEMP, G. F. and HETHCOTE, H. W. (1989). Statistical analysis of the stages of HIV-infections using a Markov model. *Statistics in Medicine* **8** 831–843.

LYNN, N. J. and SINGPURWALLA, N. D. (1997). "Burn-in" makes us feel good. Comment on "Burn-in," by H. W. Block and T. H. Savits. *Statist. Sci.* **12** 13–19.

MARTINEZ, S. and SAN MARTIN, J. (1994). Quasi-stationary distributions for a Brownian motion with drift and associated limit laws. *J. Appl. Probab.* **31** 911–920.

NIELSEN, J. P. and LINTON, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.* **23** 1735–1748.

O'CINNEIDE, C. A. (1990). Characterization of phase type distributions. *Comm. Statist. Stochastic Models* **6** 1–57.

ØKSENDAL, B. (1998). *Stochastic Differential Equations,* 5th ed. Springer, New York.

PATEL, J. K. (1983). Hazard rate and other classifications of distributions. In *Encyclopedia of Statistical Sciences* 3. Wiley, New York. pages: 590–594.

SHAKED, M. and SHANTHIKUMAR, J. G. (1991). Shock models with MIFRA time to failure distributions. *J. Statist. Plann. Inference* **29** 157–169.

SCHWARZ, W. (1992). The Wiener process between a reflecting and an absorbing barrier. *J. Appl. Probab.* **29** 597–604.

SESHADRI, V. (1998). *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Springer, New York.
WHITMORE, G. A. (1986a). First-passage-time models for duration data: regression structures and competing risks. *The Statistician* **35** 207–219.
WHITMORE, G. A. (1986b). Normal-gamma mixtures of inverse Gaussian distributions. *Scand. J. Statist*. **13** 211–220.
WHITMORE, G. A. (1995). Estimating degradation by a Wiener diffusion process. *Lifetime Data Anal*. **1** 307–319.
WHITMORE, G. A., CROWDER, M. J. and LAWLESS, J. F. (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Anal*. **4** 229–251.

# Comment

**Henry W. Block and Thomas H. Savits**

In an interesting and well-written article Aalen and Gjessing consider the shape of various hazard rates. They ask the following question: "Why, for instance, do hazard rates sometimes increase, sometimes decrease and sometimes increase and then decrease.. . ." The authors might also have asked why hazard rates decrease then increase, which is the pattern for the standard bathtub-shaped failure rate, but they do not present results which lead to this pattern.

Their elegant explanation for these phenomena comes by modeling survival distributions as first passage times of underlying processes and then explaining the monotonicities of the first passage times as a function of the behavior of the underlying processes.

The authors also mention that the shape of the hazard rate can depend on selection effects among individuals and in particular that frailty models tend to deal with this. Frailty models are special cases of mixture distributions where the conditional hazard rates are multiplicative functions of a baseline hazard. These are not, however, dealt with in the present paper. We would like to focus on these mixture models and discuss some of the various results which have been obtained in determining the shape of the hazard rate.

As mentioned by the authors, there has been a tendency in the literature to study hazard rates which are monotone. However, in practice, the shapes of hazard rates often do not follow these simple patterns. Departures from monotonicity were initially looked upon as aberrations. To initiate our discussion we look at some of these aberrations which we call "apparent anomalies."

*Henry W. Block is Professor, Department of Statistics, University of Pittsburgh, 2703 Cathedral of Learning, Pittsburgh, Pennsylvania 15260. Thomas H. Savits is Professor, Department of Statistics, University of Pittsburgh, 2734 Cathedral of Learning, Pittsburgh, Pennsylvania 15260.*

## APPARENT ANOMALIES

A much cited paper is Proschan (1963). In this paper pooled data for airplane air conditioning systems whose lifetimes are known to be exponential exhibit a decreasing failure rate. Since decreasing failure rates are usually associated with systems that improve with age, this was initially thought to be counterintuitive.

A second anomaly, at least to some, was that mixtures of lifetimes with increasing failure rates could be decreasing on certain intervals. Examples of such lifetimes can be found in Vaupel and Yashin (1985) as well as Barlow and Proschan (1975).

A variant of the above is due to Gurland and Sethuraman (1994, 1995) and gives examples of mixtures of very rapidly increasing failure rates which are eventually decreasing.

A recent paper by Wang, Muller and Capra (1998) (and many articles cited there) mentions that in many biological populations, including human ones, lifetimes of organisms at extreme old age exhibit decreasing hazard rate. A natural question to ask is whether this improvement extends to individuals in the population or not.

## SOME EXPLANATIONS

Before discussing issues and solutions raised by the above anomalies, we provide some standard language. In general we consider populations of lifetimes which are heterogeneous. Often these populations consist of a number of subpopulations (sometimes only two) having distributions of a similar type (e.g., the same distribution with different parameters). The population lifetime distribution can then be modeled as a mixture of the lifetime

distributions of its various subpopulations. The mixture can sometimes be of a specific form such as having a multiplicative failure rate (a frailty model) or an additive failure rate. We also use the terms hazard rate and failure rate interchangeably.

The apparent anomalies of the previous section lead to the following modeling issues:

- How do eventual failure rates of subpopulations compare with the eventual failure rate of the population?
- Which mixtures have failure rates that are eventually decreasing?
- Which populations experience eventual wearout?
- What causes mixtures of subpopulations with increasing failure rates to have failure rates which decrease on certain intervals and is there eventual wearout?
- When do mixtures have typical (e.g., bathtub-shaped) failure rates?

In the Proschan (1963) paper, the author resolved the apparent anomaly by showing that mixtures of exponential distributions have decreasing failure rates. In fact, it was shown that mixtures of lifetimes with decreasing failure rates have decreasing failure rates.

To discuss the other anomalies we introduce the following notation. For $w \in S$, where $S$ is a set of real numbers, we consider the lifetime density function $f(t, w)$, the survival function $\overline{F}(t, w)$ and the failure rate function $r(t, w) = f(t, w)/\overline{F}(t, w)$. The failure rate of the mixture is given by

$$r(t) = \frac{\int_S f(t, w)P(dw)}{\int_S \overline{F}(t, w)P(dw)}.$$

Block, Mi and Savits (1993, Theorem 4.1) gave a general result relating the limits of $r(t, w)$ and $r(t)$. Under the assumption

$$r(t, w) \to a(w) \quad \text{as } t \to \infty \text{ for } w \in S$$

(and some technical conditions) it is shown that

$$r(t) \to \alpha = \inf\{a(w) | w \in S\};$$

that is, the failure rate of the mixture converges to the strongest limiting failure rate. Note that the above applies to general mixtures, including both discrete and continuous mixtures.

The above result is pertinent to the first of the modeling issues. It also helps to explain the third anomaly, where a failure rate which is very rapidly increasing is mixed with an exponential distribution with stronger failure rate. By the above result, the failure rate of the mixture must eventually come down to the stronger exponential failure rate. This also gives an intuitive explanation for the last anomaly. If there are stronger and weaker components, the failure rate should eventually come down to the stronger failure rate. In other words, it is the weaker components that cause the failure rate of the mixture to decrease toward the stronger component as the weaker ones die out.

A second result (Block and Joe, 1997, Theorem 2.3) which holds for finite mixtures (we state the result for a mixture of two) is that if $r_1(t) = r(t, w_1)$ and $r_2(t) = r(t, w_2)$ are the failure rates for two subpopulations, where $r_2$ is eventually stronger than $r_1$ (i.e., $r_1(t) > r_2(t)$ for large $t$) and $r_1$ and $r_2$ are monotone and converge at polynomial rates, then the monotonicity of the mixture is eventually in the same direction as the monotonicity of the stronger component $r_2(t)$. That is, the mixture has the same eventual monotone behavior as its strongest component. In particular a mixture wears out (i.e., it has eventual increasing failure rate) if the strongest component wears out. This has bearing on the last anomaly. Assume that a population has two subpopulations (e.g., males and females). The weaker population (e.g., males) dies out first, leaving the stronger population (e.g., females). This should cause a decrease in the failure rate of the mixed population at extreme old age. However, if the stronger population by itself exhibits wearout, by the above result the mixed population should also eventually exhibit wearout. Block and Joe (1997) also consider many results where mixtures have eventually decreasing failure rates.

The comments of Lynn and Singpurwalla (1997) and the paper of Finkelstein and Essaoulova (2000) consider particular mixture models. The Lynn–Singpurwalla comments propose an additive model and the Finkelstein–Essaoulova paper considers both the additive model and a multiplicative model. These two models have the form

$$\text{(Additive)} \quad r(t, w) = w + r(t)$$

and

$$\text{(Multiplicative)} \quad r(t, w) = wr(t).$$

Notice that the multiplicative model is a frailty model.

Lynn and Singpurwalla (1997) state without proof the result that, for the additive model, $r(t)$ has a bathtub shape if certain conditions involving $E(W \mid T)$ and $\text{Var}(W \mid T)$ hold, where $T$ is the mixture lifetime and $W$ is the mixture variable. Finkelstein and Essaoulova (2000) prove this result, but require additional conditions. These latter authors have also given some results on asymptotic failure rates of the type mentioned above but under conditions involving the conditional mean and variance.

### FRAILTY

We conclude with some special cases of our results applied to the frailty case. The frailty model is a special case of the general mixture model where $r(t, w) = w\lambda(t)$, $w \in S$ and $S$ is a set of real numbers. Consequently

$$\overline{F}(t, w) = \exp\left\{-w \int_0^t \lambda(u)\, du\right\}.$$

Under the assumptions that $\lambda(t) \to a$ as $t \to \infty$ for $0 \le a \le \infty$, that $S$ is a bounded set and that for the case $a = \infty$ there are constants $K$ and $L$ such that $\lambda(t) \le K \exp Lt$, for large $t$, it is easy to show that the conditions of Theorem 4.1 of Block, Mi and Savits (1993) hold. In this case

$$r(t) \to \alpha = \inf\{wa | w \in S\}.$$

That is, the overall failure rate for the frailty model converges to the strongest limiting failure rate.

To determine the eventual direction of the monotonicity, assume that $r_1(t) = w_1\lambda(t)$ and $r_2(t) = w_2\lambda(t)$, where $w_1 < w_2$, and also that $\lambda(t)$ converges monotonically to $\alpha$, $0 \le \alpha \le \infty$, as $t \to \infty$ at a polynomial rate (see Block and Joe, 1997). Then the eventual monotonicity of $r$ is the same as the monotonicity of $\lambda$.

# Comment[1]

**Lynn E. Eberly, Patricia Grambsch and John E. Connett**

### 1. INTRODUCTION

The authors are to be congratulated; they are performing a signal service in reintroducing first-passage-time models to the statistics community. As the authors point out, these models allow predictor variables to perform two distinct roles: to mark the initial state or to influence the rate of diffusion through the state space to absorption. This flexibility sets these models apart from standard models for time-to-event data, such as proportional hazards or accelerated failure time, and gives them great appeal for disease modeling. Our discussion focuses on modeling survival in AIDS patients using the Wiener process with randomized drift, the last process described. We first summarize some simple analyses showing the promise of this model and then discuss some conceptual and computational issues which must be addressed before it can become a practical addition to the statistical toolkit.

### 2. MODELING ISSUES

Our data came from the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA), a network of 15 community-based clinical units studying HIV-related therapies in a primary-care setting. The CPCRA has an extensive database on 12,330 HIV-infected individuals who have enrolled in one or more CPCRA protocols from September 1990 through January 2000. We used a random subset from this database (15%, $n = 1{,}612$) and considered two predictors, age (in years) and CD4+ cell count (in cells/mm$^3$), measured at each individual's initial enrollment. There were 574 deaths with a median follow-up time of 565.5 days. CD4+ count, a measure of immune system strength, is a well-established predictor of survival with higher values associated with longer survival; the square root scale was used due to the skewness of its distribution. Median baseline CD4+ count was 142.5 cells/mm$^3$. Age is weekly associated with survival, where older ages are associated with shorter survival. Median age at baseline was 38 years.

The model parameters, initial state $c$ and negative expected drift $\mu$, were for most analyses taken to be exponentiated linear functions of the predictors to keep $c > 0$ and $\mu > 0$. Note that either $\tau^2 \neq 0$ or $\mu \le 0$ leads to a curve-type model with a defective survival distribution which may not be appropriate for some data. The models were fitted according to a likelihood built from Aalen and Gjessing

*Lynn E. Eberly is Assistant Professor, Division of Biostatistics, School of Public Health, University of Minnesota, 420 Delaware Street SE, MMC 303, Minneapolis, Minnesota 55455-0378 (e-mail: lynn@biostat.umn.edu). Patricia Grambsch is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota (e-mail: pat@biostat.umn. edu). John E. Connett is Professor, Division of Biostatistics, School of Public Health, University of Minnesota (email: johnc@ccbr.umn.edu).*

[1]*This is Division of Biostatistics Research Report 2000-014.*

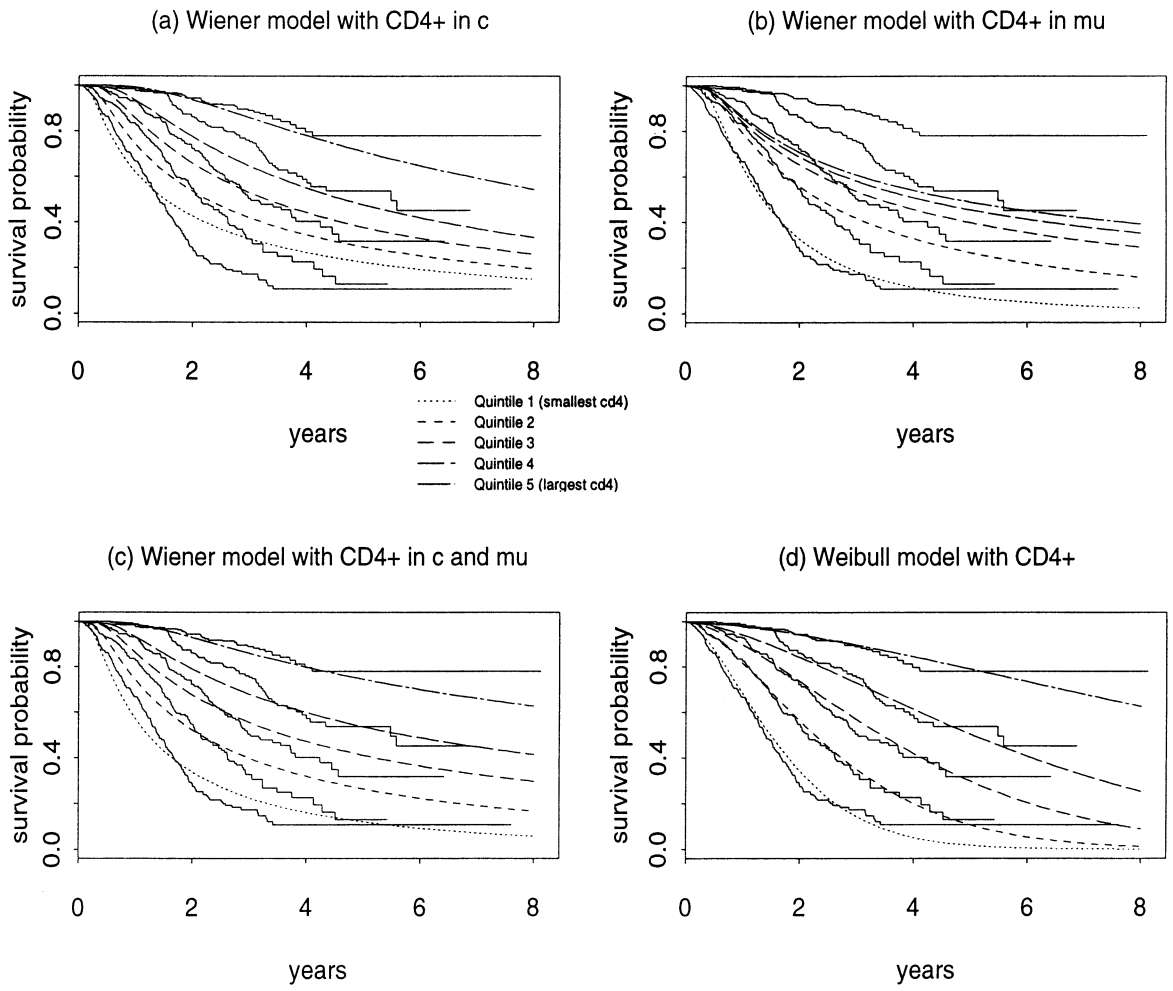


FIG. 1. *Fitted survial curves superimposed on Kaplan–Meier curves, each stratified by CD4+ quintile. Fitted curves from Wiener models with square root CD4+* (a) *in c,* (b) *in* $\mu$, (c) *in both* $\mu$ *and c and* (d) *from a Weibull regression model with square root CD4+.*

equations (6) and (7) using a standard minimization algorithm (`nlminb()` in S-PLUS version 3.4) on the outcome "years to death" (range of values roughly 0–8). We supplied a gradient but not a Hessian. The "null" model with no predictors had a maximized log likelihood of $-1590.756$ with parameter estimates $\hat{c} = 1.307, \hat{\mu} = 0.007$ and drift variance $\hat{\tau}^2 = 1 \times 10^{-7}$.

The square root of CD4+ count was a highly significant predictor (see our Table 1) whether modeled as affecting $\mu$ only or $c$ only. From a medical perspective, either makes sense, so we then modeled both $c$ and $\mu$ as functions of the square root of CD4+ count, which resulted in a significantly better fit according to the log likelihood. This finding suggests that CD4+ acts not merely as a marker of disease state but may also be predictive of how fast the disease progresses. This type of insight is not available from the typical time-to-event models and illustrates one of the most exciting aspects of the Wiener process approach.

Our Figure 1 shows Kaplan–Meier plots of survival, stratified by quintiles of CD4+ count, with superimposed fitted survival plots from each of the models considered. The fitted curves were computed by averaging fitted survival curves over the members of each quintile. We can see that having CD4+ affect only $c$ (so that all individuals are forced to have a common rate of progression $\mu$) inadequately models the lowest CD4+ quintile (Figure 1a). This supports the conjecture that those in the highest quintile progress toward absorption at a significantly slower rate, which cannot be compensated for by the dependence of the *starting point* on CD4+. Likewise, having CD4+ affect only $\mu$ (so that all individuals are forced to have a common starting state $c$) inadequately models the highest

TABLE 1
*Maximum likelihood results for four Wiener process models with parameters as exponentiated functions of CD4+*

| Model | Maximized log likelihood | Estimated slope for $\sqrt{CD4+}$ in $c$ | Estimated slope For $\sqrt{CD4+}$ in $\mu$ |
|---|---|---|---|
| Null (no CD4+) | −1590.76 | 0 | 0 |
| $\sqrt{CD4+}$ in $c$ only | −1405.62 | 0.056 | 0 |
| $\sqrt{CD4+}$ in $\mu$ only | −1489.77 | 0 | −0.207 |
| $\sqrt{CD4+}$ in both | −1388.71 | 0.044 | −0.132 |

CD4+ quintile (Figure 1b). Perhaps those in the lowest quintile start significantly closer to absorption than those in other quintles, which cannot be compensated for by the dependence of the *rate* on CD4+. Allowing both $c$ and $\mu$ to be functions of CD4+ adequately models all quintiles, as shown in Figure 1c.

The Wiener process with randomized drift also forces the hazard of death to be 0 at time 0. Although certain model parameter values can direct the hazard to increase very quickly away from 0, this may be inappropriate for HIV and AIDS survival data. We next compared the fit of our Wiener models to a Weibull fit with square root of CD4+ count as the sole predictor. Even when both $c$ and $\mu$ are defined as functions of CD4+, the fit of the Wiener model seems inferior to that of a Weibull model (Figure 1d). This is particularly true during the first four years of follow-up (when most deaths occurred). This may be an indication that the Wiener model is not flexible enough to accommodate the need in our dataset for a nonzero hazard at time 0.

When we removed the constraint that $\mu > 0$ from the model with CD4+ count in $\mu$ only (by modeling $\mu$ simply as a linear function of the square root of CD4+ count), we obtained fitted values of $\mu$ which were negative (meaning drift away from the absorbing state) for many of the patients with high CD4+ values. Our Figure 2 shows that the highest quintile is now much more accurately modeled than in Figure 1b, thus indicating that this subpopulation may require a cure-type model. This presents an interesting dilemma, since a cure-type model is not appropriate for the lowest quintile. Surprisingly, the model with $\mu > 0$ and CD4+ in both $c$ and $\mu$ somehow compensates for the highest quintiles needing a cure model without forcing $\tau^2$ to be nonzero (here, $\hat{\tau}^2 = 1 \times 10^{-7}$).

Models using age as the only predictor maximized the log likelihood at −1589.210 when $\mu$ was an exponentiated linear function of age, but at −1589.056 when $c$ was modeled as an exponentiated linear function of age. Thus the likelihood ratio test (LRT) statistics were 3.09 and 3.41, respectively, with 1 degree of freedom each. Age is thus not strongly associated with either the rate of progress toward death or the starting point ($\alpha = 0.05$). However, when $\mu$ was modeled as a (nonexponentiated) linear function of age, we obtained a log likelihood of

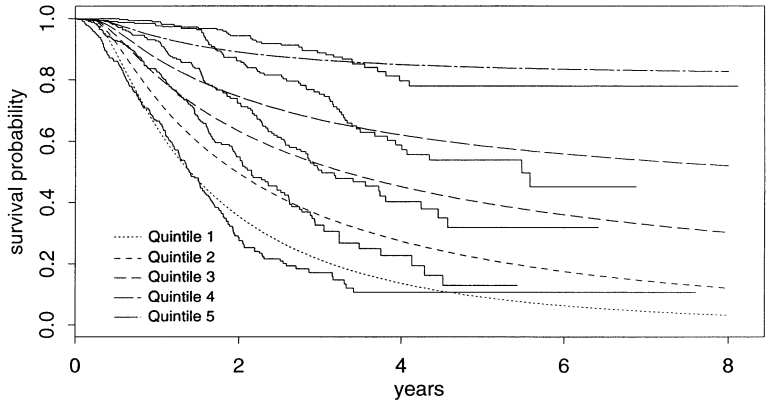


FIG. 2. *Fitted survival curves superimposed on Kaplan–Meier curves, each stratified by CD4+ quintile. Fitted curves from Wiener models with square root CD4+ in $\mu$, where $\mu$ is not constrained to be nonnegative.*

$-1588.384$ and a LRT statistic of 4.744 (p-value = 0.03). An estimated slope of $+0.03$ shows that older people progress faster toward death than younger, as expected. Here we see conclusions regarding the association of age with survival which differ depending on the scale in which the covariate appears, but with such marginal p-values (just above 0.05 for the exponentiated scale and just under for the linear scale), this example does not provide convincing evidence that this is a genuine problem.

### 3. FURTHER NEEDED DEVELOPMENTS

We have found the Wiener process model with randomized drift to be very flexible and appealing. However, there are many computational and modeling issues yet to be worked out. As with most nonlinear models, multimodal likelihoods are possible; we found at least one instance where the algorithm converged to a local rather than global maximum. Starting values are critical to the successful convergence of the numerical likelihood maximization algorithm. It is not clear if identifiability issues arise when covariates are incorporated into both $c$ and $\mu$, but we had no such difficulties. Interpretation of this model even without covariates is difficult. Note that if $\tau^2 = 0$, we have a proper inverse Gaussian with mean $c/\mu$, but it is not clear either how to interpret $c$ and $\mu$ when $\tau^2 \neq 0$ or how to interpret the magnitude of covariate regression coefficients in any model.

The scale of the covariates is something only briefly touched on by Aalen and Gjessing. The question is whether to exponentiate the linear combination of covariates, thereby forcing the parameters to be nonnegative. We have shown that this can make quite a difference, and the decision may depend on whether a cure-model makes scientific sense for all members of the population under study. We also fit models with additional covariates for Karnofsky score, progression of disease and anti-retroviral use in $c$ and for gender, ethnicity, intravenous drug use and homosexual contact in $\mu$. Using an exponentiated form for the covariates gave sometimes better and sometimes poorer results in terms of the maximized log likelihood. In general, it did seem slightly more difficult to obtain convergence in models with exponential parameterizations of $c$ and $\mu$.

Model building was straightforward, as we found during our exploration of the 9-covariate model mentioned above, since the likelihood is fully parametric and nested models can be compared with LRTs. Standard error estimates can also be obtained, as demonstrated by Aalen and Gjessing. Clearly, diagnostic tools with residuals, outlier identification and ways to determine in which parameter a particular covariate should be placed are needed. One of the attractive features of other time-to-events models is the ability to stratify on a factor such as clinical unit. It is well established that there are survival differences among clinical units in the CPCRA, but it seems excessive and perhaps risky from a numerical algorithm point of view to introduce 14 indicator variables into one or both of the covariate vectors. It would be useful for the Wiener model to be adapted to such a situation.

### ACKNOWLEDGMENTS

First author supported in part by NIH Contract 5 U01 AI42170-07. Third author supported in part by NIH Grant 5 U10 HL59275 03.

# Comment

## Niels Keiding

Hazard rate models are ubiquitous in survival analysis: Tetens (1786) formulated the explicit one-parameter (nonproportional) hazards model

$$\frac{\lambda(t)}{\lambda_0(t)} = \frac{1 + 2\alpha S(t)}{1 + \alpha S_0(t)}$$

for the ratio between the "instantaneous decrement" $\lambda(t)$ in the select group and that $(\lambda_0(t))$ of the general population. Here $S(t)$ and $S_0(t)$ are the corresponding survival functions since a fixed age at entry.

*Niels Keiding is Professor, Department of Biostatistics, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark.*

Very recently Bagdonavicius and Nikulin (1999) formulated an ambitious approach to a unified framework encompassing many survival models. Their *generalized proportional hazards model* was specified as

$$\lambda_{X(\cdot)}(t) = \lambda_0(t) r(X(t)) q(\Lambda_{X(\cdot)}(t)),$$

where $X_{(\cdot)}$ is a time-dependent covariate function, $r$ and $q$ are positive functions and

$$\Lambda_{X(\cdot)} = \int_0^t \lambda_{X(\cdot)}(s)\, ds.$$

Choosing $r(X(t)) = e^{\beta X(t)}$ and $q(\Lambda_{X(\cdot)}) = e^{\gamma \Lambda_X(t)}$ leads to what they call the generalized linear proportional hazards model.

The Cox (1972) proportional hazards model

$$\lambda_i(t) = \lambda_0(t) e^{\beta X_i}$$

and other hazard regression models allowed control for *observed* confounders in survival analysis. However, *unobserved heterogeneity* required its own development.

With the primary aim of illustrating various demographic consequences of admitting random heterogeneity Vaupel, Manton and Stallard (1979) let a positive random variable $Z$ (termed *frailty*) multiply an underlying standard hazard rate $\lambda_0(t)$. This model was motivated primarily by technical convenience, as was the case for the Cox model.

Vaupel, Manton and Stallard's main message was to formalize a concept of survivor selection deriving from the fact that the *frail die first* so the *individuals age faster than cohorts*: the hazard of a random survivor at age $a$ is $\lambda(a) E(Z \mid \text{survival to } a)$. Vaupel, Manton and Stallard even proposed a calculus for correcting cohort life tables into "individual life tables."

That both Cox (1972) and Vaupel, Manton and Stallard (1979) chose to model heterogeneity (deterministic and random, respectively) multiplicatively on the hazard scale made it obvious to combine the models into a *proportional hazards model with frailty*

$$\lambda_i(t|Z_i) = \lambda_0(t) Z_i e^{\beta X_i}.$$

Vaupel, Manton and Stallard (1979) ended on an optimistic note, having identified random heterogeneity as a possibly very important aspect of mortality variation, neglect of which might lead to serious bias. Vaupel, Manton and Stallard did note that frailty "however defined, is difficult to measure."

Subsequent literature has documented that this may be rather an understatement. In the simple one-sample problem of $n$ independent (possibly left-truncated and/or right-censored) survival times, the marginal "population" distribution is observed and no empirical distinction between individual hazard and frailty (mixing) distribution is possible. For the frailty model with proportional individual hazards the frailty distribution is in principle identifiable, at least if it has finite expectation. However, there is no way of empirically verifying the proportionality assumption, and even if it were true, Hougaard, Myglegaard and Borch-Johnsen (1994) and Keiding, Andersen and Klein (1997) provided case studies to show that the regression parameter estimates may be very dependent on the hardly identifiable choice of frailty distribution, while reparametrization in terms of accelerated failure time models yielded much better identifiability of the parameters. This situation lead Robins and Greenland (1989, 1991) to argue forcefully against practical use of the "individual hazard ratio," specifically in connection with compensation schemes in occupational insurance.

At this point it is important to step back and recall the definition of hazard. Particularly as explicitly formulated by Scheike, Petersen and Martinussen (1999), the interpretation of the regression parameters in this model requires conditioning on all observed covariates (as usual) *but also on the unobserved and unobservable frailty* $Z$. In some cases this is plainly uninterpretable in practice, and the whole concept of individual frailty is really, in Aalen and Gjessing's phrase, *elusive*.

I welcome Aalen and Gjessing's attempt at providing a fresh start for the interpretation of the hazard rate, and I look forward to learning how this approach will cope with the interplay between observed and unobserved heterogeneity.

## ACKNOWLEDGMENT

This work was supported by National Cancer Institute Contract R01 CA54706-07.

# Rejoinder

**Odd O. Aalen and Håkon K. Gjessing**

We are grateful to the discussants for taking the time and effort to comment on our paper.

Keiding reviews the use of hazard functions in the Cox model and in frailty theory, and supports our assertion that hazard rate is a concept that needs clarification. Frailty theory has been an attempt at achieving a deeper understanding of the hazard rate and of the selection effects which influence the hazard. It is our opinion that frailty theory, in spite of its somewhat speculative character, has made an important contribution to a deeper understanding of issues in survival analysis. However, as pointed out by Keiding, the situation is far from satisfactory, and we agree with him that a fresh start is needed and hope that the one we offer here will be of use.

Block and Savits present a reliability point of view. In reliability theory there is a long and strong tradition of studying the shapes of hazard rates with a number of elegant mathematical results. It is important that biostatisticians also be aware of this literature, and the review by Block and Savits is a good starting point. Our approach differs from the main tradition in this field in that we focus on properties of the underlying process, in particular the quasistationary distribution.

Block and Savits mention bathtub-shaped hazard rates, which is an important theme in reliability theory, because technical products often have a higher incidence of failures early in their life, whereafter the rate decreases, before increasing again due to ageing. In fact, bathtub-shaped hazards may easily be derived within our framework. An example within the phase type distributions is given by Aalen (1995, Figure 4).

Eberly, Grambsch and Connett give a detailed discussion of our regression model for analyzing the effect of covariates. The model using randomized drift for a Wiener process is primarily intended as an illustration of the principles discussed in the paper. However, we see it as important that it should be possible to implement the ideas in specific models, and we need more practical knowledge about relevant models. It is thus very interesting to see the model applied to a different type of data, and we appreciate the effort Eberly, Grambsch and Connett have taken in exploring the model. Several of the issues concerning the application of this particular parametric model to real data are of course well known from other parametric modeling situations. The model has a limited range of possible hazards, and inherently the hazards are nonproportional, especially when varying $c$. Clearly, the successful fit of a Weibull model suggests that the proportional hazards assumption is valid throughout the time period and that in this case the randomized drift model should not be expected to be an improvement.

Scale issues are essential to any modeling discussion. For our data we did not encounter problems with the restriction $c > 0$, so exponentiation was not needed to avoid impossible likelihood values. Our experience from other datasets suggests that $\mu$ should not be restricted, thus allowing for a substantial probability of never hitting the barrier. We note that Eberly, Grambsch and Connett have restricted $\mu$ to be positive in most of their analyses, and their Figure 2 shows that dropping this restriction improves the situation. With an unrestricted $\mu$ one would expect a considerably better fit than is indicated in Figure 1 of their discussion. Whether or not, say, exponentiation (or other transformations) yields a better fit than the additive is another matter. We found the additive to provide the best fit, but this is a discussion that will apply to any modeling situation, and more experience needs to be accumulated for this particular model.

Eberly, Grambsch and Connett comment on the interpretation of the parameters in the model. Since the distribution, with $\tau^2 > 0$, is defective, it follows that expectation and variance are not well defined. However, the main parameters, $\mu$ and $c$, are well defined in terms of the underlying process and have a simple intuitive meaning. Furthermore, changing the time scale of the model has a simple effect on the parameter estimates. Scaling time to a new variable $t^* = t/k$, where $k$ is a constant, yields the exact same survival function when the parameters $\mu$, $c$ and $\tau$ are replaced by $\mu^* = \mu\sqrt{k}$, $c^* = c/\sqrt{k}$ and $\tau^* = \tau\sqrt{k}$, respectively. Thus, $B(t, c, \mu, \tau) \equiv B(t/k, c^*, \mu^*, \tau^*)$ as can be verified from (6) or numerically. [Of course, if an exponential model $c = \exp(a + b'z)$ is used, this will correspond to a shift $-(1/2)\log(k)$ of $a$ and leave $b$ unchanged, whereas in an additive formulation $c = a + b'z$ both $a$ and $b$ will be scaled.] This scaling property can be understood from the fact that if $W_t$ is a Wiener process, then so is $W^*_{t^*} = (1/\sqrt{k})W_{t^*k}$; that is, a scaling of $1/k$ in time can be offset by a corresponding scaling $1/\sqrt{k}$ of the distance axis, which in turn leads

to a scaling $(1/\sqrt{k})/(1/k) = \sqrt{k}$ of $\mu$ and $\tau$, which are in units of distance/time.

In our analysis we used a (rather arbitrary) scale of days/100. To recover the results for the original time scale the value for $\mu$ in Table 1 should be divided by 10, the value for $\tau^2$ divided by 100, and the estimates for $a$ and the covariates should be multiplied by 10. Apart from that, the attained maximum likelihood is identical.

For the datasets where we have successfully fitted the randomized drift model, our experience has been that it is important to obtain good initial values primarily for $\mu$ and $c$. The value of $\tau$ is usually found during the estimation from a starting value, say, between 0 and $\mu$. When covariates are present, obtaining a good starting value for the intercept $a$ in the centered model $c = a + b'(z - \bar{z})$ seems sufficient; $b$ is entered with starting value 0. To obtain the starting values for $\mu$ and $a$ we first computed the Kaplan–Meier curve for the data, disregarding covariates. From the Kaplan–Meier curve we estimated values for the mean and harmonic mean of the survival time. Finally, these values were used in the explicit formulas for the maximum likelihood estimators available in the case of no censoring (Chhikara and Folks, 1989). Although far from perfect, it at least gives starting values of the correct order of magnitude. We have made a set of simple S-PLUS functions for computing the initial values and performing the maximum likelihood estimation available at http://www.uio.no/~hakong/invgauss/.

## ADDITIONAL REFERENCES


BAGDONAVICIUS, V. B. and NIKULIN, M. S. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Anal.* **5** 329–350.

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability*. Holt, Rinehart & Winston, New York.

BLOCK, H. and JOE, H. (1997). Tale behavior of the failure rate functions of mixtures. *Lifetime Data Anal.* **3** 269–288.

BLOCK, H. W., MI, J. and SAVITS, T. H. (1993). Burn-in and mixed populations. *J. Appl. Probab*. **30** 692–702.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

FINKELSTEIN, M. S. and ESSAOULOVA, V. (2000). Modeling a failure rate for a mixture of distribution functions. Technical Report 279, Univ. Orange Free State.

GURLAND, J. and SETHURAMAN, J. (1994). Reversal of increasing failure rates when pooling failure data. *Technometrics* **36** 416–418.

GURLAND, J. and SETHURAMAN, J. (1995). How pooling failure data may reverse increasing failure rates. *J. Amer. Statist. Assoc*. **90** 1416–1423.

HOUGAARD, P., MYGLEGAARD, P. and BORCH-JOHNSEN, K. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics* **50** 1178–1188.

PROSCHAN, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* **5** 373–383.

ROBINS, J. and GREENLAND, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics* **45** 1125–1138.

ROBINS, J. and GREENLAND, S. (1991). Estimability and estimation of expected years of life lost due to a hazardous exposure. *Statistics in Medicine* **10** 79–93.

SCHEIKE, T. H., PETERSEN, J. H. and MARTINUSSEN, T. (1999). Retrospective ascertainment of recurrent events: an application to time to pregnancy. *J. Amer. Statist. Assoc.* **94** 713–725.

TETENS, J. N. (1786). *Einleitung zur Berechnung der Leibrenten und Anwarschaften II*. Weidmanns Erben und Reich, Leipzig.

VAUPEL, J. W., MANTON, K. G. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.

VAUPEL, T. W. and YASHIN, A. I. (1985). Heterogeneity's ruse: some surprising effects of selection on population dynamics. *Amer. Statist*. **39** 176–185.

WANG, J., MULLER, H. and CAPRA, W. B. (1998). Analysis of oldest-old mortality: lifetimes revisited. *Ann. Statist*. **26** 126–133.