BIOMETRICS 42, 693–734
December 1986

A BIOMETRICS INVITED PAPER WITH DISCUSSION

The Natural Variability of Vital Rates and Associated Statistics

David R. Brillinger
Statistics Department, University of California, Berkeley, California 94720, U.S.A.

SUMMARY

The first concern of this work is the development of approximations to the distributions of crude mortality rates, age-specific mortality rates, age-standardized rates, standardized mortality ratios, and the like for the case of a closed population or period study. It is found that assuming Poisson birthtimes and independent lifetimes implies that the number of deaths and the corresponding midyear population have a bivariate Poisson distribution. The Lexis diagram is seen to make direct use of the result. It is suggested that in a variety of cases, it will be satisfactory to approximate the distribution of the number of deaths given the population size, by a Poisson with mean proportional to the population size. It is further suggested that situations in which explanatory variables are present may be modelled via a doubly stochastic Poisson distribution for the number of deaths, with mean proportional to the population size and an exponential function of a linear combination of the explanatories. Such a model is fit to mortality data for Canadian females classified by age and year. A dynamic variant of the model is further fit to the time series of total female deaths alone by year. The models with extra-Poisson variation are found to lead to substantially improved fits.

1. Introduction

Vital statistics are data on the fundamental events of human lives—events such as birth, death, marriage, and the like. They usually take the form of counts or rates and are often collected via censuses and legally required registrations. They are used for summarization, comparison, forecasting, detection of change, hypothesis generation, surveillance, and studying public health generally.

A continuing presence is a wish to make comparisons—comparisons between regions, (comparisons between) time periods, (comparisons between) social groups, and so on. Now, in many circumstances the data are virtually complete so that it is a fact that death rates differ for two counties or two years or two races. What is more likely of concern then is: do two death rates differ by more than some level of natural fluctuations? The purpose of this study is the stochastic conceptualization and formalization of the natural variability of vital statistics to support their use in comparisons and other analyses. The particular cases of mortality and of groups specifically delineated in age and time will be emphasized; however, results for a broad variety of other cases should be apparent.

Quite a number of distinct vital statistics are in common use. These include counts, rates, and ratios, the emphasis being on the latter two.

Counts. One sets down the total number of deaths in a given time period for a population of interest, perhaps separately by age, region, or cause. Figure 1, middle, gives this data for the population of all Canadian females annually for the time period 1926–1982. Figure 2 gives the counts of deaths, again for Canadian females, but now for the years 1950–1972

Key words: Age-adjusted death rate; Dynamic model; Extra-Poisson variation; Point process; Poisson regression; Standardized mortality ratio; Uncertainty estimation; Vital statistics; Weights.

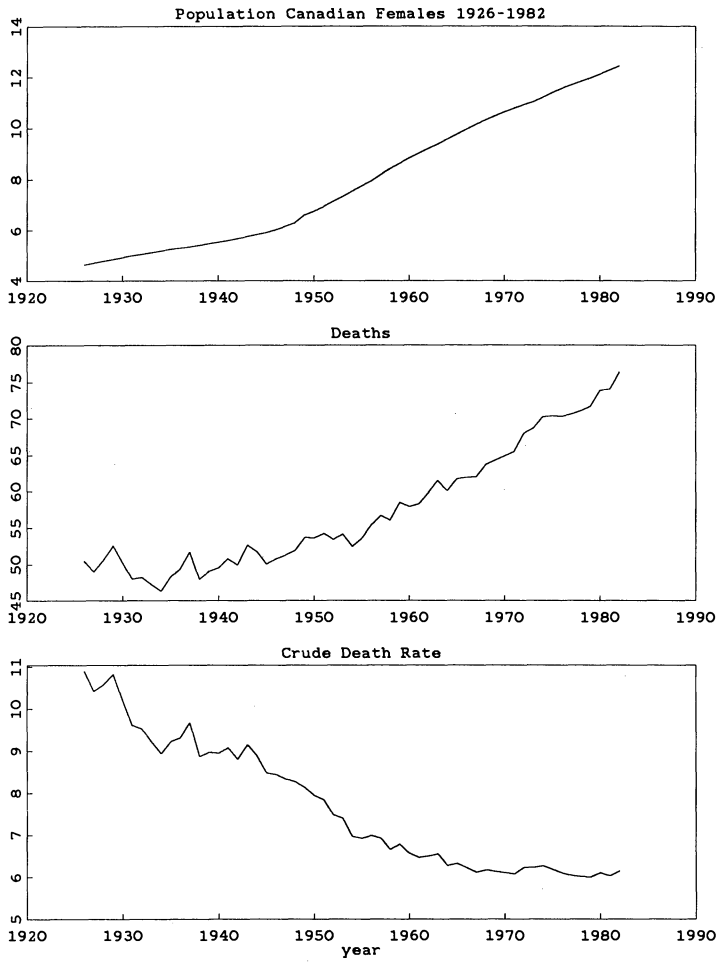


Figure 1. The top graph provides the estimated 1 June number of females in Canada for 1926–1982. The middle graph provides the year's total number of female deaths for the same time period. The bottom graph is the ratio of the previous two, the crude death rate.

and separately by age. The radii of the circles plotted here are proportional to the corresponding numbers of deaths. An issue that typically arises in mortality studies is whether the population whose deaths have been recorded is *closed* or *open.* In the latter case the group membership changes continually because of emigration, immigration, and the like. In the former, changes arise solely from births, deaths, and birthdays. The theorems of the paper are for closed populations. The data sets are for open populations.

Rates. Rates are relative frequencies. For example, the *crude death rate* is the number of deaths in a population of interest during a specified time period, divided by the number of person-years lived by the population during the time period. A complication that arises often is that person-years lived has to be estimated. In the case of annual rates, an estimate of the midyear population is often used. [This procedure "is (largely?) confined to English-speaking countries," remarks a referee.] Figure 1 gives the Canadian 1 June population estimate for all females for the period 1926–1982 and also the corresponding crude death rate. The three graphs of Figure 1 display rising numbers of deaths and population members, but a falling death rate. (The kink in the population series in 1949 resulted in part from

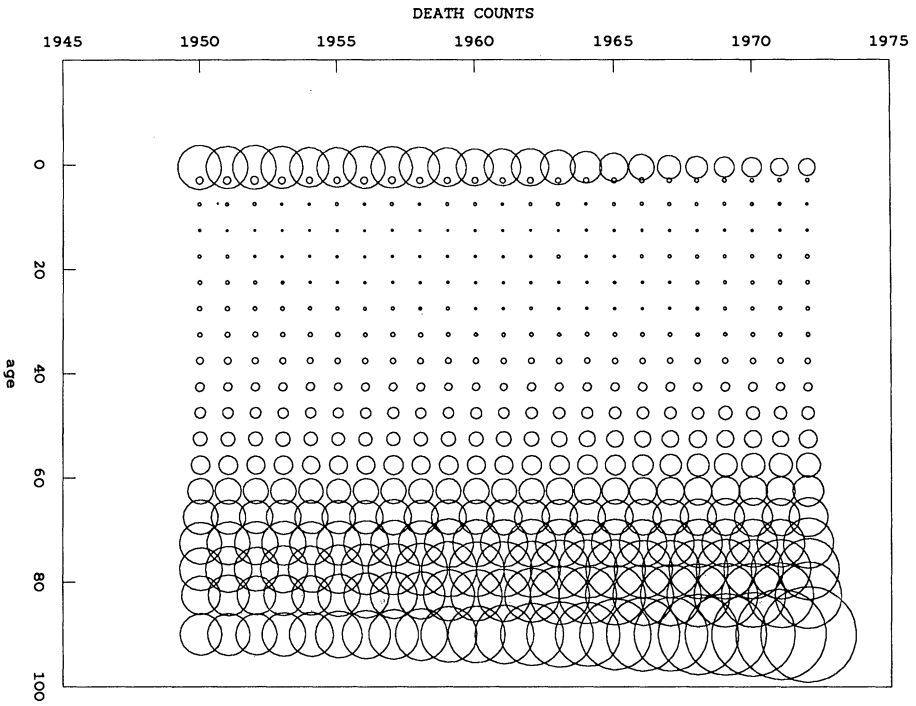


Figure 2. A circle diagram to represent the counts of females dying annually from 1950 to 1972 separated into 19 age groups (age intervals: 0–1, 1–4, 5–9, 10–14, . . . , 80–84, 85+). The radius of the circle plotted is proportional to the corresponding count (and thus an estimate of the count's variance in the Poisson case).

Newfoundland's entering Confederation. That naive use of the 1 June figure can lead to biased results, should be noted.) In the case that a death rate is given for a specified age group, it is referred to as an *age-specific mortality rate.* These rates are important because mortality experience usually varies substantially with age and a crude rate may not display important phenomena. Figure 3 gives the age-specific death rate for the Canadian females for the period 1950–1972; again the radii are proportional to the variate of concern. It evidences interesting trends in mortality. *Age-adjusted rates* are an attempt to provide single rates that allow direct comparison of populations with differing age compositions. They are weighted combinations of age-specific rates. (For example, the weights may correspond to the composition of some standard population.)

Ratios. The *standardized mortality ratio* (SMR) may be mentioned. It is the ratio of observed total deaths to "expected" deaths using the rates of some standard population and the given person-years lived. It is often used in making comparisons. A useful survey of the SMR is given in Breslow and Day (1985).

The purpose in setting down the above material has been to bring out the basic quantities involved in constructing vital statistics—counts and estimates of population size. These are the quantities whose variability will be fundamental. Discussion of vital statistics generally and details concerning particular cases may be found in Chiang (1961), Keyfitz (1966), Benjamin (1968), Fleiss (1981), Benjamin and Pollard (1980), Inskip, Beral, and Fraser (1983), and Hoem (1976, 1978, 1984a), for example.

The structure of this paper is the following. Section 2 sets the scene and presents some variability measures in common use; Section 3 sets down a conceptual model for the biological process of concern and shows how the Lexis diagram and the methodology of point processes may be used. Section 4 presents specific formulas for a number of cases of

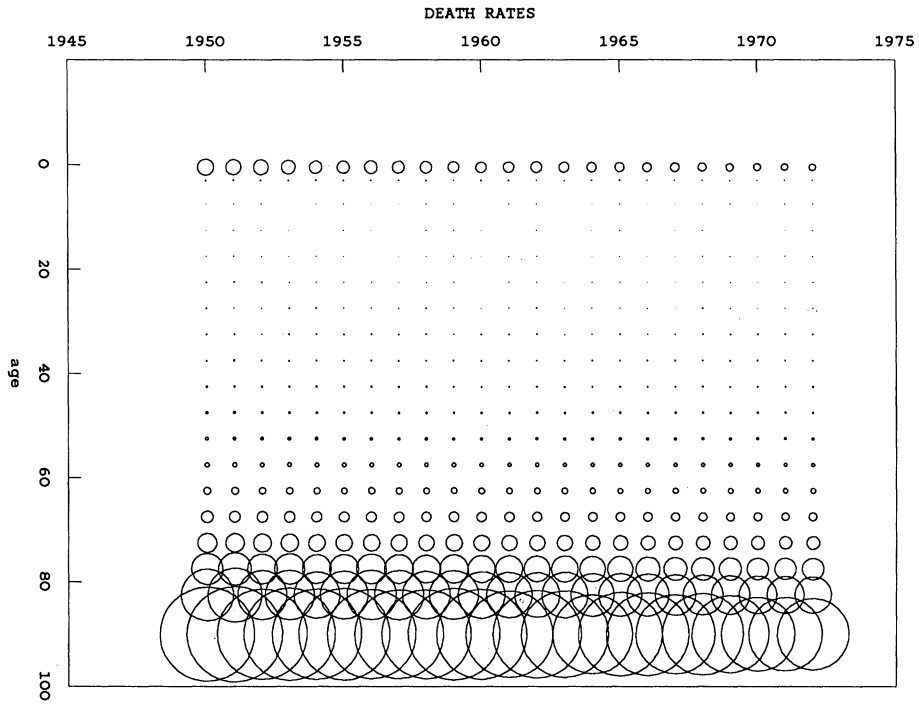


Figure 3. A circle diagram of the age-specific death rates similar to Figure 2.

interest; the following section discusses the results obtained and describes a simplifying approximation. Section 6 turns to the (regression) case where measurements of explanatory variables are available and indicates a model fitting procedure. Section 7 makes use of that fitting procedure for the two Canadian data sets mentioned above. The final section draws some conclusions and indicates some problems for further study.

2. Some Background

Discussion will focus on the case of an age-specific death rate for a given year. Let D_x denote the number of deaths in the year for the age group x to $x + K$, say, and let P_x denote the midyear population for that age group. Then the age x death rate is (usually taken to be)

$$M_x = D_x/P_x. \tag{2.1}$$

(In practice, P_x has to be estimated, but D_x may be obtained from official records. For the moment, however, P_x will be assumed available.)

In statistical studies, D_x is often assumed to be distributed as a binomial variate with parameter $n = P_x$ and its variance is estimated by $D_x(1 - M_x)$. [See, for example, Pollard (1970), Daw (1974), Mosteller and Tukey (1977, §11C).] Conceptually, however, this assumption has to be viewed as an approximation for our case of a closed population (as opposed to a cohort of individuals). Here some individuals enter the population during the year, when they reach age x, others leave during the year, when they reach age $x + K + 1$. The exposures of the individuals are not all the same and the realizations of the individual life histories are not identically distributed, as is required for the binomial. Further, P_x is not the number of individuals in the study; rather, it is an estimate of the average number alive aged x to $x + K$ during the year.

In research to obtain a more valid variance estimate, Chiang (1961) created a hypothetical cohort of size $N_x = [P_x + K(1 - a_x)D_x]/K$ and assumed D_x binomial with parameter $n = N_x$. (Here a_x is taken to represent the average fraction lived in the year by an individual who dies then. It is usually taken to be $\frac{1}{2}$.) His estimate of the variance of D_x is

$$D_x(1 - D_x/N_x). \tag{2.2}$$

Chiang develops his result by replacing the lives of those in the population aged, say, $x + 1, x + 2, \ldots$, by later stretches of life of those aged x at the beginning of the period. The value for N_x results from equating expressions for person-years lived. In the case that the death rate is low, both Chiang's and the preceding estimate are approximately D_x. The latter corresponds to Poisson variation. It will be argued in this paper that it is the Poisson estimate that should be employed generally, whether or not the death rate is low. Further, no artificial cohorts will be created in our analysis. It may be remarked that others have suggested the use of the Poisson; however, their arguments seem always to be accompanied by a remark that this follows from deaths being rare.

Death rates are often subjected to regression analyses when explanatory variables are available. The discussion of what is an appropriate variance estimate there becomes the question of what weights to employ in the regression analysis. Some references are Fryer et al. (1979), Hogan et al. (1979), and Pocock, Cook, and Beresford (1981). This issue will be returned to later in the paper.

3. A Conceptual Model

An issue that arises with death counts and rates is: are these not facts (that is, exact values), and hence not subject to uncertainty? There are, however, various conceptual bases for treating birthtimes and lifetimes as random. Among justifications that may be provided are: moments of birth and death appear unpredictable; there exists an immense biological variability; there exists substantial environmental diversity; there are epidemics; there are medical advances, accidents, violent deaths; periods of extreme weather occur; and finally, researchers have constructed useful chance mechanisms for fertility and disease. At the same time it may be mentioned that there do exist some near-deterministic aspects; in particular, babies may be induced or born via Caesareans and there exist seasonal fluctuations. (This last issue will be returned to later.) In the framework to be presented, both times of birth and lifetimes will be assumed stochastic, thence leading to a natural variability in vital statistics. We mention that other philosophical attitudes have been adopted regarding the issue of stochastic models. For example, Keyfitz (1966) suggests that "a census may be regarded as a sample drawn in time from all the times in which substantially the same conditions prevailed."

Before sampling results are developed, some notation and assumptions will be set down. It will be convenient to display individuals' life histories by slope 1 lines in the age vs time-of-death plane, i.e., to employ Lexis diagrams. [This technique is discussed and employed in Benjamin and Pollard (1980), for example.] In the diagram the axes have matched scales; lifelines begin at birth and end at death. For a set A of the plane, the number of deaths in A is given by the number of lines ending in A. Figure 4 provides an example of a Lexis diagram with the lifelines indicated. Suppose one is interested, for example, in the age 40–44 death rate of 1980. It is given by $N(B)/N(C)$, with B and C the regions indicated in Figure 5.

In setting down definitions and developing results, it will be convenient to make use of the mathematical machinery of stochastic point processes. Cox and Lewis (1966) is one reference to this material. Briefly, a *linear point process* is a random scattering of points along the real line. Its realizations may be denoted by $\{\sigma_j\}$ with σ_j the coordinate of the jth

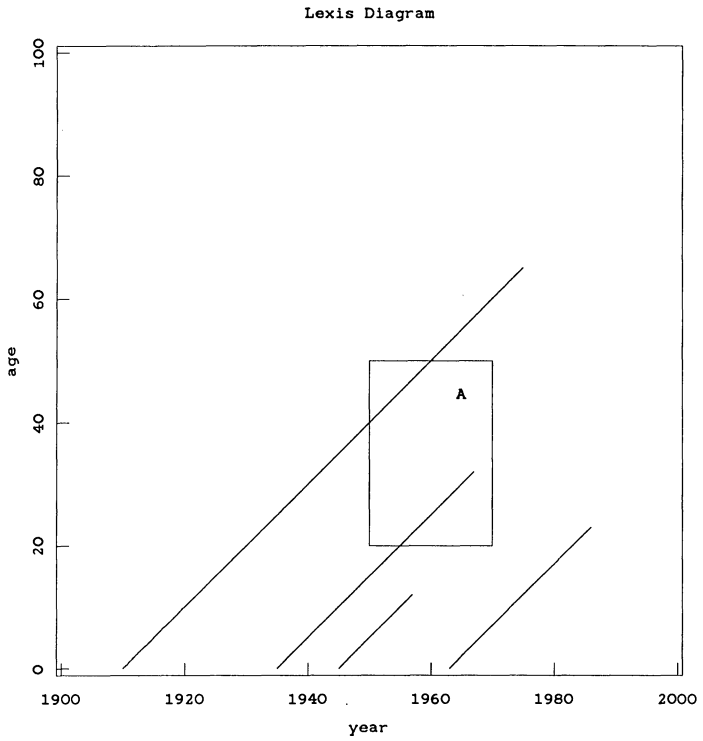


Figure 4. A Lexis diagram, with the sloping lines representing individuals' lifetimes. The lines begin at the moment of birth and end at death. Those ending in region A represent individuals dying in the corresponding age and time intervals.

point. (Shortly the σ_j will be taken to be the birthtimes of the population members.) An important parameter of a linear point process is its *intensity* or *rate function*, $\beta(\cdot)$, given by

$$\Pr\{\text{Point in } (t, t + h)\} \approx \beta(t)h \tag{3.1}$$

for h small. Supposing I to be an interval, and $M(I)$ to be the number of points in I, then

$$\mathrm{E}\{M(I)\} = \int_I \beta(t)\, dt = B(I), \tag{3.2}$$

say. The (linear) Poisson process with intensity $\beta(\cdot)$ may now be defined by the requirement that for disjoint intervals $I_1, \ldots, I_K$, the counts $M(I_1), \ldots, M(I_K)$ are independent Poisson variates with means $B(I_1), \ldots, B(I_K)$, respectively, $K = 1, 2, \ldots$.

A *planar point process* is a random scattering of points in the plane. Its intensity function $\lambda(\cdot)$ is given by

$$\Pr\{\text{Point in } (t, t + h) \cdot (x, x + \bar{h})\} \approx \lambda(t, x)h\bar{h} \tag{3.3}$$

for h, $\bar{h}$ small. If A is a region of the plane and $N(A)$ the number of points in A, then

$$\mathrm{E}\{N(A)\} = \int\int_A \lambda(t, x)\, dt\, dx = \Lambda(A) \tag{3.4}$$

here.

The *planar Poisson process* with intensity $\lambda(\cdot)$ is given by the requirement that for

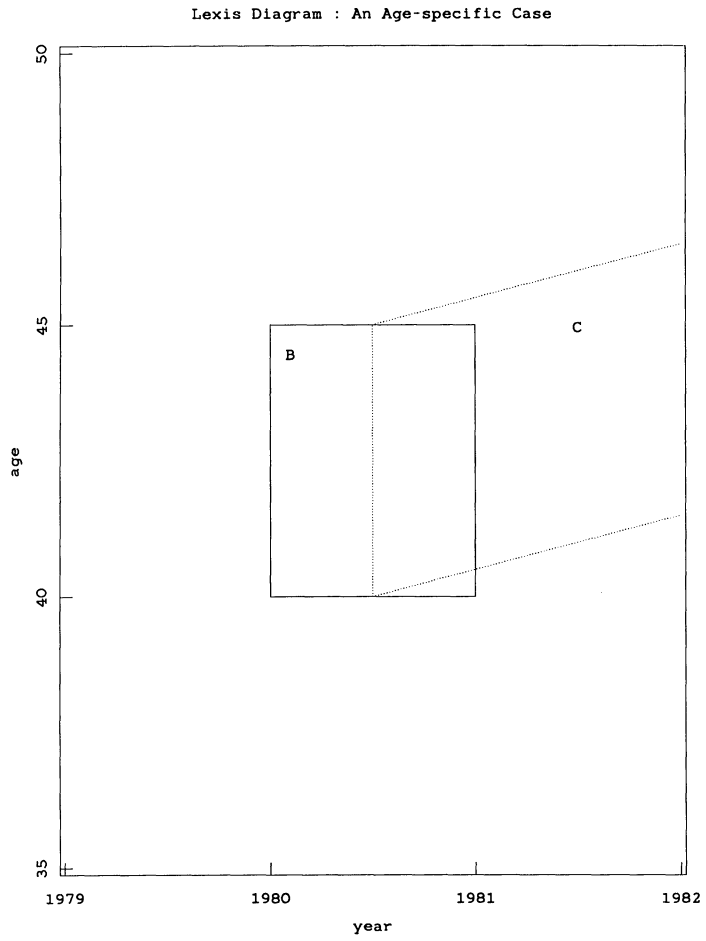


Figure 5. A Lexis diagram representing the death rate, $N(B)/N(C)$, for the age group 40–44 and the year 1980, taking $N(C)$ to be the midyear population.

disjoint regions $A_1, \ldots, A_K$, the counts $N(A_1), \ldots, N(A_K)$ are independent Poisson variates with means $\Lambda(A_1), \ldots, \Lambda(A_K)$, respectively, $K = 1, 2, \ldots$. Here $\Lambda(A)$ is given by expression (3.4).

Properties of this Poisson process include: (i) $\text{var}\{N(A)\} = \text{E}\{N(A)\} = \Lambda(A)$; and (ii) for A contained in a region B, the distribution of $N(A)$ given $N(B) = n$ is binomial with parameters n and $\Lambda(A)/\Lambda(B)$.

Returning to the discussion of vital statistics, suppose that the times of birth of the members of the population of concern are $\sigma_1, \sigma_2, \ldots$. Let $M(I)$ denote the number of σ_j in the interval I. Supposing that $M(I)$ is a stochastic point process, its intensity function $\beta(t)$ will be referred to as the *birth intensity.* (We remark that, for example, $\beta(t)$ would be periodic were there a weekly effect present.) Next, suppose that individuals live random lengths of time. Let X denote the lifetime of an individual born at time σ_0. The distribution of X is conveniently described by the *force of mortality,* $\mu(t, x)$, defined by

$$\Pr\{x < X \leqslant x + h \mid X \geqslant x\} \approx \mu(t, x)h \tag{3.5}$$

with $t = \sigma_0 + x$ and h small. For example, the probability that an individual born at time

σ_0 survives to age x is given by

$$\exp\left\{-\int_0^x \mu(\sigma_0 + y, y)\, dy\right\}. \tag{3.6}$$

The *death process* is defined to be the planar point process with points at the positions (date of death, age at death); specifically, supposing the jth individual is born at time σ_j and dies aged x_j, then $N(A)$ denotes the number of points $(\sigma_j + x_j, x_j)$ in the region A. It corresponds to the endpoints of the lifelines of the Lexis diagram. Let $\lambda(t, x)$ denote the *death intensity* [i.e., the intensity of the point process $N(\cdot)$]; then $\Lambda(A)$ is given by expression (3.4) with

$$\lambda(t, x) = \beta(t - x)\mu(t, x)\exp\left\{-\int_0^x \mu(t - x + y, y)\, dy\right\}. \tag{3.7}$$

This last follows from first principles; the three factors on the right have the respective interpretations: "born at $t - x$," "die at t age x," "survive to t," assuming appropriate statistical independence.

A theorem describing the distribution of the death process may now be stated. A proof is provided in the Appendix.

Theorem 1. If (a) the birth process $\{\sigma_j\}$ is Poisson with intensity $\beta(t)$, (b) the lifetimes $\{X_j\}$ of the individuals are independent of each other, independent of the birth process, and correspond to the force of mortality $\mu(t)$, then the death process $N(\cdot)$ is planar Poisson with intensity (3.7).

This result may be used to derive the distributions of various vital statistics. It is particularly convenient because for the Poisson process, counts corresponding to disjoint regions are statistically independent. It is worth remarking specifically that the resulting Poisson distributions for death counts arise not from rarity (small numbers), but rather from the assumed total randomness (Poisson) of the birth process and the assumed randomness of lifetimes.

The assumptions of Poisson births and independent lifetimes were essential to the derivation of the Poisson result. In fact, the birth process may be expected to show some clustering because of twin births. Further, lifelines will not be completely independent because of the existence of multiple deaths (in accidents, for example). However, these particular phenomena may be expected generally to have small effects. There have been a number of empirical studies of time series of birth data. In particular, we mention Cohen (1983), containing a variety of references, and Izenman and Zabell (1976). A prominent phenomenon brought out in the references is the weekly period present in a variety of data sets. As mentioned above, inductions and Caesarians are obvious explanations of the period. Izenman and Zabell (1976) also carried out a spectral analysis of their data (births in New York City—a very open population!). Their estimated power spectrum shows peaks corresponding to a weekly period superposed on a fairly flat spectrum. This is what could be expected were the process periodic Poisson. Other evidence consistent with a Poisson was found in unpublished work by LeRoy (course project, Statistics Department, University of California, Berkeley, 1983). She studied the births for 512 consecutive days (a total of 1777 births) at Highland Hospital. This is the county hospital for Alameda County and there is little intervention in the birth process. No pronounced weekly period was found and spectrum analysis results were consistent with a pure noise process.

The Poisson process results from various limiting operations. These include rare events, random thinning, superposition, and large translations [see, for example, Daley and Vere-Jones (1972)]. One can argue here that many birth processes will result from the superposition of many separately evolving sub-birth processes, for example, and thereby motivate a Poisson assumption. However, ultimately any justification of the Poisson has to be

empirical. Some other works presenting stochastic formulations of the population process include Jagers (1974, 1975), Keiding and Hoem (1976), Braun (1978), Brillinger (1981), and Jagers and Nerman (1984). One of the referees remarked on the possibility of developing a full model for the population process of concern, taking note that births are to members of the population and so on. An issue was whether or not a Poisson process would/could result for the birth process. It may be noted that Kendall (1948) developed a result showing that this is not the case for a pure birth-and-death process. In Section 4 of his paper, Kendall developed the distribution of the total number of births. It was not Poisson. If a Poisson is to result from axioms, it seems it must come from many processes being superposed, from events being rare, or some such. Since some of the births in most human populations are typically occurring to members of the population and lead to later births into the population, it cannot be expected that the birth process will ever be exactly Poisson. It appears sensible and topical for someone to follow up the referee's suggestion and to build an econometric-type model of an evolving population. A situation in which the Poisson birth assumption is clearly inappropriate occurs with manpower studies. Here there is a steady (deterministic) recruitment to the population so the results derived will most likely be inappropriate. Batholomew and Forbes (1979) is one reference concerned with statistical aspects of manpower.

The following result will be used in the next section to set down the distributions of various statistics of interest. It follows by writing the regions involved in terms of disjoint subregions and from the fact that for a Poisson process, counts for disjoint regions are statistically independent.

Corollary. Under the conditions of the theorem, for any regions in the Lexis diagram: (a) $\{N(B), N(C)\}$ is distributed as $\{U + W, V + W\}$, where U, V, W are independent Poissons with means $\Lambda(B - BC)$, $\Lambda(C - BC)$, $\Lambda(BC)$; (b) $N(B)/N(C)$ is distributed as $(U + W)/(V + W)$. Further, (c) $N(B)$ given $N(C)$ is distributed as $U + S$, where U is Poisson with mean $\Lambda(B - BC)$ and S is independently binomial with mean $n = N(C)$ and proportion $\Lambda(BC)/\Lambda(C)$.

4. Some Examples

The preceding theorem and corollary will now be used to set down distributions for various vital statistics.

Example 1. *Crude death rate* Let D denote the number of deaths in a given year and P the corresponding midyear population. Then the crude death rate is D/P. It may be represented by $N(B)/N(C)$ with B, C regions of the Lexis diagram, analogous to those of Figure 4, but for the whole age range.

Assuming the complete randomness of births and independent lifetimes as required in the theorem, it follows that $\{D, P\}$ has a bivariate Poisson distribution. Specifically, $\{D, P\}$ is distributed as $\{U + W, V + W\}$ of the theorem, with $\Lambda(\cdot)$ given by (3.4), (3.7). The crude death rate D/P, is therefore distributed as $(U + W)/(V + W)$. (Incidentally, this representation shows that there is a chance that the denominator of this ratio may be 0 when the numerator is not. This happens when all die in the first half of the year.) The bivariate Poisson is discussed in Haight (1967).

On some occasions one is interested in conditional distributions. It follows from the corollary that the distribution of D given P is that of $U + S$ with U Poisson and S independently binomial. In particular, this gives

$$\mathrm{E}\{D \mid P\} = \Lambda(B - BC) + P\frac{\Lambda(BC)}{\Lambda(C)}, \tag{4.1}$$

$$\mathrm{var}\{D \mid P\} = \Lambda(B - BC) + P\frac{\Lambda(BC)}{\Lambda(C)}\left[1 - \frac{\Lambda(BC)}{\Lambda(C)}\right]. \tag{4.2}$$

Restating (4.1), the regression coefficient of D on P is $\Lambda(BC)/\Lambda(C)$. This is initially surprising because the region BC refers only to deaths occurring in the second half of the year to persons born before 1 July of the year. The constant term provides specific information in this case of concern—a closed population. The relations (4.1) and (4.2) may be used to guide regression analyses. The conditional variance of D, and indeed its distribution, are made up of a Poisson and a binomial part. In the case that the expected number of deaths in $B - BC$ is small, the distribution is approximately binomial and one is led to the traditional assumption of binomial variation. In the case that B is contained in C, $B - BC$ is empty and the binomial is exact.

Example 2. *Age-specific death rates* The age x to $x + 4$ death rate for 1980 has the form $M_x = D_x/P_x = N(B_x)/N(C_x)$, where B_x is the set (1980, 1981) $\cdot$ $(x, x + 5)$ and C_x is the set of (t, y) satisfying $1980.5 < t$ and $x \leqslant y - (t - 1980.5) < x + 5$. (See Fig. 5.) $N(B_x)$ and $N(C_x)$ count, respectively, how many die aged x to $x + 5$ in 1980 and how many were alive and aged x to $x + 5$ on 1 July 1980. Because of the (planar) Poisson nature of the death process, a variety of distributions are now apparent here from the theorem and its corollary. The distribution of D_x is Poisson with mean

$$\Lambda(B_x) = \int_{1980}^{1981} \int_{x}^{x+5} \lambda(t, y)\, dt\, dy. \tag{4.3}$$

Its variance may be estimated by D_x. The distribution of $\{D_x, P_x\}$ is bivariate Poisson. The distribution of D_x given P_x is not generally simple. An approximation to the distribution will be presented in the next section.

One simple result is that M_x statistics for disjoint age intervals are statistically independent.

Example 3. *Age-standardized rates* These have the form

$$\sum_x w_x N(B_x)/N(C_x) \tag{4.4}$$

for given weights w_x [see, for example, Chiang (1961)]. The distribution may be described in terms of Poisson variates. It is generally nonelementary.

Example 4. *Ratios* These are generally based directly on counts. For example, the standardized mortality ratio (SMR) is given by

$$N(B) \Big/ \sum_x M_{xs} N(C_x), \tag{4.5}$$

with the M_{xs} the rates of a selected standard population. The distribution here is nonelementary, but it may be represented directly in terms of statistically independent Poisson variates. An approximation to its variance will be suggested in the next section.

5. Some Discussion and an Approximation

The principal purpose of this paper has been to provide a conceptual basis on which sampling uncertainties of various vital statistics might be derived. Assuming birthtimes in accordance with a Poisson process, and assuming independent lifetimes, it has been found that the points (time of death, age at death) are distributed in the Lexis diagram in accordance with a planar Poisson process. This means, for example, that counts corresponding to disjoint regions of the Lexis diagram are independent Poissons. As many vital statistics may be written as functions of such counts, a representation for their distribution has therefore been constructed. The results are found to differ from those of Chiang (1961);

surprisingly, the results here are simpler. Chiang's results typically involve Poisson terms *and* correction terms, such as the $(1 - D_x/N_x)$ of expression (2.2). One implication of the difference is that variances estimated under the present framework will generally be larger. An extreme case of this is provided by that of the rate for those aged 85 and over. Chiang (1961, p. 281) suggested estimating this variance by 0. Here it would be estimated by D_{85}/M_{85}^2 (see below).

The exact distribution of an age-specific death rate has been found to involve the bivariate Poisson. This is generally an inconvenient distribution to work with. In the case that the coefficient of variation of the population size variable is small, as the following theorem shows, an elementary approximation may sometimes be employed usefully. A further advantage occurring here is that the particular choice made for the denominator (person-years lived) becomes not so crucial. Briefly, the approximation is to replace the denominator by its expected value.

Theorem 2. Suppose that D is Poisson with mean λ and that P has mean Λ and variance σ^2; then

$$\left|\Pr\left\{\left(\frac{D}{P} - \frac{\lambda}{\Lambda}\right)\Big/\frac{\sqrt{\lambda}}{\Lambda} \leqslant x\right\} - \Pr\left\{\left(\frac{D}{\Lambda} - \frac{\lambda}{\Lambda}\right)\Big/\frac{\sqrt{\lambda}}{\Lambda} \leqslant x\right\}\right| \leqslant 3\left[\frac{\sigma}{\Lambda}(\sqrt{\lambda} + x)\right]^{2/3} + \frac{1}{\sqrt{\lambda}} \tag{5.1}$$

for $\sqrt{\lambda} + x > 0$.

The proof of this result is given in the Appendix. The essence is that P may be replaced by its expected value when its coefficient of variation, σ/Λ, is small.

This result, via a first substitution of Λ for P, then one of P for Λ, leads one, for example, to estimate the variance of the crude death rate D/P by D/P^2; to estimate the variance of an age-specific death rate $M_x = D_x/P_x$ by D_x/P_x^2; and to estimate the variance of an age-standardized rate $\sum w_x M_x$ by $\sum w_x^2 D_x/P_x^2$. This last is to be contrasted with expression (18) in Chiang (1961)—namely, $\sum w_x^2 D_x(1 - D_x/N_x)/P_x^2$. For the SMR $= D/\sum M_{xs}P_x$, the estimate of variance obtained is $D/(\sum M_{xs}P_x)^2 = \text{SMR}/(\text{expected deaths})$.

In some situations one may have a parametric model of interest. One may then be able to set down a likelihood function and proceed to compute, say, maximum likelihood estimates. In particular cases, that likelihood may factor in a pertinent fashion, leading one to make inferences conditionally. This happens, for example, in the case of a cohort. Another situation in which things simplify is when individuals' person-years lived values are known. Hoem (1984b) discusses this case and presents variance estimates.

Theorem 2 provides a result of use in developing approximations to desired results. Rosenblatt et al. (1983) develop a uniform approximation to the distribution of a Poisson variate by a normal. This approximation seems likely to prove of use in the present situation.

6. Regression

In many studies of mortality, measured explanatory variables are available. The most common of these are age and (time) period. Others include race, sex, and region. An individual's mortality may be expected to depend on various of these. The measurements may be included, in a quantitative manner, by setting down a functional form for the force of mortality or a related parameter. In this section, the case of Poisson regression will first be mentioned; then the case of extra-Poisson variation will be studied.

6.1 *Poisson Regression*

The stochastic model of mortality, presented earlier in this paper, led to a Poisson distribution for the number of deaths. In the case that the population size is P and that a (possibly vector-valued) explanatory variable $\mathbf{x}$ is available, one might assume that the number of deaths, D, is Poisson with mean $P \exp\{\mathbf{x}'\boldsymbol{\beta}\}$, $\boldsymbol{\beta}$ being a (possibly vector-valued) parameter to be estimated. For example, Frome (1983) sets down such a model for lung cancer deaths of British physicians, taking P to be man-years at risk, and years of smoking and number of cigarettes per day as explanatories. The Poisson model is found to fit well, a deviance of 51.47 based on 48 degrees of freedom obtained. [See also Frome and Checkoway (1985).]

6.2 *Extra-Poisson Variation*

As is often the case in ordinary regression analyses, it is to be expected that in many situations essential explanatory variables will not have been observed. Were they all available, a rate $\exp\{\mathbf{x}'\boldsymbol{\beta}\}$ might be appropriate. In the case of omitted explanatories, we are led to consider a rate

$$\exp\{\mathbf{x}'\boldsymbol{\beta} + u\}, \tag{6.1}$$

with u normal, mean 0 and variance σ^2. The parameter σ^2 provides a measure of the extra-Poisson variation. Following the work of Bock and Aitkin (1981), Hinde (1982), and Brillinger and Preisler (1983), the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ here may be determined by a combination of numerical integration and the EM algorithm. That the error u has been assumed normal is not crucial to the technique; rather, that the distribution of the error should be "known" up to a finite-dimensional parameter is. Of course, the estimates will depend critically on the assumed distribution for u.

Briefly, the approach is as follows. Let U denote a latent variate with density function $f(u \mid \gamma)$ depending on a parameter γ. (In the present case, U is u and γ is σ.) Let Y be an observable variate with probability mass (or density) function, given $U = u$, $f(y \mid u, \beta)$ depending on the parameter β. (In the present case Y is Poisson and β is $\boldsymbol{\beta}$.) Then the marginal probability mass function of Y is

$$f(y \mid \beta, \gamma) = \int f(y \mid u, \beta) f(u \mid \gamma)\, du. \tag{6.2}$$

Let $\boldsymbol{\theta} = (\beta, \gamma)$ and

$$\psi(y \mid \boldsymbol{\theta}) = \frac{\partial \log f(y \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{6.3}$$

Supposing that observations $y_1, \ldots, y_n$ are available, the maximum likelihood equation for estimating $\boldsymbol{\theta}$ is given by

$$\sum_{i=1}^{n} \psi(y_i \mid \hat{\boldsymbol{\theta}}) = 0. \tag{6.4}$$

Elementary manipulations allow this last to be written

$$\sum_{i=1}^{n} \int \frac{\psi(y_i \mid u, \hat{\beta}) f(y_i \mid u, \hat{\beta}) f(u \mid \hat{\gamma})}{f(y_i \mid \hat{\beta}, \hat{\gamma})}\, du = 0, \tag{6.5}$$

$$\sum_{i=1}^{n} \int \frac{\psi(u \mid \hat{\gamma}) f(y_i \mid u, \hat{\beta}) f(u \mid \hat{\gamma})}{f(y_i \mid \hat{\beta}, \hat{\gamma})}\, du = 0. \tag{6.6}$$

The difficulty that arises in many cases, particularly the present one, is that the integrations in (6.5) and (6.6) may not be carried out analytically. An effective approach, however, is

to carry out the integrations numerically, replacing the probability element $f(u \mid \gamma)\, du$ by a discrete approximation

$$f(u \mid \gamma)\, du = \sum_{m=1}^{M} p_m \delta\{u - u_m\}, \tag{6.7}$$

$\delta\{u\}$ denoting the unit mass at $u = 0$. The u_m are nodes and the p_m are corresponding weights. This all leads to the approximate likelihood equations

$$\sum_{i=1}^{n} \sum_{m=1}^{M} \psi(y_i \mid u_m, \hat{\beta}) w_m(y_i \mid \hat{\beta}, \hat{\gamma}) = 0, \tag{6.8}$$

$$\sum_{i=1}^{n} \sum_{m=1}^{M} \psi(u_m \mid \hat{\gamma}) w_m(y_i \mid \hat{\beta}, \hat{\gamma}) = 0, \tag{6.9}$$

where the w_m are weight functions given by

$$w_m(y \mid \beta, \gamma) = f(y \mid u_m, \beta) p_m \Big/ \sum_{k=1}^{m} f(y \mid u_k, \beta) p_k. \tag{6.10}$$

These equations are conveniently solved iteratively or via GLIM directives [see Hinde (1982) and Brillinger and Preisler (1983) for these last].

Other papers addressing the issue of extra-variation include Vaupel, Manton, and Stallard (1979), Manton, Woodbury, and Stallard (1981), Hougaard (1984), Clayton and Cuzick (1985), and Yashin, Manton, and Vaupel (1985). The regression approach is studied, for cohorts, in Breslow et al. (1983). There are a number of papers devoted to the use of mixture models for lifetime data. We mention Farewell (1982), Heckman and Singer (1982, 1984), and Greenhouse and Wolfe (1984).

7. Two Examples

The modelling and fitting procedure of the previous section will now be illustrated by two sets of computations. One set involves the fitting of a dynamic (time series) model to the historical data on Canadian female mortality given in Figure 1. However, the first set discussed refers to the data of Figures 2 and 3, where both age and period are explanatories. The computations make use of Gauss–Hermite integration with 11 nodes [see Davis and Rabinowitz (1975) for the weights and nodes]. The standard errors were estimated as in Brillinger and Preisler (1983).

Examination of Figures 2 and 3 shows high death counts and rates for the 0–1 age group, with both counts and rates then falling as time passes. The figures show death counts at the higher ages increasing (as the population size increases with time), but the death rates themselves are falling.

Let D_{ij} denote the number of deaths in age group i for year j and let P_{ij} denote the corresponding (midyear) population. The model fit is one of D_{ij} given u_{ij} being Poisson of rate

$$P_{ij} \exp\{\beta_{i\cdot} + \beta_{\cdot j} + u_{ij}\}, \tag{7.1}$$

with the u_{ij} independent normals of mean 0 and variance σ^2. The $\beta_{i\cdot}$ and $\beta_{\cdot j}$ are age and period effects, respectively. On the basis of Theorem 2, the fitting is carried through as if P_{ij} is constant. The deviance obtained for a pure Poisson fit ($\sigma^2 = 0$) was 4443 on 396 degrees of freedom. The deviance with the extra-Poisson variance was 1429 on 395 degrees of freedom—a substantial reduction for the inclusion of a single further parameter. The estimate of σ was .072. It is to be expected that the deviance may be driven down substantially further by including other explanatories, such as province or other distributions

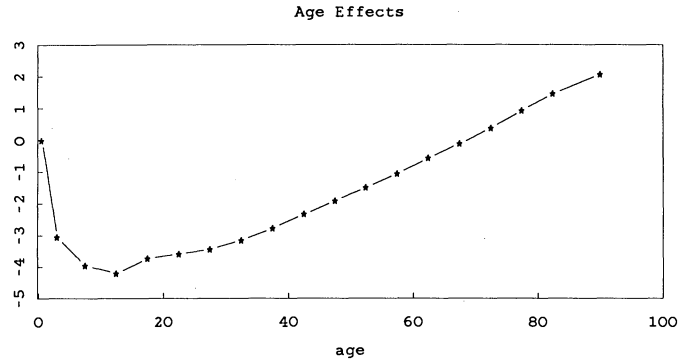


Figure 6. The estimated age effects for the data of Figure 2 and the model (7.1). The values are plotted against the average age of the corresponding age interval and the value for age 0 is set to 0.

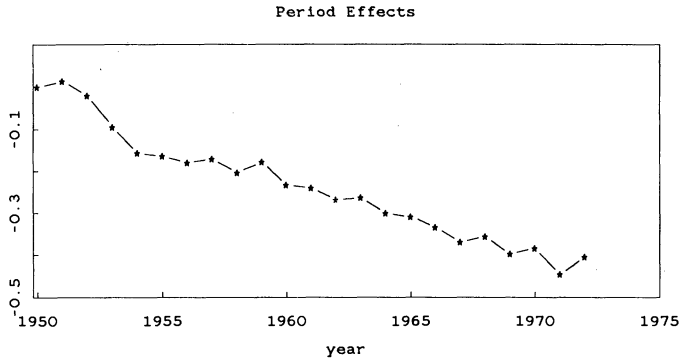


Figure 7. The estimated period (year) effects for the data of Figure 2. The 1950 value is set to 0.

for u_{ij}; however, the principal purpose of the present study is to illustrate that mortality data can be non-Poisson and that a direct procedure is available to handle the extra-variation, rather than to provide a "complete" analysis of these data sets. Figures 6 and 7 provide the estimated age and period effects, $\beta_{i\cdot}$ and $\beta_{\cdot j}$. (Actually, the model was re-parametrized to $\mu + \beta_{i\cdot} + \beta_{\cdot j}$, with $\beta_{1\cdot}, \beta_{\cdot 1} = 0$ to avoid aliasing.) The age effects show a "bathtub" shape—corresponding to high infant mortality, then a drop followed by a steady increase with age. There is a "bump" around age 17.5. This phenomenon has commonly been associated with high accident rates for this particular age group. In contrast, the period effects evidence an essentially steady decrease of mortality with time. It seems inappropriate to speculate on causes for its changes of direction just now.

The fit of a model is often conveniently studied by standardized residuals as well as the deviance. These are defined as $(D - m)/s$, where m and s are estimates of $\mathrm{E}\{D\}$ and $\sqrt{\mathrm{var}(D)}$, respectively, under the model considered. Figure 8 is an estimate of the density of the standardized residuals under the simple Poisson model. The distribution is exceedingly broad. Figure 9 is the estimated density for the model (7.1). This last figure provides further evidence of substantial improvement in fit resulting from employing the model with extra-Poisson variation.

When the computations were carried out with the original data, the residual plots brought

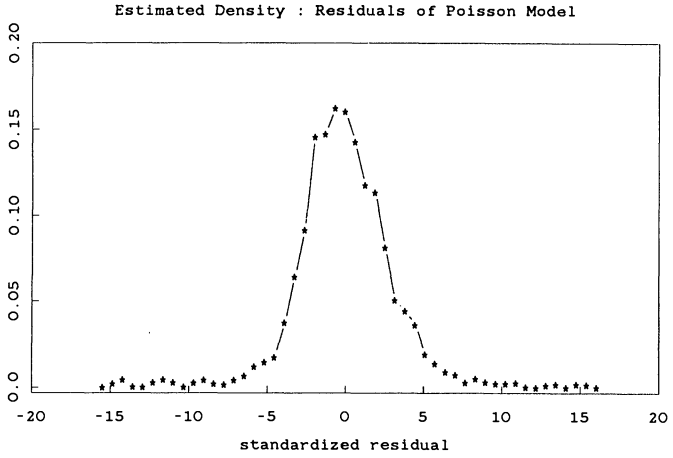


Figure 8. An estimate of the density function of the standardized residuals resulting from fitting a simple ($\sigma = 0$) Poisson model to the data of Figure 2.

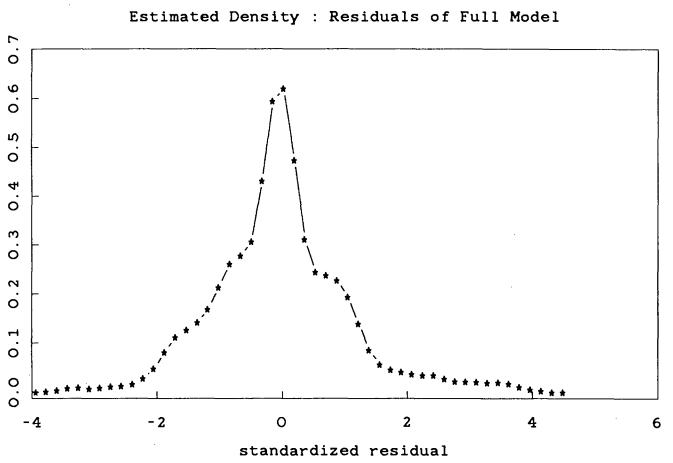


Figure 9. An estimate of the density function of the standardized residuals resulting from fitting the model (7.1) to the data of Figure 2.

out a clear outlier in the published values (Statistics Canada, 1976, Table 6)—namely, the 2.9 rate per thousand for those aged 35–39 in 1951. When Statistics Canada did a search in response to a request, they found that the value should have been 1.9 in fact. The results presented here are based on the corrected data set.

It is perhaps worth remarking that when the model with extra-Poisson variation was fit to the Frome data mentioned in Section 6, there was no real reduction in the deviance; it just fluctuated about with round-off error.

Our second example involves a time series modelling of the data on all Canadian female deaths during the time period 1926–1982, presented earlier in Figure 1 and taken from Table 1 of Statistics Canada (1976) and a supplement provided by D. Nagnur. The model fit is analogous to that for an autoregressive time series of order 1. Let d_t denote the number of deaths in time period (year) t and let p_t denote a corresponding measure of population

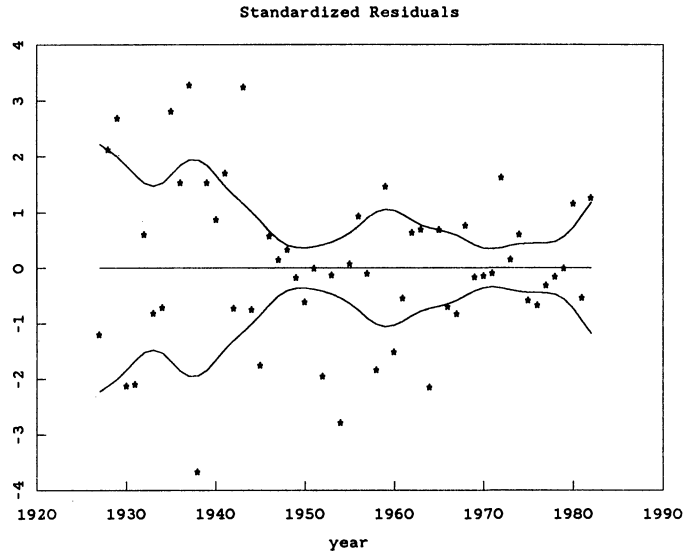


Figure 10. The standardized residuals, plotted versus year, resulting from fitting the dynamic model (7.2) to the total counts of Canadian female deaths during 1926–1982, Figure 1. Also plotted are heavily smoothed versions of the absolute values of the residuals.

size. Let u_t denote the death rate at time t and suppose that it evolves in accordance with

$$\log u_t = \gamma_0 + \gamma_1 \log u_{t-1} + \epsilon_t, \tag{7.2}$$

the ϵ_t being independent normal variates with mean 0, variance σ^2. Suppose further that given u_t and the past, d_t is distributed as Poisson with mean $p_t u_t$. A model of this sort may be expected to be of some use in forecasting.

The model (7.2) was fit by maximum likelihood as in the earlier example. The fitting was done as if the p_t were constant. The deviance for a Poisson model ($\sigma^2 = 0$) was 1844 based on 54 degrees of freedom. For the dynamic model ($\sigma^2 \neq 0$) it was 276.6 based on 53 degrees of freedom—a substantial improvement of fit for 1 degree of freedom. The estimates (and their estimated standard errors) are $\hat{\gamma}_0 = .062$ (.007), $\hat{\gamma}_1 = .964$ (.003), $\hat{\sigma} = .0168$ (.0005). The first two were highly correlated. Figure 10 is a plot of the (conditionally) standardized residuals versus time. It evidences a suggestion of variability reducing with time. The solid curves are the results of heavily smoothing the absolute values of the standardized residuals and are meant as guides for examining the points plotted.

8. Concluding Remarks

The goals of this paper have been to provide a conceptual basis for the description of the natural variability of certain vital statistics and to make use of that description in the analysis of two data sets. It was found that under two elementary assumptions (one pertaining to the birth process, the other to lifetimes) that basic counts of deaths were Poisson, with those counts corresponding to disjoint regions of the Lexis diagram independent. It was further demonstrated that sometimes, perhaps because of omitted explanatory variables, Poisson variability was insufficient. A general model involving extra-variability was set down and fit the two data sets. These data sets were found to evidence substantial variability beyond the Poisson.

A continuing issue in analyses of mortality rates, with measured explanatory variables, by linear regression has been: what are the appropriate weights for the observations?

Different choices are made in Fryer et al. (1979), Hogan et al. (1979), and Pocock et al. (1981), for example. Employing a full likelihood analysis, as is proposed in this paper, is clearly an alternate way to address the issue. Noting that the present computations were in fact carried out by iteratively reweighted least squares makes the connection even more apparent.

Next, it is to be noted that this paper has taken the basic quantities to be analyzed to be simple counts and rates. Clearly other quantities, perhaps specific estimates of probabilities as in Hoem (1984b) or subtle variants such as the Mosteller (1969) rate $D/(P + cD)$ will be of interest. It would also be of interest to extend the approach of this paper to handle the case of dependent lifetimes.

Acknowledgements

The author would like to thank J. W. Tukey for getting him interested in this problem. He would further like to thank C. L. Chiang, J. G. Fryer, J. W. Tukey, and the referees for helpful comments. Jan Hoem made many comments that substantially improved the structure of the paper and its completeness. Finally, the author would like to thank D. Nagnur of Statistics Canada for providing the data analysed. The research was supported in part by the National Science Foundation Grant DMS-8316634.

Résumé

L'objet principal de ce travail est le développement d'approximations des distributions des taux brutes de mortalité, des taux de mortalité par âges, des taux par âges standardisé, des rapports de mortalité standardisée et des paramètres analogues dans le cas de populations fermées. En supposant un processus Poissonnien sur les instants de naissance, et des durées de vie indépendantes, on trouve que le nombre de morts et la population correspondante à mi-année ont une distribution bivariée de Poisson. Le diagramme de Lexis utilise directement ce résultat. On suggère que dans des situations différentes, il serait satisfaisant d'approcher la distribution du nombre de morts conditionnellement à la taille de la population, par une loi de Poisson de moyenne proportionnelle à la taille de la population. De plus, on suggère que les situations où des variables explicatives existent, pourraient être modélisées par une distribution composée d'une loi de Poisson pour le nombre de décès, avec une moyenne proportionnelle à la taille de la population, d'une fonction exponentielle d'une combinaison linéaire des variables explicatives. Un tel modèle est ajusté aux données de mortalité des femmes canadiennes classées par âge. Une variante dynamique de ce modèle est de plus ajustée sur la série des décès totaux des femmes par année. Le modèle avec des variations Poissonniennes conduit à une amélioration sensible de l'ajustement.

References

Bartholomew, D. J. and Forbes, A. F. (1979). *Statistical Techniques for Manpower Planning.* New York: Wiley.

Benjamin, B. (1968). *Health and Vital Statistics.* London: Allen and Unwin.

Benjamin, B. and Pollard, J. H. (1980). *The Analysis of Mortality and Other Actuarial Statistics.* London: Heinemann.

Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46,** 443–460.

Braun, H. (1978). Stochastic stable population theory in continuous time. *Scandinavian Actuarial Journal,* 185–203.

Breslow, N. E. and Day, N. E. (1985). The standardized mortality ratio. In *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences,* P. K. Sen (ed.), 55–74. Amsterdam: North-Holland.

Breslow, N. E., Lubin, J. H., Marek, P., and Langholz, B. (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association* **78,** 1–12.

Brillinger, D. R. (1981). Some aspects of modern population mathematics. *Canadian Journal of Statistics* **9,** 173–194.

Brillinger, D. R. and Preisler, H. K. (1983). Maximum likelihood estimation in a latent variable

problem. In *Studies in Econometrics, Time Series and Multivariate Statistics*, T. Amemiya and S. Karlin (eds), 31–65. New York: Academic Press.
Chiang, C. L. (1961). Standard error of the age-adjusted death rate. *Vital Statistics—Special Reports*, Vol. 47, No. 9, Washington, D.C.: U.S. Public Health Service.
Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A* **148,** 82–117.
Cohen, A. (1983). Seasonal daily effect on the number of births in Israel. *Applied Statistics* **32,** 228–235.
Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events.* London: Methuen.
Daley, D. J. and Vere-Jones, D. (1972). A summary of the theory of point processes. In *Stochastic Point Processes*, P. A. W. Lewis (ed.), 299–383. New York: Wiley.
Davis, P. J. and Rabinowitz, P. (1975). *Methods of Numerical Integration.* New York: Academic Press.
Daw, R. H. (1974). A study of the variance of mortality rates. *Journal of the Institute of Actuaries* **101,** 415–434.
Doob, J. L. (1953). *Stochastic Processes.* New York: Wiley.
Farewell, V. T. (1982). The use of mixture models for the analysis of survival data and long-term survivors. *Biometrics* **38,** 1041–1046.
Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions.* New York: Wiley.
Frome, E. L. (1983). The analysis of rates using Poisson regression models. *Biometrics* **39,** 665–674.
Frome, E. L. and Checkoway, H. (1985). Use of Poisson models in estimating incidence rates and ratios. *American Journal of Epidemiology* **121,** 309–323.
Fryer, J. G., Harding, R. A., Macdonald, M. D., Read, K. L. Q., Crocker, G. R., and Abernathy, J. (1979). Comparing the early mortality rates of the local authorities in England and Wales. *Journal of the Royal Statistical Society, Series A* **142,** 199–209.
Greenhouse, J. B. and Wolfe, R. A. (1984). A competing risks derivation of a mixture model for the analysis of survival data. *Communications in Statistics—Theory and Methods* **13,** 3133–3154.
Haight, F. A. (1967). *Handbook of the Poisson Distribution.* New York: Wiley.
Heckman, J. and Singer, B. (1982). Population heterogeneity in population models. In *Multidimensional Mathematical Demography*, K. C. Land and A. Rogers (eds), 587–599. New York: Academic Press.
Heckman, J. and Singer, B. (1984). Econometric duration analysis. *Journal of Econometrics* **24,** 63–132.
Hinde, J. (1982). Compound Poisson regression models. In *GLIM82*, R. Gilchrist (ed.), 109–121. Lecture Notes in Statistics **14**. New York: Springer.
Hoem, J. M. (1976). The statistical theory of demographic rates. *Scandinavian Journal of Statistics* **3,** 169–185.
Hoem, J. M. (1978). Demographic incidence rates. *Theoretical Population Biology* **14,** 329–337.
Hoem, J. M. (1984a). Statistical analysis of a multiplicative model and its application to the standardization of vital rates: A review. *Stockholm Research Reports in Demography* **16**. Stockholm: University of Stockholm.
Hoem, J. M. (1984b). A flaw in actuarial exposed-to-risk theory. *Scandinavian Actuarial Journal*, 187–194.
Hogan, M. D., Chi, P-Y., Hoel, D. G., and Mitchell, T. J. (1979). Association between chloroform levels in finished drinking water supplies and various site-specific cancer mortality rates. *Journal of Environmental Pathology and Toxicology* **2,** 873–887.
Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71,** 75–83.
Inskip, H., Beral, V., and Fraser, P. (1983). Methods for age-adjustment of rates. *Statistics in Medicine* **2,** 455–466.
Izenman, A. J. and Zabell, S. L. (1976). Babies and the blackout: A time series analysis of New York City birth data. In *Proceedings of the American Statistical Association Section on Social Statistics*, 427–433. Washington, D.C.: American Statistical Association.
Jagers, P. (1974). Convergence of general branching processes and functionals thereof. *Journal of Applied Probability* **11,** 471–478.
Jagers, P. (1975). *Branching Processes with Biological Applications.* New York: Wiley.
Jagers, P. and Nerman, O. (1984). The growth and composition of branching populations. *Advances in Applied Probability* **16,** 221–259.
Keiding, N. and Hoem, J. M. (1976). Stochastic stable population theory with continuous time. I. *Scandinavian Actuarial Journal*, 150–175.
Kendall, D. G. (1948). On the generalized "birth-and-death" process. *Annals of Mathematical Statistics* **19,** 1–15.

Keyfitz, N. (1966). Sampling variance of standardized mortality rates. *Human Biology* **38,** 309–317.
Manton, K. G., Woodbury, M. A., and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics* **37,** 259–269.
Mosteller, F. M. (1969). Estimation of death rates. In *The National Halothane Study*, J. P. Bunker (ed.), 234–235. Bethesda, Maryland: National Institutes of Health.
Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression.* Reading, Massachusetts: Addison-Wesley.
Pocock, S. J., Cook, D. G., and Beresford, S. A. A. (1981). Regression of area mortality rates on explanatory variables: What weighting is appropriate? *Applied Statistics* **30,** 286–295.
Pollard, A. H. (1970). Random mortality fluctuations and the binomial hypothesis. *Journal of the Institute of Actuaries* **96,** 251–264.
Rosenblatt, J., Jackson, D. L., Dole, W. P., and Thompson, W. L. (1983). On uniform approximation of Poisson variates by normal ones. *Journal of Statistical Planning and Inference* **8,** 281–300.
Statistics Canada (1976). *Vital Statistics: General Mortality 1950–1972.* Ottawa: Statistics Canada.
Tsaregradskii, I. P. (1958). On uniform approximations of the binomial distribution by infinitely divisible laws. *Theory of Probability and Its Applications* **10,** 472–479.
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.
Vere-Jones, D. (1968). Some applications of probability generating functionals to the study of input–output streams. *Journal of the Royal Statistical Society, Series B* **30,** 321–333.
Yashin, A. I., Manton, K. G., and Vaupel, J. W. (1985). Mortality and aging in a heterogeneous population: A stochastic process model with observed and unobserved variables. *Theoretical Population Biology* **27,** 154–175.

Received June 1985; *revised March and April* 1986.

Appendix

The proof of Theorem 1 will make use of the method of probability generating functionals. The pertinent methodology may be found in Vere-Jones (1968) and Daley and Vere-Jones (1972), for example, To begin, note that for a general stochastic point process with points located at positions $\mathbf{r}_j$, the probability generating functional (p.g.fl.) is defined as

$$\mathrm{E}\left\{\prod_j \xi(\mathbf{r}_j)\right\}$$

for a general function $\xi(\cdot)$. The p.g.fl. characterizes a point process. The p.g.fl. of a Poisson process with intensity function $\nu(\cdot)$ is given by

$$\exp\left\{\int [\xi(\mathbf{r}) - 1]\nu(\mathbf{r})\, d\mathbf{r}\right\}. \tag{A.1}$$

Proof of Theorem 1. The ends of lifelines in the Lexis diagram occur at the positions $(\sigma_j + X_j, X_j)$ with σ_j denoting the birthtime and X_j the lifetime of the jth individual. The p.g.fl. of the death process is therefore

$$\mathrm{E}\left\{\prod_j \xi(\sigma_j + X_j, X_j)\right\}.$$

Now

$$\mathrm{E}_X\{\xi(t + X, X)\} = \int \xi(t + x, x)\mu(t + x, x)\exp\left\{-\int_0^x \mu(t + x - y, y)\, dy\right\} dx, \tag{A.2}$$

with X denoting the lifetime of an individual born at time t. Let $\eta(t)$ denote expression (A.2). It has been assumed that the σ_j correspond to a Poisson process of intensity $\beta(\cdot)$. Therefore, from (A.1),

$$\mathrm{E}\left\{\prod_j \eta(\sigma_j)\right\} = \exp\left\{\int [\eta(t) - 1]\beta(t)\, dt\right\}.$$

Combining these last expressions, one sees that the death process is (planar) Poisson with the indicated intensity, as was to be proved.

Doob (1953, pp. 405–407) is an early reference to this type of result—that random translations of Poisson processes are themselves Poisson.

Proof of Theorem 2. Let

$$X = \left(\frac{D}{\Lambda} - \frac{\lambda}{\Lambda}\right) \Big/ \frac{\sqrt{\lambda}}{\Lambda}$$

and

$$\epsilon = (\sqrt{\lambda} + x)\left(\frac{P}{\Lambda} - 1\right);$$

then elementary manipulations show that the quantity to be bounded is

$$\begin{aligned} &\leqslant \Pr\{x - |\epsilon| < X \leqslant x + |\epsilon|\} \\ &\leqslant \Pr\{|\epsilon| > \delta\} + \Pr\{|X - x| \leqslant \delta\} \end{aligned} \tag{A.3}$$

for any $\delta > 0$. Now, by Tchebycheff's inequality, the first probability here is

$$\leqslant \operatorname{var} \epsilon/\delta^2 = (\sqrt{\lambda} + x)^2\sigma^2/\delta^2\Lambda^2. \tag{A.4}$$

For D a Poisson variate of mean λ it is the case that

$$|\Pr\{D \leqslant u\} - \Pr\{D \leqslant v\}| \leqslant (u - v)c/\sqrt{\lambda} \tag{A.5}$$

for u, v integers $u > v$ with $c = \pi\sqrt{2\pi}/8$. This comes from Theorem 2 of Tsaregradskii (1958) and from bounding the absolute value of the characteristic function of D by $\exp(-2\lambda t^2/\pi^2)$.

The second probability in (A.3) for $\sqrt{\lambda} + x > 0$ is $\leqslant \Pr\{v < D \leqslant u\}$ with $u - v \leqslant 2\delta\sqrt{\lambda} + 2$. The result of the theorem now follows by adding (A.5) and (A.4) and then choosing δ to give the smallest total.

DISCUSSION ON THE PAPER BY DAVID R. BRILLINGER

Norman Breslow (Department of Biostatistics, SC-32, University of Washington, Seattle, Washington 98195, U.S.A.)

Professor Brillinger has provided an elegant theory for the occurrence of vital events in open populations from which he derives approximations to the "sampling" variability of widely used statistics such as age-standardized death rates and standardized mortality ratios. His development is a welcome step toward putting the statistical treatment of such data on a rigorous footing. My discussion concentrates on two issues: (i) the generality of his fundamental assumptions; and (ii) the availability of simple, alternative methods to handle the extra-Poisson variation often encountered in this domain.

In practice, standardized death rates and mortality ratios are routinely calculated for specific causes of deaths for open population units that change membership not only via births and deaths, but also via immigration, emigration, and loss due to death from competing causes. (In spite of his claim that the development is concerned with closed populations, the example and discussion make clear the intention to apply the theory to open populations.) Theorem 1, however, involves entry into the population only via the birth process and exit only via death. One wonders what additional assumptions may be required to justify the representation of (D, P) as a bivariate Poisson process when emigration, immigration, and loss are taken into account, or whether such a representation is even possible. These quibbles aside, I was pleased to see that the theory led to the standard approximate representation of D given P as Poisson, and to approximate formulas for the standard error of the death rate in agreement with those given by Breslow and Day (1986) among others.

An alternative approach to the distribution theory that does accommodate the usual features of open populations is of course via the theory of point processes applied to censored survival data (Aalen, unpublished Ph.D. dissertation, Department of Statistics, University of California, Berkeley, 1975; Andersen and Gill, 1982; Borgan, 1984; Kalbfleisch and Prentice, 1980). Such realistic features of individual life histories as left truncation (immigration) and right censoring (emigration or loss) are easily dealt with. Furthermore, this theory provides a stochastic mechanism for the occurrence of the vital events (deaths) even when the times of entry into the population are considered fixed rather than random. Of course, it also presumes that individual times of entry and exit into the population are available.

Provided that the age × time cells are sufficiently small that the death rates within each one can be assumed constant, the likelihood kernel that arises from a parametric description of the unknown rate parameters (Borgan, 1984) is identical to that which results by regarding the number of deaths in separate cells as having independent Poisson distributions with means equal to the (unknown constant) rates times (fixed) person-years denominators. This fact provides an alternate justification for the Poisson approximation of Theorem 2, at least so far as large-sample likelihood inferences about the rates are concerned.

Although this is not explicitly stated, Theorem 2 would seem to invoke the same rare-event hypothesis that Brillinger has faulted in the work of others. For P Poisson, we have $\sigma = \Lambda^{1/2}$. Thus, the right-hand side of (5.1) will be small only if $\lambda^{1/2}$ is large and $(\lambda/\Lambda)^{1/2}$ is small. Roughly speaking, this means that the expected number of deaths must be large, while the (average) death rate λ/Λ must be small. Thus, it is still not clear (to me at least) that the expressions given in Section 5 to estimate variances of the crude or adjusted death rates are valid when the component rates are large, no matter what the coefficient of variation of P.

An alternative method for fitting the log-linear model (6.1) with a random error term to accommodate extra-Poisson variation was described by Breslow (1984). Briefly, one considers a generalized linear model (McCullagh and Nelder, 1983) where the mean μ is related to the linear predictor $\mathbf{x}'\boldsymbol{\beta}$ via the log-link $\log(\mu) = \mathbf{x}'\boldsymbol{\beta}$ and where the variance function satisfies $V(\mu) = \mu + \sigma^2\mu^2$. No distributional assumptions are made about the random perturbations except that they have mean 0 and variance σ^2. Estimates of $\boldsymbol{\beta}$ and σ^2 are obtained via an iterative procedure whereby, for fixed σ^2, quasilikelihood is used to estimate $\boldsymbol{\beta}$ whereas, for fixed $\boldsymbol{\beta}$, σ^2 is estimated by setting the Pearson sum of squared standardized residuals equal to its degrees of freedom. This procedure may be implemented easily in GLIM without having to expand the data vector as in Hinde's (1982) adaptation of the normal theory model. [Incidentally, an early reference to this model is found in unpublished work by Pierce and Sands (Technical Report 46, Oregon State University, 1975).] Large-sample properties of the resulting estimates have been derived by Moore (1986), who shows in particular that the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is unaffected by the need to estimate the dispersion parameter σ^2. Thus, the standard errors resulting from the usual quasilikelihood formulation (which assumes σ^2 known) are applicable.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large-sample study. *Journal of the American Statistical Association* **10,** 1100–1120.

Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics* **11,** 1–16, 275.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33,** 38–44.

Breslow, N. E. and Day, N. E. (1986). *Statistical Methods in Cancer Research.* II: *Design and Analysis of Cohort Studies.* Lyon: International Agency for Research on Cancer (in press).

Hinde, J. (1982). Compound Poisson regression models. In *GLIM82*, R. Gilchrist (ed.), 109–121. Lecture Notes in Statistics **14**. New York: Springer.
Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.
McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* New York: Chapman and Hall.
Moore, D. F. (1986). Asymptotic properties of moment estimates for overdispersed counts and proportions. *Biometrika* (in press).

Chin Long Chiang (School of Public Health, University of California, Berkeley, California 94720, U.S.A.)

In this article, David Brillinger has presented an extensive discussion on the variability of vital rates. His conceptual model of the planar Poisson process represents a new approach to the problem. But I still cannot reconcile the differences between the results in his article and those presented in my 1961 paper. In this discussion I shall make two comments on Brillinger's article and clarify two points from my 1961 paper.

The main issue in question is the variability of the number of deaths D_x in the death rate

$$M_x = \frac{D_x}{P_x} \tag{1}$$

for an age group $(x, x + n_x)$ in a current population.

When should we begin counting the variability associated with D_x? Brillinger's planar Poisson process begins to operate prior to the time of birth of individuals in age group $(x, x + n_x)$, or $x + n_x$ years before the current calendar year. If we consider age group (40, 45) in the year 1980, according to the planar process, we need to keep track of the variability from the year 1935 of everyone born between 1935 and 1940. But the death rate M_x in (1) is a measure of the mortality experience of a group of people during the age interval $(x, x + n_x)$ in a current year. The death process prior to age x has already taken place. Randomness should no longer be attached to those who have already died prior to x. Only the variability of the death process of those living in the age interval $(x, x + n_x)$ during the current year need be considered. In other words, we consider the variability associated with D_x from exact age x on.

The binomial distribution In the corollary to Theorem 1, Brillinger states: "...(c) $N(B)$ given $N(C)$ is distributed as $U + S$, where U is Poisson with mean $\Lambda(B - BC)$ and S is independently binomial with mean $n = N(C)$ and proportion $\Lambda(BC)/\Lambda(C)$." He further illustrates this assertion with crude death rates and age-specific death rates. In Example 2, age-specific death rates, we find: "$N(B_x)$ and $N(C_x)$ count, respectively, how many die aged x to $x + 5$ in 1980 and how many were alive and aged x to $x + 5$ on 1 July 1980." These points are all good and clear. But they seem to be inconsistent with one's intuition. The $N(C_x)$ people are of different ages on 1 July 1980 and are subject to different probabilities of dying. The number of people among $N(C_x)$ who die in the second half of the year 1980 will not have a binomial distribution.

An interpretation of N_x In Chiang (1961) we introduced a number N_x as "a hypothetical population at age x, in which D_x deaths will occur in the age interval (x to $x + n_x$) ..."; we considered D_x a binomial variable and $\hat{q}_x = D_x/N_x$ a binomial proportion so that the variance of D_x is $D_x(1 - \hat{q}_x)$. Now we argue that N_x may be taken as the total number of individuals who reached exact age x during the current calendar year. This may be verified by means of a Lexis diagram. Figure 1 illustrates the situation for age interval (40, 45) during the year 1980. The rectangle (40, 45) × 1980, denoted by A, corresponds to region A in Brillinger's planar process. The parallelogram to the right of the rectangle, denoted by

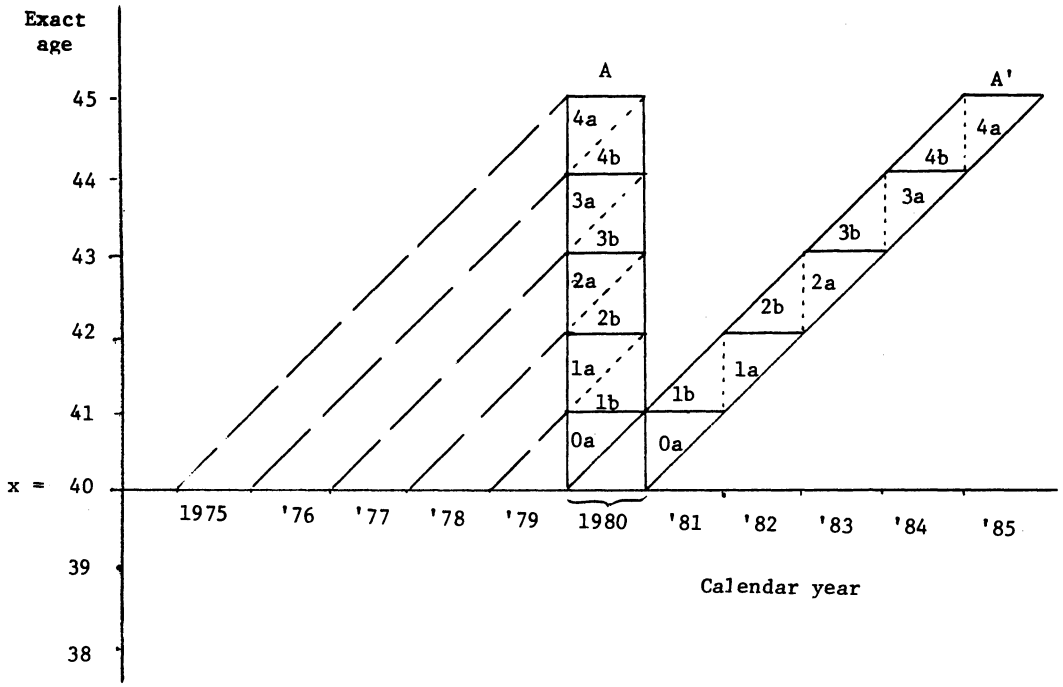


Figure 1. Lexis diagram.

A', also has 1980 as its base and its area is equal to the area of A. The lifeline of an individual beyond age 40 starts at a point on the line $x = 40$ extending diagonally upward. Everyone who dies at an age in the interval (40, 45) during the year 1980 will have his lifeline ending in the rectangle. For example, a person who has attained his 40th birthday in 1975 and dies at age 44 in 1980 will have his lifeline ending in the triangle 4a. There are D_x lifelines ending in A.

Parallelogram A' contains the lifelines of all individuals beyond age 40 who attained exact age 40 during the calendar year 1980. Those who die in the age interval (40, 45) will have their lifelines ending in the parallelogram. An individual who attained age 40 in 1980 and dies at age 44 in 1985, for example, will have his lifeline ending in the triangle 4a, at the upper right corner of region A'. The two triangles denoted by 4a in the two regions A and A' contain lifelines of people of the same age. Other pairs of triangles in A and A' with the same designation also contain lifelines of people of the same age.

Now we use the concept of stationary population and assume that the number of lifelines ending in rectangle A is equal to the number of lifelines ending in parallelogram A'. Then the number of lines ending in the parallelogram is also equal to D_x. That is to say, the number of people who attained exact age $x = 40$ during the year 1980 and die of age (40, 45) is also equal to D_x. Thus, in general, if we let N_x be the number of people who attain exact age x during the current calendar year, D_x of the N_x individuals will die during the age interval $(x, x + n_x)$. Furthermore, using the age-dependent time-homogeneous Brillinger planar process, one can show that the expected number of deaths in region A' is equal to the expected number of deaths in region A. That is, $\mathrm{E}(D_x)$ in A' is equal to $\mathrm{E}(D_x)$ in A.

Including the variability of P_x in the variance of M_x There are two measures in general use that summarize the mortality experience of a given population during the age interval $(x, x + n_x)$: the death rate M_x and the conditional probability q_x that an individual alive at

exact age x will die in the age interval $(x, x + n_x)$. Studies of the relationship between the two measures have generated rich literature in life-table construction for current populations. King (1914), Reed and Merrell (1939), Greville (1943), Wiesler (1954), Sirken (1964), and Keyfitz (1966), to name a few, have all contributed to the methods of life-table construction. The following formula was suggested in Chiang (1961) to compute $\hat{q}_x$ from M_x:

$$\hat{q}_x = \frac{n_x M_x}{1 + (1 - a_x) n_x M_x}, \tag{2}$$

where a_x is the fraction of the interval $(x, x + n_x)$ lived by an individual who dies in the interval. Justification of this formula also has been given in Chiang (1972), Elandt-Johnson and Johnson (1980), and Golbeck (1986). See also Keyfitz (1968). Solving (2) for M_x yields

$$M_x = \frac{\hat{q}_x}{(1 - \hat{q}_x) n_x + \hat{q}_x a_x n_x}. \tag{3}$$

Formula (3) is easier to interpret when M_x and $\hat{q}_x$ are considered as their corresponding theoretical quantities. Thus, for an individual alive at exact age x, q_x in the numerator is equal to the expected number of deaths in the interval $(x, x + n_x)$. For the denominator, we realize that if a person is to survive the interval, with a probability $(1 - q_x)$, his period of exposure to the risk of dying is the entire length of the interval—namely, n_x. If he is to die in the interval, with a probability q_x, his expected length of exposure is $a_x n_x$. Therefore, the right-hand side of formula (3) is the ratio of the expected number of deaths in the interval $(x, x + n_x)$ to the expected length of exposure to the risk of dying in the interval. That is the definition of the age-specific death rate.

We can make use of (3) to include the variability of P_x in the variance of the death rate M_x. Substituting $\hat{q}_x = D_x/N_x$ in (3) yields

$$M_x = \frac{D_x}{n_x N_x - (1 - a_x) n_x D_x}, \tag{4}$$

which contains D_x as the only random variable. And the midyear population P_x in (1) becomes an estimate of $[n_x N_x - (1 - a_x) n_x D_x]$ in the denominator in (4). An approximate formula for the variance of M_x can be obtained from

$$V(M_x) = \left(\frac{d}{dD_x} M_x\right)^2 V(D_x), \tag{5}$$

namely,

$$V(M_x) = \frac{1}{P_x} M_x (1 - \hat{q}_x)(n_x M_x / \hat{q}_x)^2. \tag{6}$$

The variance in (6) is greater than the variance

$$V(M_x) = \frac{1}{P_x} M_x (1 - \hat{q}_x) \tag{7}$$

by a factor $(n_x M_x/\hat{q}_x)^2$, which is always greater than unity.

Remark Brillinger used a planar Poisson process to derive a formula for the variance of the age-specific death rate, while I considered D_x as a binomial random variable among N_x people who have reached exact age x. The two approaches, however, have resulted in two different formulas for the variance of the death rate.

References

Chiang, C. L. (1961). Standard error of the age-adjusted death rate. *Vital Statistics—Special Reports*, Vol. 47, No. 9, 275–285. Washington, D.C.: U.S. Public Health Service.

Chiang, C. L. (1972). On constructing current life tables. *Journal of the American Statistical Association* **67,** 538–541.

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis*. New York: Wiley.

Golbeck, A. L. (1986). Probabilistic approaches to current life table estimation. *The American Statistician* **40,** 185–190.

Greville, T. N. E. (1943). Short methods of constructing abridged life tables. *Record of American Institute of Actuaries* **32,** 29–42.

Keyfitz, N. (1966). A life that agrees with data. *Journal of the American Statistical Association* **61,** 305–311.

Keyfitz, N. (1968). *Introduction to the Mathematics of Population.* Reading, Massachusetts: Addison-Wesley.

King, G. (1914). On a short method of constructing an abridged mortality table. *Journal of the Institute of Actuaries* **48,** 294–303.

Reed, J. L. and Merrell, M. (1939). A short method for constructing an abridged life table. *American Journal of Hygiene* **30,** 33–62.

Sirken, M. G. (1964). Comparison of two methods of constructing abridged life tables by reference to a "standard" table. *Vital and Health Statistics* Series 2, No. 4, 1–11. Washington, D.C.: National Center for Health Statistics, U.S. Department of Health, Education and Welfare.

Wiesler, H. (1954). Une méthode simple pour la construction de tables de mortalité abrédées. *Proceedings of the World Population Conference*, Vol. IV, 1063–1073. New York: United Nations.

Jan M. Hoem (Section of Demography, University of Stockholm, S-106 91 Stockholm, Sweden)

Brillinger's paper is an appealing reminder of the message that one should recognize the natural variability in vital statistics, and that better insight into population processes can be gained by describing such data by proper statistical tools. The pleasing simplicity of several of his results will facilitate the analysis of data for which his assumption of Poisson births is a reasonable approximation, such as for data about a single cohort or a few consecutive cohorts, or for an age-specific period death rate, such as in Example 2 of his Section 4.

Unfortunately, the assumption of Poisson births disregards essential aspects of the internal dynamics of real-life populations. Results based on the Poisson assumption must have little relevance for the analysis of statistics which involve data for several generations, such as the crude death rate in Example 1 of Section 4. The Poisson assumption could be all right in a population where "births" are generated by an external mechanism in the manner sometimes used to model immigration, which would then occur at "age" 0. In normal populations, children are borne by population members, however, so the occurrence and timing of births would depend on the size and the (age) composition of the population itself—i.e., births in one period are generated by population events in previous periods and in the same period. The consequences are immediately apparent in any population data which involve both parent and offspring cohorts, at least when there are parent cohorts of both normal and deviant sizes. To elaborate the evident, in a population where births were generated by a Poisson process, there would be no reflection in the number of births a generation after births were depleted by a famine or a depression, or swelled by a baby boom, and the reduction of the size of a parent generation by the toll of a major war would have no consequences for the number of subsequent births. In such a Poisson world, stable population theory would be of no relevance and there would be no room for economic explanations such as the Easterlin hypothesis [see, e.g., Easterlin, Wachter, and Wachter (1978)].

Conversely, since real populations *are* subject to the internal dependencies which are made manifest by population waves, the assumption of Poisson births is untenable in multigenerational analyses.

These commonplace dependencies are of course recognized in the usual population mathematics, both in deterministic and stochastic formulations. In Scandinavia, the discovery of population waves (and their use to explain population behaviour) is attributed to Sundt (1855, Chaps 5–7); see Vogt (1968). For appreciations of Sundt's pioneering work as a sociologist and a statistician, see Otnes (Working Paper 86, Department of Sociology, University of Oslo, 1977), Seip (1986), and Iversen (1983).

Westergaard (1880) realized already a century ago that there can be stochastic variation in vital statistics even when the data do not come from a sample survey; see Keiding (1987). Since promotion of this idea is still needed, however, there may be some use for an illustration documenting random fluctuations even at the national level. The two curves in Figure 1 are plots of ordinary occurrence/exposure rates for first births to married women born in 1945, for two groups of ages at marriage, by duration of first marriage in 1-month intervals (2-month intervals after 60 months). The data are for women of Swedish citizenship according to the census of 1960, resident in Sweden, and nulliparous at first marriage. The solid curve is for 17,211 women first married at ages 20–22. The broken curve is for 9755 women first married at ages 23–25. The latter curve has been plotted only up to a duration of 30 months, for after that the two curves overlap extensively. These are complete segments of a total national population of some 8.3 million people. No sampling procedure is involved.

Figure 1. Marital first birth rates to Swedish women born in 1945, by duration of first marriage, for two groups of ages at marriage. One-month intervals up to a marital duration of 60 months; 2-month intervals afterward.

Each plot gives a striking impression of a curve with a marked systematic structure, overlaid with random fluctuations. There is an obvious structural change as we move from the younger to the older age group at marriage, and it is largely interpretable in terms of a reduced tendency for premarital conceptions. This fits in a pattern covering neighbouring age groups and extending to other cohorts. Nevertheless, the presence of random variation is evident, and it masks the trends at longer durations than about a year of marriage. Note how the appearance of random fluctuations is dampened as we pass a duration of 60 months and switch from 1-month to 2-month intervals.

Leif Johansson had the data extracted from the Swedish Central Population Register. They are presumed to be of unsurpassed quality. The curves of Figure 1 have been selected from a larger material for their nice features, but many other empirical plots from our own experience would have served our purposes equally well. The computations for Figure 1 were made by Fjalar Finnäs.

References

Easterlin, R. M., Wachter, M. L., and Wachter, S. M. (1978). Demographic influences on economic stability. *Population and Development Review* **4**, 1–22.

Iversen, G. (1983). Eilert Sundt as a statistician (in Norwegian with a summary in English). *Tidsskrift for samfunnsforskning* **24**, 511–536. Universitetsforlaget, Oslo.

Keiding, N. (1987). The method of expected number of deaths 1786–1886–1986. *International Statistical Review* **55** (to appear).

Seip, A. (1986). Eilert Sundt: A founding father of the social sciences in Norway. *Scandinavian Journal of History* **11** (to appear).

Sundt, E. (1855). *Om Giftermaal i Norge.* Mallings, Christiania. Reissued in 1967 by Universitetsforlaget, Oslo. Translation (*On Marriage in Norway*) by M. Drake published in 1980 by Cambridge University Press.

Vogt, J. (1968). Eilert Lund Sundt. In *International Encyclopedia of the Social Sciences* **15**, 409–411. D. L. Sills (ed.). New York: Macmillan & Free Press.

Westergaard, H. (1880). An exposition and evaluation of recent investigations on the mortality of the various classes of society (in Danish). University of Copenhagen Prize Paper. Published 1882 as *Die Lehre von der Mortalität und Morbilität: Anthropologisch-statistische Untersuchungen* by Gustav Fischer, Jena.

Peter Jagers (Department of Mathematics, Chalmers University of Technology and University of Göteborg, S-412 96 Göteborg, Sweden)

Problems of population growth and composition meet in quite different contexts, and are analyzed by quite different means. There are human populations, studied in a demographic tradition. There is the very active area of population dynamics, essentially of a zoological or ecological nature. There are cell kinetics, bacteriology, population genetics, and several types of particle multiplication in physics. And in all cases there is an interplay between special attributes of the particular system under study, and general properties common to all sets of (more or less) freely reproducing individuals under (more or less) stable or at least well-described conditions. There is finally a more abstract, mathematical study of precisely these last properties, epitomized into the somewhat Platonic free populations of branching process theory.

Time is more than ripe for bringing all these cultures a little bit closer to one another. Probabilists and statisticians should learn the problems of empirical population science, and empirical scientists should be relieved of the unnecessary burden of refinding what has long been known.

From this point of view the present paper is more than welcome. It is written by a well-known statistician and connects a demographic problem with more general properties of

point processes. Brillinger's observation, that even in a not necessarily time-homogeneous setting a Poisson property of the population birth process enhances a two-dimensional "Poisson-ness" of the population death process, seems to the point and might well prove very useful. (Mathematically it is of course, as Dr Brillinger points out, just an example of the well-known translation properties of Poisson processes.) For my part, however, I shall not discuss the proposed applications but rather concentrate on the more theoretical aspirations of the paper: "the stochastic conceptualization and formalization of the natural variability of vital statistics" as the author unhesitatingly formulates his purpose.

The first issue to be tackled is then how to explain this variability in terms of stochastic phenomena, the basic events of life seeming, as Dr Brillinger puts it, "facts (that is, exact values) and hence not subject to uncertainty." The paper presents some (in my view rather weak) arguments in favour of treating birthtimes and lifespans as random. Here is my try:

When an individual is born, his or her life could take many different paths. Let us denote the enormous space of all these perceivable life careers by Ω. An element $\omega \in \Omega$ is thus a complete biography giving information about anything occurring in a life, such as at which ages an individual leading the life ω begets children, and at what age he or she dies. In other words, the basic demographic events at an individual level are functions defined on the abstract space Ω. For example, the age at death X (in Brillinger's notation) is a function from Ω into the nonnegative half-line. Hence, if Ω is equipped with some σ-algebra $\mathcal{A}$ of relevant events and a probability measure P, telling how likely different demographic events are, then X is a random variable on a life space $(\Omega, \mathcal{A}, P)$, quite according to the textbook pattern. As usual, this does not at all contradict that for ω given, $X(\omega)$ is a number, i.e., "has an exact value." And obviously the same goes for entities such as the number of births to be given, or ages at childbearing and any other example you could think of (such as duration of mitosis in cell kinetics).

In this way what seems a philosophical paradox—how can there be randomness when everything is determined—dissolves itself by a careful mathematical formulation (as is so often the case with statistical paradoxes). To sum up, there is no more problem with randomness in population dynamics than in coin tossing.

But not only does this construction save us from philosophical headache, but it also guides us into a strict construction of the whole population process. From $(\Omega, \mathcal{A}, P)$ (or slightly more complicated versions, if individuals can be of various types and time inhomogeneity is allowed), a population space can be built by some Ulam–Harris device [cf. Jagers (1975) or Jagers and Nerman (1984b)] or by Neveu's tree technique (Neveu, 1986). On this space the vital statistics such as counts, rates, or ratios of various kinds are then well-defined random variables.

In the more precise framework thus arising one could therefore also study the assumption boldly postulated in the paper now under discussion. Under what circumstances is a process of births (or deaths) (approximately) Poisson? For the closed, time-homogeneous case the issue has actually been analyzed by Härnqvist (1981), proving a local Poisson tendency as the population grows. Related Poisson properties for quite general populations have been exhibited by Jagers and Nerman (1984a).

However, these are results that must be used with much care. Strictly speaking, reproductive systems can never be Poisson due to the feedback that occurs and generates dependence between occurrences in disjoint sets of the age–time space. What we can hope for is limit theorems, for closed populations, as in Härnqvist's paper, or for open populations by results for closed populations combined with classical theorems on superposition of many independent sources of immigration. And such limit theorems will more easily be proved directly for the various point processes (of births, deaths, or other events) than via the translation property of Poisson processes.

REFERENCES

Härnqvist, M. (1981). Limit theorems for point processes generated in a general branching process. *Advances in Applied Probability* **13,** 650–668.

Jagers, P. (1975). *Branching Processes with Biological Applications.* New York, Chichester: Wiley.

Jagers, P. and Nerman, O. (1984a). Limit theorems for sums determined by branching and other exponentially growing processes. *Stochastic Processes and Their Applications* **17,** 47–71.

Jagers, P. and Nerman, O. (1984b). The growth and composition of branching populations. *Advances in Applied Probability* **16,** 221–259.

Neveu, J. (1986). Arbres et processus de Galton–Watson. *Annales de l'Institut Henri Poincaré. Probabilités et Statistique* **22,** 199–208.

Niels Keiding (University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark)

Rates as estimators of parameters Brillinger's main interest is to describe the variability of rates such as the "age x death rate" $M_x = D_x/P_x$ specified in his formula (2.1). This is taken right out of demographical practice, using the midyear population P_x as denominator. A more satisfactory theoretical–statistical analysis would, however, be to consider such rates as estimators resulting from well-defined statistical estimation problems. Thus, in the problem of estimating an (age-specific) death intensity μ_x (assumed constant for the relevant age group) under full continuous observation, the maximum likelihood estimator is the occurrence/exposure rate D_x/A_x with A_x = exposure to death = the total time lived in the age interval by the members of the population. In demographic practice, A_x often cannot be observed, but to go from A_x to various common demographic approximations should to the theoretical statistician be considered a problem of estimation under incomplete observation (Keiding, 1976). In this systematic statistical approach there is no doubt that the rates, as well as the intensities they estimate, are in units time^{-1}, and also the inherently random nature of the denominator (simply due to the fact that deaths happen at "random" times over the year, producing a random total exposure A_x) is obvious.

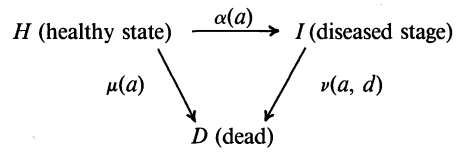
The standard error of the standardized mortality ratio (SMR) Since Brillinger's point of departure seems to have been related to computing the standard error of a standardized mortality ratio (SMR), a few more detailed comments may be appropriate here.

A modern statistical discussion is again phrased naturally in terms of an ordinary estimation problem, this time of a *relative mortality* θ compared to some known standard age-dependent mortality (Breslow and Day, 1975). Ordinary large-sample maximum likelihood theory yields the simple estimator SMR/(expected deaths) of var(SMR), also obtained by Brillinger in Section 5. [Breslow and Day (1985) present the state of the art concerning alternative estimators of the standard error of the SMR.]

The simple estimator is often ascribed to Yule (1934), who obtained it by approximating with a deterministic denominator (as in Brillinger's Theorem 2) and using a Poisson approximation to the "canonical" binomial distribution of the observed number of deaths in the numerator. At Yule's time the calculation of expected deaths, and hence of the SMR, was a very well-established method in official statistics (it may in fact be traced back at least to 1777), but it is particularly interesting in the present context that the view of death counts, rates, and so forth as essentially random phenomena already penetrated the textbook by Westergaard (1882), who, like Brillinger, provided fresh input to current demographic and official statistical methodology by implementing ideas from mathematical statistics. Westergaard derived the same simple standard error estimator as mentioned above and used it extensively in his many comparative studies of mortality and morbidity. Further details on the history of the method of expected number of deaths were given by Keiding (1987).

The fact that the denominator of the SMR ("the expected number of deaths") is essentially random has caused confusion in quite recent medical statistics (cf. Keiding and Væth, 1986), although Berry (1983) has now provided, in this journal, an authoritative survey of its use, will full theoretical–statistical documentation.

Epidemiological information in a cross-sectional sample Finally, let me indicate how the idea of feeding the Lexis diagram with a homogeneous Poisson process of births may prove useful in developing a consistent mathematical foundation for the interplay between the cohort and the cross-sectional viewpoints in epidemiology—that is, when the individuals may enter an irreversible disease state before death. Assume that the life history of an individual is governed by the semi-Markov process specified by the diagram



where the *incidence* $\alpha(a)$ and the *mortality* $\mu(a)$ of healthy individuals depend on age a only whereas the mortality $\nu(a, d)$ of diseased individuals may in addition depend on d, the duration in the diseased state I. If the birth process is Poisson with intensity λ, then similar Poisson process considerations as used by Brillinger yield the following quantities relevant to a cross-section of the population at a particular time, say $t = 0$:

E(# ind. in I at time 0 aged $[z, z + dz)$ with first appearance $[y, y + dy)) = h(y, z)\, dy\, dz$,

where $h(y, z) = 0$ for $y > z$ and

$$h(y, z) = \lambda \exp\left\{-\int_0^y [\mu(u) + \alpha(u)]\, du\right\}\alpha(y)\exp\left[-\int_y^z \nu(u, u - y)\, du\right]$$

when $0 < y < z$;

and similarly

$$\text{E(\# ind. in } H \text{ at time 0 aged } [z, z + dz)) = k(z)\, dz,$$

where

$$k(z) = \lambda \exp\left\{-\int_0^z [\mu(u) + \alpha(u)]\, du\right\}.$$

It is readily seen that

$$\lambda^{-1} \int_0^\infty \int_y^\infty h(y, z)\, dz\, dy = M_I,$$

the *mean diseased lifetime*, and that

$$\lambda^{-1} \int_0^\infty k(z)\, dz = M_H,$$

the mean healthy lifetime, so that the joint age–status–duration distribution in a cross-sectional sample is given by the density

$$\begin{cases} h(y, z)/M, & \text{diseased individual with first occurrence at } y \\ & \quad \text{and current age } z, \quad y < z, \\ k(z)/M, & \text{healthy individual with current age } z, \end{cases}$$

$$M = M_H + M_I = \text{mean lifetime.}$$

This representation makes it immediate to derive likelihood functions corresponding to various sampling plans for a cross-sectional sample, containing information on (i), current age for all and (retrospectively collected) age at first occurrence for the diseased; or (ii) current age and age at first occurrence from the prevalent cases only; or (iii) the distribution of current age for all.

The likelihoods simplify into easily tractable examples of current survival analysis problems under various special statistical models, such as (i) $\nu(a, d) = \mu(a)$, that is, nondifferential mortality for the diseased; or (ii) $\nu(a, d) = \nu(d)$, that is, age-independent but duration-dependent mortality for the diseased; or (iii) one or more intensities constant.

These details are presented in unpublished work by Keiding (Research Report 86/3, Statistical Research Unit, University of Copenhagen, 1986), as are the following interpretations of the well-known relation in epidemiology:

$$\text{Prevalence} = \text{Incidence} \times \text{Duration}.$$

A first (calendar-time dynamics) version is to interpret *prevalence* as the expected number of diseased individuals at time 0, that is,

$$\int_0^\infty \int_y^\infty \lambda h(y, z)\, dz\, dy = \lambda M_I,$$

incidence as the expected number of new cases per calendar-time unit:

$$\lambda \int_0^\infty \alpha(y) \exp\left\{ - \int_0^y [\mu(u) + \alpha(u)]\, du \right\} dy = \lambda A,$$

and *duration* as the conditional expectation Δ of remaining life, given that disease has just occurred. It is then elementary that

$$\lambda M_I = \lambda A \Delta$$

so that this relation is true in the time-homogeneous regime with no restriction on the age- and duration-dependence of the intensities.

A second interpretation (cf. Miettinen, 1976) states that

$$\text{Prevalence odds [= Prevalence/(1 - Prevalence)]} = \text{Incidence} \times \text{Duration}.$$

Interpreting prevalence odds as the ratio between the expected number of diseased individuals and the expected number of healthy individuals will yield

$$\lambda M_I / \lambda M_H = M_I / M_H,$$

so that if duration is interpreted as Δ as above, one would have

$$\text{Incidence} = \frac{M_I}{\Delta M_H} = \frac{A}{M_H} = \frac{\int_0^\infty \alpha(s) \exp\{-\int_0^s [\alpha(u) + \mu(u)]\, du\}\, ds}{\int_0^\infty \exp\{-\int_0^s [\alpha(u) + \mu(u)]\, du\}\, ds},$$

which is the mean incidence (intensity) with respect to the (expected) distribution of current age in the healthy state H. In the particular case of age-independent incidence α, this of course reduces to $A/M_H = \alpha$, so that prevalence odds = $\alpha\Delta$. We note that this relation concerns the (mean) *age-dynamics*.

A heuristic, discrete-time survey of prevalence, incidence, and duration concepts was given by Freeman and Hutchison (1980), and the device of Poisson input was used for studying length-biased sampling by Simon (1980).

References

Berry, G. (1983). The analysis of mortality by the subject-years method. *Biometrics* **39,** 173–184.

Breslow, N. E. and Day, N. E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases* **28,** 289–303.

Breslow, N. E. and Day, N. E. (1985). The standardized mortality ratio. In *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*, P. K. Sen (ed.), 55–74. Amsterdam: North-Holland.

Freeman, J. and Hutchison, G. B. (1980). Prevalence, incidence and duration. *American Journal of Epidemiology* **112,** 707–723.

Keiding, N. (1976). Contribution to the discussion of J. M. Hoem: The statistical theory of demographic rates: A review of current developments. *Scandinavian Journal of Statistics* **3,** 179–180.

Keiding, N. (1987). The method of expected number of deaths 1786–1886–1986. *International Statistical Review* **55** (to appear).

Keiding, N. and Væth, M. (1986). Calculating expected mortality. *Statistics in Medicine* **5,** 327–334.

Miettinen, O. (1976). Estimability and estimation in case-referent studies. *American Journal of Epidemiology* **103,** 226–235.

Simon, R. (1980). Length-biased sampling in etiologic studies. *American Journal of Epidemiology* **111,** 444–452.

Westergaard, H. (1882). *Die Lehre von der Mortalität und Morbilität.* Jena: Fischer.

Yule, G. U. (1934). On some points relating to vital statistics, more especially statistics of occupational mortality (with Discussion). *Journal of the Royal Statistical Society* **97,** 1–84.

Leslie Kish (Institute for Social Research, The University of Michigan, Ann Arbor, Michigan 48106, U.S.A.)

This is a stimulating and important work that combines theoretical and empirical contributions, usually more difficult and more valuable than most of those confined to one of those ends. It would be difficult for me to add to the technical aspects of this fine paper, but I shall remark briefly on the philosophical issues it stimulates.

Let us begin with two contradictory definitions:

A: "Statistics, the science concerned with the collection and analysis of numerical information in order to answer questions wisely."

B: "Statistics and statisticians deal with the effects of chance events on empirical data."

Definition A admits the counts of complete censuses as *fixed* statistical data, but B views them as products of stochastic processes. Both of them include census statisticians to the fold, but only B insists that we draw boundaries between their work and those of accountants, cashiers, and bankers. And let us not base that distinction only on the errors of measurements; let us concentrate instead on the basic philosophical aspects of stochastic variations. It is curious that definition A was written by Brillinger (1985), whose article strongly supports B, written by me (1978).

Yet I (and other practical survey samplers) write in terms of finite populations of fixed N members. Thus, there exists widespread belief that the theory of sampling concerns chiefly the fixed parameters of *finite* populations. Indeed, *probability sampling* requires randomized selections from *frames* that define finite populations for statistical inferences from sample statistics to the parameters of *frame populations.*

Most survey samplers recognize, sometimes in writing, the existence of and the need for "inferential" populations under (or above, behind, or beyond) the frame populations; that goes both for samples and for complete censuses. Sometimes the terms "target population" or "superpopulation" have been used, but these have also been loaded and confused with other meanings and I prefer to avoid them, as I explain elsewhere (Kish, 1987, §2.1).

The fundamental aspect consists of separating the two steps of inference. The first step involves statistical inference from the sample to the frame populations based on mechanical randomization, and it is well defined, objective, and public. The second step from the

frame population to the inferential populations can be multiform, multipurpose, mathematical or scientific, subjective, and private.

Some statisticians advance models depending on "superpopulations," and they view sample selections as if taken directly from them, thus merging into one step the two steps separated above. However, I distrust this combined view that would merge the defined, public, objective, complex probability selections with some subjective, changeable models concerning realizations from superpopulation models. The probability selections often, or usually, use complex procedures of clustering, subsampling, and stratification. The field of survey sampling owes its separate existence mainly to those practical, necessary complexities. On the other hand, sampling theorists tend to propose models of superpopulations that are relatively simple and mathematical. The clustered structures of samples and of populations is especially important to survey sampling. For consistency, such clustering and complexities should also be reflected in the relations of superpopulations to finite populations. Brillinger also notes the need for going beyond a simple Poisson model.

We want more models of superpopulations like Brillinger's, that arise out of actual, scientific problems. Discussions like his about the structure or process that underlie a definite realization in a finite population should help our theoretical understanding of the relation of samples to finite populations and censuses and of these to superpopulations.

Let me admit that I do not know which survey samplers share the strong feelings I expressed above about: (a) complex, clustered superpopulation models, (b) the two separate steps to frame and then to inferential populations, and (c) the distinction between sampling theory and survey sampling. I also propose for consistency similar views both for analytical statistics for relationships and for descriptive statistics.

References

Brillinger, D. R. (1985). Statistics. In *The Canadian Encyclopedia.* Edmonton: Hurtig.

Kish, L. (1978). Chance, statistics, and statisticians. *Journal of the American Statistical Association* **73**, 1–6.

Kish, L. (1987). *Statistical Design for Research.* New York: Wiley-Interscience.

Kenneth G. Manton (Duke University, Center for Demographic Studies, Durham, North Carolina 27706, U.S.A.)

The paper by Dr Brillinger is a clear and thoughtful presentation of a topic that has often caused confusion among epidemiologists, demographers, and biostatisticians, i.e., selection of the appropriate statistical model of the variation of vital rates. Dr Brillinger argues that, assuming Poisson birthtimes and independent lifetimes, the number of deaths and the corresponding midyear population in an open population have a bivariate Poisson distribution and not a binomial distribution. He further shows in a theorem that, when the coefficient of variation of the population size variable is small, one can use a Poisson approximation for the distribution of the mortality rate.

The implications of Brillinger's arguments are important and practical. For example, he shows that the Poisson is likely to be a better model of mortality rates in open populations than the commonly employed binomial distribution. He suggests that the binomial distribution will probably tend to systematically underestimate the variance of mortality rates. He suggests that the ultimate test of which model is superior, the Poisson or binomial, is empirical. We can report, consistent with the examples in this paper, that in many empirical mortality analyses, the variance of the distribution of rates over areal or other population replicates tends to be at least Poisson—often we find evidence of super-Poisson variability (e.g., Manton, Woodbury, and Stallard, 1981; Manton, Stallard, and Vaupel, 1986). Indeed,

it is our concern that perhaps Brillinger does not stress strongly enough the importance of making empirical tests of the variance of the vital statistics rates as recommended by Collings and Margolin (1985).

Brillinger argues that certain critical assumptions of the binomial are not very realistic for the case of an open population (i.e., for persons entering and leaving the population that individual exposures are the same and that the realizations of individual life histories are the same). He also shows that formulas due to Chiang [e.g., Brillinger's equation (2.2)] constructed to create a more valid variance estimate by constructing a hypothetical cohort will probably not perform as well as the conceptually and computationally simpler Poisson model. Finally, he deals directly with the misconception of many population scientists that since death counts and rates are population values, they are "exact values" and therefore not stochastic.

Although the paper deals with important issues, several points are raised in the discussion that deserved closer attention. First, Brillinger suggests that when the Poisson is presented as a model for mortality rates it "seems always" to be based on the rationalization that it is for a situation where deaths are rare and, consequently, on the assumption that the Poisson is a good "approximation" to the binomial. This assumption is not "always" made by demographers. In Manton and Stallard (1984, pp. 285–287), we show that the Poisson is an excellent approximation to the variance of the rate estimate even when the rate is large (e.g., .30) when the denominator is based on person-years of exposure. Furthermore, it can be shown that, for rates based on midyear population counts, the variance of the rate estimate is at least Poisson with negative binomial variance to midyear and binomial thereafter. Specifically, if the force of mortality for persons age t is λ for $a \leqslant t < a + 1$ and $P = e^{-\lambda/2}$, the number of deaths during the second half of the year, d_2, is binomially distributed with parameters N and Q, where N is the midyear population and $Q = 1 - P$. The distribution of deaths in the first half of the year, d_1, is negative binomial with the same parameters N and P. Since d_1 and d_2 are assumed independent, the total number of deaths d $(= d_1 + d_2)$ is distributed as the convolution of the two distributions or

$$\Pr[d \mid N, \lambda] = [\exp(-N\lambda)/d!][N\lambda + O(d^3/N^2)]^d[1 + O(d^3/N^2)].$$

(See unpublished paper of Manton et al. presented at International Symposium on Small Area Estimation, Statistics Canada, Ottawa, 1985.)

One can see that the accuracy of the approximation to the Poisson is of the order of the term d^3/N^2. The expression for the variance of the death count is shown by Manton et al. (in work as yet unpublished) to be

$$\begin{aligned}\operatorname{var}(d) &= Ne^{\lambda/2}(1 - e^{-\lambda})(e^{-\lambda} + e^{\lambda/2})/(1 + e^{-\lambda/2}) \\ &\approx \mathrm{E}(d)(1 + \lambda^2/4).\end{aligned}$$

A further potentially useful extension of Brillinger's results would have been to the case where cohort as well as period and age effects are represented. This gives way to an age–period–cohort model of mortality producing a Poisson field with parameters (discussed in work as yet unpublished by Woodbury, Manton, and Blazer)

$$\begin{aligned}\lambda(a, t) = \Lambda(t - a)\exp&\left\{-\beta(t - a)\int_0^a \alpha(u)\gamma(u + t - a)\,du\right\} \\ &\times \alpha(a)\beta(t - a)\gamma(t).\end{aligned}$$

This expression replaces Brillinger's equation (3.7).

A second area where further discussion would have been useful is in the area of extra-Poisson variation. We have found that in most of our empirical studies of population replicates for mortality data there is considerable extra-Poisson variability. Three issues

that could have been more completely discussed with respect to such extra-Poisson variability are (i) the robustness of parameter estimates in models which describe this extra-Poisson variability as a function of unobserved covariates; (ii) the role of empirical Bayes procedures in assessing the effect of extra-Poisson variability; and (iii) the implications of failing to model this extra-variability (e.g., Collings and Margolin, 1985).

Brillinger comments that "of course, the estimates will depend critically on the assumed distribution for u," i.e., on the distribution of the unobserved variable. This is an unfortunate assertion that is made without proof or a citation. The issue of the robustness of estimates for models with extra-Poisson variability has been raised by Heckman and Singer (1984), for example, in the case of mixture models for lifetime data. Actually, the Heckman and Singer papers do not deal specifically with extra-Poisson variability though they do deal with the effects of unobserved covariates on hazard functions—the topic raised in Section 6.2.

In hazard models for heterogeneous populations one must postulate (i) a class of "mixing" distributions to represent the distribution of unobserved covariates affecting the hazard function, and (ii) the form of the hazard function for the individual. Heckman and Singer did find sensitivity to the parametric specification of the underlying mixing distribution in the specific data on unemployment durations that they were analyzing. The result of the analysis of that specific data set, however, has been overgeneralized to suggest that the estimation of hazard models adjusted for unobserved covariates will be unstable in any data set (Trussell and Richards, 1985; Hobcraft and Murphy, 1986). There are several reasons to question such an assumption—and Brillinger's apparently derivative statement.

First, there are questions about the use of the EM algorithm for certain hazard model specifications. Specifically, the EM algorithm does not explicitly contain the constraints on the tail behavior that were shown necessary by Heckman and Singer to achieve consistency with the Weibull hazard function in the case of a mixing distribution represented by a number of discrete mass points. Indeed, this concern is not limited to the case of EM algorithm. In general, the conditions necessary to ensure consistency of estimates are not embodied in numerical algorithms though it might be possible to introduce such constraints by appropriately penalizing the likelihood function for certain conditions.

A second more general concern is that the characteristics of the data under analysis may affect the degree of instability of the estimate—as they might in any statistical model. In the case of hazard models adjusted for unobserved covariates, it appears that the survival curve for certain types of data may manifest a near discontinuity that is the cause of instability in the maximum likelihood estimation. Examples of problematic data are apparently the duration of unemployment data used by Heckman and Singer, where reports of duration were apparently "clumped" (i.e., there were preferences to report specific durations) and in infant mortality data where rates drop off rapidly after birth (Trussell and Richards, 1985). In contrast, Manton et al. (1986) found that, for total or cause-specific mortality at extreme ages, estimates of parameters in a mixture model for lifetime data were relatively robust to selection of either mixture distributions (e.g., inverse Gaussian vs gamma) or hazard functions (Weibull vs Gompertz) of similar types. Marini (in a paper presented at Population Association of America meetings, Boston, March 1985) also found stability in analyses of marital duration. In work as yet unpublished, Heckman and Walker show that it is possible to discriminate between various model specifications using a χ^2-based measure of fit. Interestingly, they find that modeling unobserved heterogeneity gives better, and more stable, parameter estimates than the use of different lag functions of observables—a strategy that has been recommended as a way of avoiding the necessity of modeling unobservable heterogeneity.

It may be that one can assess the nature of the survival function describing a given set of data using something analogous to an "influence function" to determine if a particular hazard model specification will be unduly sensitive. Such a measure could be used as a

diagnostic tool to evaluate the likely performance of a particular hazard model specification in a particular data set. The blanket assertions about instability seem unfounded and have inappropriately discouraged evaluation of hazard models with unobserved variables—even when failure to deal with such models can lead to seriously biased regression coefficients and erroneous conclusions (as discussed in Heckman and Walker's unpublished paper).

The second issue to be addressed in the area of extra-Poisson variability is the role of empirical Bayes procedures. Again, in reference to the statement about the criticality of assumptions about the distribution of unobserved covariates, one can refer to the work of Morris (1983). Morris demonstrates that, when the mean and variance of the distribution of events are known (or are estimated and treated as known), an empirical Bayes model where the prior distribution and the distribution of the mortality outcomes of individuals are natural conjugate distributions will have certain optimality properties, i.e., they will "minimize the maximum expected squared error loss." This follows from the properties of the natural conjugate distributions and the assumption of a quadratic loss function and not from a justification of the physical properties of the distributions. This property implies a robustness to model misspecification for empirical Bayes models where rates are affected by unobserved variables when the model specification uses distributions that are natural conjugates.

In this context, then, more discussion of the negative binomial, which can be viewed as arising from a gamma mixed Poisson process (Manton et al., 1981) might have been useful. A failure to discuss the model representing extra-Poisson variation arising from use of the natural conjugate distribution to the Poisson distribution seems an important omission in the paper if discussion of the effects of unobserved covariates is to be undertaken. Furthermore, in the earlier spirit of the paper, it would have been useful to "correct" certain misperceptions in the literature based on misunderstanding of the stochasticity of mortality processes, to identify and stress the potential consequences of simply "ignoring" extra-Poisson variability (Collings and Margolin, 1985). That is, in analogy to the argument that rate estimates for populations are nonstochastic, the effect of extra-Poisson variation is often ignored by demographers and epidemiologists on the rationalization that though such variation is likely to exist, models that describe such additional sources of variation are not "robust." Empirical evidence, the analysis of data characteristics that jeopardize "robustness," and the empirical Bayes approach all suggest that hazard models reflecting the effect of unobserved variables are not unstable in all types of data. What has not been balanced against an appropriate concern about the robustness of the model to misspecification, which should be an issue when applying any statistical model, however, are the errors that result from retaining an incorrectly specified model. This point was made clearly by Heckman and Singer (1984), who, even when their model exhibited sensitivity to parametric assumptions, argued that a failure to include parameters representing heterogeneity could lead to badly biased results. Thus, they suggest that, when the model is sensitive, heterogeneity be adjusted for by a set of nuisance parameters. Furthermore, even if a model representing the effects of population heterogeneity were sensitive to the specification of either a mixing distribution or a hazard function, this may represent important information about the nature of the process being modeled.

A third issue that was probably beyond the scope of the current paper, but which is at least alluded to in the references, is where mortality is generated by a Poisson process conditional on covariates whose values are generated by multivariate diffusion processes. The theoretical paper by Yashin, Manton, and Vaupel (1985) could also have been supplemented by reference to the empirical application of Woodbury, Manton, and Stallard (1979) and Manton and Woodbury (1985). Andersen (1986) has identified this as currently an important area in survival analysis that has not received adequate attention. The limited discussion by Brillinger does not identify the nature of the statistical issues that are raised

by this rich class of models. For example, Yashin et al. (1985) extend the basic model of Woodbury and Manton (1977) by considering the additional effect of an unobserved covariate process on the basic hazard and diffusion processes.

Overall, the paper addressed a number of important issues on which there is considerable confusion in current statistical practice in dealing with vital event data and suggested some fundamental changes in current thinking and practices. The author is to be applauded for the elegance in which he performed these difficult tasks. It does seem that, at times, some related topics (e.g., the robustness of mortality models in heterogeneous populations) were alluded to but not fully discussed. This was unfortunate because there are important misconceptions among population scientists on many of these topics as well and, though it may not have been practical to deal with those issues in this one paper, the allusions to these issues sometimes served to confuse, rather than to clarify, certain issues.

References

Andersen, P. K. Time-dependent covariates and Markov processes. (1986). In *Modern Statistical Methods in Chronic Disease Epidemiology*, S. H. Moolgavkar and R. L. Prentice (eds), 82–103. New York: Wiley.

Collings, B. J. and Margolin, B. H. (1985). Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association* **80,** 411–418.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52,** 271–320.

Hobcraft, J. and Murphy, M. (1986). Demographic event history analysis: A selective review. *Population Index* **53,** 3–27.

Manton, K. G. and Stallard, E. (1984). *Recent Trends in Mortality Analysis.* San Francisco: Academic Press.

Manton, K. G., Stallard, E., and Vaupel, J. W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association* **81,** 635–644.

Manton, K.G. and Woodbury, M. A. (1985). A continuous-time multivariate Gaussian stochastic process model of change in discrete and continuous state variables. In *Sociological Methodology 1985*, N. Tuma (ed.), 277–315. San Francisco: Jossey-Bass.

Manton, K. G., Woodbury, M. A., and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics* **37,** 259–269.

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical Theory. *Annals of Statistics* **11,** 515–529.

Trussell, J. and Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman–Singer procedure. In *Sociological Methodology 1985*, N. Tuma (ed.), 242–276. San Francisco: Jossey-Bass.

Woodbury, M. A. and Manton, K. G. (1977). A random walk model of human mortality and aging. *Theoretical Population Biology* **11,** 37–48.

Woodbury, M. A., Manton, K. G., and Stallard, E. (1979). Longitudinal analysis of the dynamics and risk of coronary heart disease in the Framingham study. *Biometrics* **35,** 575–585.

Yashin, A. I., Manton, K. G., and Vaupel, J. W. (1985). Mortality and aging in a heterogeneous population: A stochastic process model with observed and unobserved variables. *Theoretical Population Biology* **27,** 154–175.

John W. Tukey (Princeton University, Fine Hall, Washington Road, Princeton, New Jersey 08544, U.S.A.)

My interest in this problem area stems from a request from James A. McCullough, of the California Center for Health Statistics, for some advice about a simple multiplicity allowance, applicable to the comparison of age-adjusted death rates. The querist very kindly furnished me a write-up (McCullough, Report No. 83-05070 on Data Matters, California Center for Health Statistics, 1983) providing two sets of standard errors, one tracing to Chiang (1961) and the other to Kleinman (1976, 1977). On occasion, these differed widely,

leaving one with the feeling something was wrong. (We can now see that Kleinman's formulas are reasonably close to a careful approximation.)

David Brillinger, who was a natural interpreter between this discussant and Chiang, because he combined geographical proximity to the latter and much past collaboration to the former, became involved and, fortunately, decided to work things out from first principles. To see this paper appearing with discussion heartens me very much.

The original problem The bulk of my own comments go to the essential element of the original problem: "What is an appropriate standard error for a 'period study' based on the number of deaths, in a specified population during a certain interval of calendar time, by persons of specified age—the number of deaths in a certain rectangle in the date–age plane?" I shall treat the denominator of the rate—the population size—as fixed and known, so only the variability of deaths concerns us.

While I agree with Brillinger that using Poisson variances for deaths and zero variances for populations is *the* best *simple* analysis, I would look to a different approximation if I wished to refine that simple approximation somewhat. (Looking at a better approximation *may well strengthen our resolve to use the Poisson*!)

The accompanying Figure 1 shows a more detailed version of Brillinger's Figure 5, for a period study in which the age range exceeds the calendar range. (Similar arguments lead to related results when the opposite is true.) The death-counting rectangle is divided into four parts, for which the exposures and variances (all conditional on population P) seem to me to be as follows (observed deaths in B_i are denoted D_i).

B_0 consists of deaths where those exposed to risk had no chance of being counted in P, and thus the corresponding number of deaths earns a Poisson variance, estimated by D_0.

B_1 consists of deaths to individuals either counted in P or dying in D_1. We show below the appropriate estimated variance is closely enough, $D_1(1 - 4D_1/P)$. (This is four times the binomial correction.)

B_2 consists of deaths to individuals counted in P, and thus deserves a binomial variance, naturally estimated by $D_2(1 - D_2/P)$.

B_3 consists of deaths to those escaping P, and thus deserves a Poisson variance, naturally estimated by D_3.

If the increase in death rate from the bottom to the top of the age range and the rate of growth of the population are negligible, then D_1 and D_2 have the same average value. The estimated variance for the total number of deaths is then, nearly enough,

$$D_0 + D_1 + D_2 + D_3 - \frac{D_2^2}{P} - \frac{4D_1^2}{P} = \text{Total deaths} - 5\frac{[(D_1 + D_2)/2]^2}{P} \quad \text{crudely} \approx D - \frac{5D^2}{4P},$$

where $D = D_1 + D_2 + D_3 + D_4$ is the total number of deaths. (The last form would be $\frac{5}{4}$ of an overall binomial correction.)

If a is the length of the age interval, c is the length of the calendar interval, and $a = c/2$ (so that Fig. 1 applies), then, if deaths were (numerous and) uniform over the rectangle,

$$\frac{D_1 + D_2}{D_0 + D_1 + D_2 + D_3} = \frac{ac - (c/2)^2}{ac} = 1 - \frac{c}{4a},$$

and a better approximation for the estimated variance for the total number of deaths would be

$$D - \frac{5(1 - c/4a)^2 D^2}{4P} = (\text{Total deaths}) - \frac{5}{4}\left(1 - \frac{c}{4a}\right)^2 (\text{binomial correction})$$

For the situation pictured in Figure 1, $a = 5$, $c = 1$, and the factor multiplying the binomial

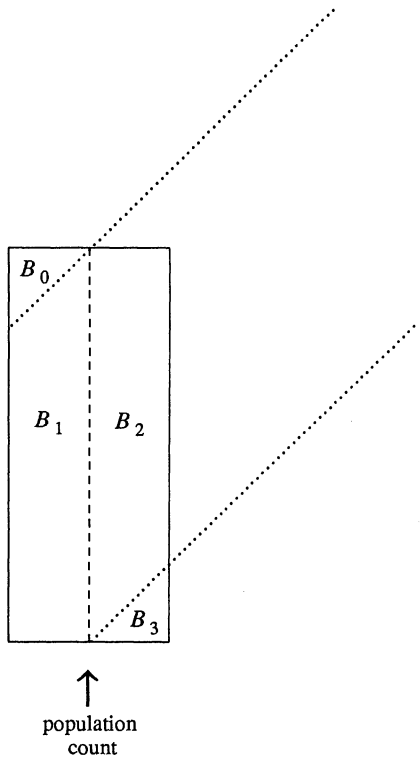


Figure 1. More detailed version of Brillinger's Figure 5, with fourfold division of the study rectangle (5-year age interval, 1-year calendar interval).

correction is $\frac{5}{4}(.9025) = 1.13$. (For $c = 2$ and $a = 5$, the factor drops to 1.01; for $c = 3$ and $a = 5$ to .90.)

If we must correct the Poisson variance, multiplying the binomial correction by

$$1 - \frac{5}{4}\left(1 - \frac{c}{2a}\right)^2 \frac{D}{P}$$

will be a *simple* and closely approximate overcorrection (so long as $c \leqslant 2a$). We may well do quite well enough by using the binomial formula.

It would clearly be worthwhile to carry out similar calculations for a variety of patterns of population counting or estimation and for more general assumptions, such as

$$\text{Force of mortality} = \mu(t, x) = \mu_0(t)e^{\phi(x-x_0)},$$

$$\text{Population size} = P_0(x)e^{\psi(t-t_0)}$$

but this is not the place for such details.

Derivation We now derive the approximation used above.

Consider the ratio of one Poisson, y, to the sum of that Poisson and a constant, E. This corresponds to those exposed to risk being those who die in B_1, y_1 in number, and those who are included in the count, E. Since we are conditioning on the count, we treat E as

fixed. We have

$$\frac{y}{y+E} = \frac{y}{E}\left(1+\frac{y}{E}\right)^{-1} = \frac{y}{E} - \left(\frac{y}{E}\right)^2 + \left(\frac{y}{E}\right)^3 \cdots$$

and

$$\operatorname{var}\left(\frac{y}{y+E}\right) = \operatorname{var}\left(\frac{y}{E}\right) - 2\operatorname{cov}\left(\frac{y}{E}, \frac{y^2}{E^2}\right) + \operatorname{var}\left(\frac{y^2}{E^2}\right) + 2\operatorname{cov}\left(\frac{y}{E}, \frac{y^3}{E^3}\right) + \cdots$$

$$= \frac{1}{E^2}\operatorname{var}(y) - \frac{2}{E^3}\operatorname{cov}(y, y^2) + \frac{1}{E^4}[\operatorname{var}(y^2) + 2\operatorname{cov}(y, y^3)] + \cdots.$$

If y is Poisson with parameter λ, then its moments about zero are λ, $\lambda^2 + \lambda$, $\lambda^3 + 3\lambda^2 + 2\lambda$, $\lambda^4 + 6\lambda^3 + 11\lambda^2 + 6\lambda$, ... so that $\operatorname{var}(y) = \lambda$, $\operatorname{cov}(y, y^2) = 2\lambda^2 + 2\lambda$, and $\operatorname{var}(y^2) + \operatorname{cov}(y, y^2) = 10\lambda^3 + 28\lambda^2 + 18\lambda$. To estimate these we naturally use $\operatorname{ave}(y) = \lambda$, $\operatorname{ave}(y^2 - y) = \lambda^2$, $\operatorname{ave}(y^3 - 3y^2 + 2y) = \lambda^3$. Substitution gives, for the estimated variance of $y/(y+E)$,

$$\frac{1}{E^2}(y) - \frac{2}{E^3}(2y^2) + \frac{1}{E^4}(10y^3 + 2y + 10y) + \cdots$$

$$= \frac{y}{E^2}\left[1 - \frac{4y}{E} + \frac{10y^2}{E^2}\left(1 - \frac{y}{6E} + \frac{y^2}{E^2}\right) + \cdots\right].$$

In most practical situations y/E will be small, and the first correction term will suffice, giving the result used above.

ACKNOWLEDGEMENT

This work was prepared in part in connection with research at Princeton University sponsored by the U.S. Army Research Office (Durham), DAAL03-86-K-0073.

REFERENCES

Chiang, C. L. (1961). Standard error of the age-adjusted death rate. *Vital Statistics—Special Reports*, Vol. 47, No. 9. Washington, D.C.: U.S. Public Health Service.

Kleinman, J. C. (1976). Infant mortality. *Statistical Notes for Health Planners*, No. 2. Washington, D.C.: National Center for Health Statistics.

Kleinman, J. C. (1977). Age-adjusted mortality indexes for small areas: Applications to health planning. *American Journal of Public Health* **67**, 834–840.

The author replied as follows:

I wish to thank each of the discussants for his thought-provoking analysis of the paper and suggestive comments. I am sure that *Biometrics* readers will find the package of paper and discussion to be a dramatic improvement over that of paper alone. The wealth of contemporary and historical references that the discussants add, does this alone. The issues elucidated by the discussants are central. My discussion of the discussion is grouped by topic.

Stochastic description We all know the advantages that flow from setting down a random mechanism for measurements and we further know how sensitive analyses can be to the choice of description made. It is not surprising therefore that various of the discussants (particularly Professors Jagers and Kish) focus on this aspect. Professor Kish addresses the case of data gathered in a sample survey (as vital statistics often are) and makes an essential distinction between the frame of the survey and the inferential population. In this paper it

is the latter that is of concern. The randomization is not under the statistician's control and he is left with the question of what population are the data representative of and what other data values might plausibly have been obtained. Stochastic description is an invaluable crutch for seizing hold of countless problems, but it is an assumption that has to be introduced and one that I do not believe statisticians are honor bound to.

Poisson births Professors Hoem and Jagers attack the Poisson assumption for the birth process for the case of a population with reproduction, i.e., with births dependent on births of the previous generation. I do not argue with this and indeed quote Kendall (1948) as evidence. I make two comments, however. First, the Poisson may still show itself empirically as a reasonable description. Second, what is really needed for the Poisson distribution of deaths in an age-period rectangle, is that the points of initial intersection of the lifelines with the rectangle (i.e., the ages at entry or dates of entry) be Poisson. (That the deaths in the rectangle are Poisson then follows from the same invariance of the Poisson under random translations as before.) Now, despite the obvious serial dependence in the population, the Poisson may prove a reasonable empirical approximation for the distribution of these points, particularly when the rectangle is not too large. We suggest that this result addresses Professor Chiang's wish not to be concerned with what has taken place before an individual arrives in the rectangle of concern and also provides a partial answer to Professor Breslow's query as to when the Poisson remains reasonable for an open population with emigration, immigration, and loss. It should when these are occurring randomly. In other cases one may need to introduce a further (error) variate.

Theorem 2 The essence of Theorem 2 is that in some circumstances one may act as if the population total, P, is fixed. This is meant to be the crudest approximation and to show the appearance of historically employed results. For situations in which more detail is required one needs to employ the bivariate Poisson, or a bivariate Poisson with extra-variation, or to use another denominator, or to develop a result of a different character. Professor Tukey commences the development of what could be the next approximation. Professor Breslow focuses on for just what ranges of values of the parameters is the approximation of Theorem 2 useful. I had in mind moderate to large λ. As stated just above it is the crudest approximation (but as the discussants mention, often employed).

Extra-Poisson variation In practice, perhaps because of omitted variables, variation beyond the Poisson may be present. In the paper such variation is handled by modelling and inferences are then based on the likelihood function. Professor Breslow describes a "method-of-moment/quasilikelihood"-type approach that requires only a modelling of the variance. Being based on weaker assumptions, this approach will be more broadly applicable. However, the likelihood approach may be anticipated to be more efficient when the assumed distribution is reasonable. Further, it may well be close to a physical description of the situation and so allow natural interpretation of the parameters appearing. These last remarks simply parallel corresponding remarks for the ordinary-linear-regression/generalized-least-squares case. Study of the robustness of the likelihood approach is needed. Finally, the likelihood approach has an elementary extension to the case of dependent responses.

Likelihood analysis Professor Keiding propounds taking a likelihood approach to the problem. Professor Breslow also mentions it. This is an attitude pioneered in the demographic context by Professor Hoem. I strongly support it. Among its advantages is the direct ability to recognize ancillary statistics on which to condition. The likelihood approach would appear particularly pertinent in the case that individual data are available, rather than the aggregate case of principal concern here.

Robust/resistant Professor Manton somewhat takes me to task for being concerned about the robustness of the methods of Sections 6 and 7 to an assumed distribution for the latent variable u. I remain concerned, though I note his listing of evidence for robustness and I remember that in Brillinger and Preisler (1983), we found that switching from the normal to the log-normal or to the gamma had little effect for the data studied. Noting what has been learned in the case of ordinary regression, however, my feeling is that it would be better to develop and employ new robust/resistant procedures in practice rather than to stick to elementary ones. I have a student at work on developing such methods.

Professor Chiang's formula I am glad that Professor Chiang has set down some further words regarding the reasoning behind his result. I note his assumption of a stationary unchanging population size. I note his indicating that the expected deaths occurring in rectangle A will be the same as the expected deaths occurring in rectangle A'. My concerns are that this will generally not be so for the actual numbers of deaths, and to develop expressions for the variability of those numbers of deaths. The fact that his variance estimate is zero for the final age group remains problematic. I am sorry that we do not end in agreement.

Concluding remarks Readers of the discussion can see that a variety of interesting problems remain to be studied—for example, taking a systems approach to the problem, as implied by the remarks of Professors Hoem and Jagers; studying influence measures as mentioned by Professor Manton; and developing approximations going beyond those recorded in Section 5. Professor Tukey commences this task in his contribution. It may be remarked that the results of this paper are quite general, with no assumptions of stationarity or constant death rates, for example. Improved approximations will come from realistic particular assumptions.