DIPARTIMENTO DI SCIENZE DEMOGRAFICHE, UNIVERSITÀ DI ROMA « LA SAPIENZA »
Via Nomentana 41, 00161 Roma, Italie. Tél. : 00 39 06 49919598, fax : 00 39 06 85303374
DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE, UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
Via G. Sanfelice, 47, 80134 Napoli, Italie. Tél. : 00 39 081 5515333, fax : 00 39 081 5512216

Séminaire international

Démographie, analyse et synthèse

Rome et Naples-Pozzuoli
"Accademia Aeronautica"
26-29 mai 1999

I-E-3

FROM HOMOGENEITY TO HETEROGENEITY

James Vaupel and Anatoli Yashin
Max Plank Institute for Demographic Research, Rostock

INED, Paris, mars 1999
133, Boulevard Davout, 75020 Paris. Tel. : 33-1-56 06 20 00, fax : 33-1-56 06 21 94

Draft: JWV 19 May 1999

Unobserved Population Heterogeneity

by

James W. Vaupel and Anatoli I. Yashin

Max Planck Institute for Demographic Research

1. Population Heterogeneity

All populations are heterogeneous. In demographic analyses, two dimensions of individual differences—age and sex—generally are observed. Many other characteristics may be observed, including date and place of birth, urban vs. rural residence, marital status, nationality, religious affiliation, number of children, number of siblings, age of mother and father at an individual's birth, household structure, socio-economic status, educational achievement, occupation, spouse's occupation, smoking behavior, diet, height, environmental quality at current residence, health status, cognitive and physical functioning, genotype, etc. In even the most thorough study, however, most attributes of individuals are not measured. Indeed, most studies focus on only a handful or two of the multitudinous dimensions of differences that distinguish one individual from another.

Observed heterogeneity creates various analytical opportunities for demographers. Multiple regression analysis, logit and probit analysis, survival analysis, and other statistical methods have been developed to estimate the impact of observed covariates. These methods are treated in standard textbooks and will not be reviewed here.

Unobserved heterogeneity creates analytical problems rather than analytical opportunities: unobserved heterogeneity is a nuisance, a headache, a bête noire. Unobserved heterogeneity creates difficulties for demographers because demographers study how population characteristics change over age and time and place—and unobserved heterogeneity distorts observed patterns of change.

2. Compositional Change

The root of the problem is that the members of population cohorts gradually die off or drop out. Animals and plants die, machines break down, bachelors marry, the married divorce, the childless give birth, those with one child have a second, children leave parental homes, students complete their education, the unemployed find jobs, the well get sick and the ill recover. Much of demographic analysis focuses on the transition rates associated with such changes. In many instances, demographers are interested in how transition rates vary with age: they study, for instance, age-specific death rates and marriage rates. In other cases, duration matters, as in studies of recovery rates from an illness or divorce rates as a function of the duration of a marriage. In analyses of first, second, and subsequent births, birth rates by parity and time since last birth are of interest.

Hence, much of demographic analysis concerns the estimation and comparison of drop-out rates in cohorts that are changing because their members are dropping out. The problem is that those who drop out probably have a greater tendency to drop out than those who do not. People who die at some age tend to be frailer or more susceptible or at higher risk than those who survive to an older age. Couples who conceive after a month or two of trying may be more fecund than those who first conceive after many months. Marriages that quickly end in divorce may have been shaky marriages from the start. To put this another way, the composition or structure of a heterogeneous cohort changes as the cohort dies off. The frail or susceptible tend to die first, leaving a more robust surviving cohort.

3. Three Levels of Explanation

Age or duration-specific changes in birth, marriage, death, and other transition rates can be interpreted in three alternative ways that might be called level-0, level-1, and level-2 explanations. A level-0 explanation is that the data are erroneous. A level-1 explanation is that the observed change is produced by a corresponding change at the individual level. A level-2 explanation is that the observed change is an artifact of a change in the structure of the population, i.e., a change in the composition of a heterogeneous cohort.

Consider the report that the increase in mortality with age slows at the oldest ages (Vaupel et al. 1998; Thatcher, Kannisto, Vaupel 1998). A direct, level-1 explanation would be that for individuals at advanced ages the probability of death increases relatively gradually with age. A level-0 explanation (bad data) would be that death rates at advanced ages are distorted by age-misreporting problems and that the apparent deceleration of mortality is a consequence of age exaggeration. Finally, a level-2 explanation would be that the leveling off of death rates after age 100 might be "caused by decreases in the average frailty of a population cohort at later ages as frailer members are removed by mortality" (Vaupel, Manton, and Stallard 1979).

Observed patterns of mortality deceleration in different populations are almost certainly due to a mix of these three levels of explanation, with the importance of the different explanations differing from population to population. In almost all populations there are problems with age-misreporting at advanced ages and in many populations such misreporting is very severe (Jeune and Vaupel 1999). All populations are heterogeneous, so level-2 explanations must have some validity, although it is currently unclear how much of the deceleration can be explained by compositional change. The level-1 explanation that individuals age more slowly at advanced ages may be partially right—or completely wrong. There is some suggestive evidence that for individuals the chance of death actually rises faster than exponentially at advanced ages, even though population death rates are increasing slower than exponentially (Yashin and Iachine 1997).

At least since Edmund Halley (1693), demographers have recognized the importance of level-0 and level-2 explanations as alternatives to direct level-1 explanations. All careful demographers are aware of the prevalence of bad data and all well-trained demographers know that demographic rates can differ because of differences in population composition. Nonetheless, level-1 explanations—that what is observed on the population level also holds on the individual level—seem so natural that even careful demographers often find themselves naively and uncritically slipping into direct interpretations of population changes and differences (Vaupel and Carey 1993).

4. Frailty Models

Demographers try to distinguish between type-1 and type-2 explanations by using frailty models (Vaupel, Manton, Stallard 1979) and the statistical methods of survival analysis (Cox and Oakes 1984). In this approach the trajectory a cohort's rate of death or exit is usually captured by either the survival function $s(x)$ or the hazard function $\mu(x)$. Demographers call this hazard function the force of mortality when they are studying death rates and in some contexts the term intensity is used instead of hazard. The survival function and the hazard function are inter-related by the following two formulas:

$$\mu(x) = -\frac{ds(x)/dx}{s(x)} \tag{1a}$$

and

$$s(x) = e^{-\int_0^x \mu(t)dt}. \tag{1b}$$

In the simplest case there is no information about the characteristics of the individuals in the cohort except age (and whatever characteristics describe the cohort as a whole, such as "males born in France in 1948").

Because all populations are heterogeneous, it makes sense to model the population as a mix of homogeneous sub-populations (which might each consist of a single individual). Let $s(x,z)$ be the survival function for the sub-population with "frailty" z , where frailty in this context simply refers to the susceptibility or liability of the sub-population to the hazard. In general, frailty models are designed such that the greater an individual's frailty, the greater the individual's susceptibility or liability to the hazard of interest.

Let $\bar{s}(x)$ be the survival function for the population as a whole, such that

$$\bar{s}(x) = \int_o^\omega s(x,z)g(z)dz \tag{2a}$$

in the continuous case, where g(z) is the probability distribution of z at age zero and

$$\bar{s}(x) = \sum_z \pi(z)s(x,z) \tag{2b}$$

in the discrete case, where $\pi(z)$ is the proportion of the cohort in sub-population z at age zero. This general frailty model can be more specifically formulated in several ways.

a) Relative-Risk Models

One specification is the proportional-hazards or relative-risk model

$$\mu(x,z) = z\mu(x), \tag{3a}$$

or, equivalently,

$$s(x,z) = s(x)^z, \tag{3b}$$

which was suggested by Vaupel, Manton, and Stallard (1979). In this model, $\mu(x)$ is the baseline, standard, underlying hazard for individuals of frailty one and $s(x)$ is the corresponding survival function. Vaupel, Manton, and Stallard (1979) show that

$$\overline{\mu}(x) = \overline{z}(x)\mu(x), \tag{4}$$

where $\overline{z}(x)$ is the average frailty of those alive at age x. Because z is fixed and does not vary with age, $\overline{z}(x)$ declines with age as the frail drop out of the cohort. Hence, $\overline{\mu}(x)$ increases more slowly than $\mu(x)$ does. Indeed, $\overline{\mu}(x)$ can be declining even though $\mu(x)$ is rising.

For this model z is often taken to be gamma distributed with mean 1 and variance σ^2, because this gamma distribution leads to convenient mathematical relationships. In particular, for gamma-distributed frailty

$$\overline{z}(x) = (1+\sigma^2\int_o^x \mu(t)dt)^{-1} = \overline{s}(x)^{\sigma^2}, \tag{5}$$

where $\overline{s}(x)$ is the survival function for the population as a whole. It follows from (5) that

$$\overline{s}(x) = (1-\sigma^2 \ln s(x))^{-1/\sigma^2}. \tag{6}$$

As a specific example of this kind of gamma-frailty relative-risk model, suppose that mortality on the individual level follows a Gompertz trajectory:

$$\mu(x) = ae^{bx}. \tag{7}$$

Then it follows from (4) and (5) that the population trajectory of mortality will follow the logistic pattern

$$\bar{\mu}(x) = \frac{ae^{bx}}{1 + \frac{a\sigma^2}{b}(e^{bx} - 1)}, \tag{8}$$

leveling off at a value of b / σ^2.

b) Accelerated-Aging Models

Another specification is the accelerated-aging model

$$\mu(x, z) = \mu(xz), \tag{9}$$

which is analogous to the accelerated-failure model used in reliability engineering. In the special case where $\mu(x)$ follows the Weibull trajectory

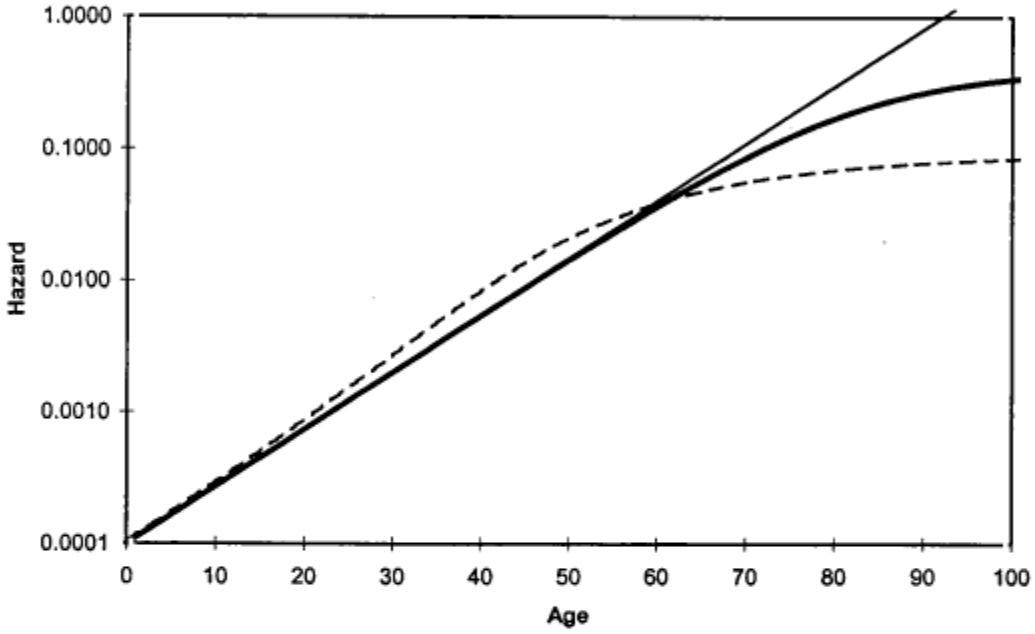
$$\mu(x) = ax^b, \tag{10}$$

where a and b are parameters, this model is equivalent to the relative-risk model, because

$$a(zx)^b = z^b ax^b = z^b \mu(x). \tag{11}$$

In the special case where $\mu(x)$ follows the Gompertz trajectory, given in formula (7), the accelerated-aging model is of the form ae^{zbx} whereas the relative-risk model is of the form zae^{bx}. Small changes in the slope parameter b can have larger effects on mortality at older ages than big changes in the level parameter a. Hence, much less heterogeneity is needed in an accelerated-aging Gompertz model than in a relative-risk Gomperz model to produce substantial differences between $\mu(x)$ and $\bar{\mu}(x)$ at older ages. This is illustrated in Figure 1.

FIGURE 1: The accelerated-aging model can produce greater mortality deceleration with less heterogeneity than the relative-risk model. Baseline Gompertz hazard with a=.0001 and b=.1 (thin line) compared with population hazard in relative-risk model with σ^2=.25 (thick line) and in accelerated-aging model with σ^2=.05 (dashed line). Note that hazards are shown on a log scale.



c) Discrete Frailty Models

The discrete frailty model is also a useful specification of the general frailty approach, as discussed by Vaupel and Yashin (1985). In this case,

$$\mu(x,z) = \mu_z(x). \tag{13}$$

That is, z is now an index for the different subpopulations, each of which has a hazard function. Let π_z be the proportion of the population in subpopulation z at age zero. Then

$$\bar{s}(x) = \sum_z \pi_z s_z(x) \tag{14}$$

and

$$\bar{\mu}(x) = \sum_z \pi_z s_z(x) \mu_z(x) / \sum_z \pi_z s_z(x). \qquad (15)$$

If it is assumed that z is a relative-risk factor, then $\mu_z(x) = z\mu(x)$ and $s_z(x) = s(x)^z$. Heckman and Singer (1984) suggested that this specification be used to control for the effects of hidden heterogeneity when fitting models to data. More generally, however, $\mu_z(x)$ can take on a different functional form for each value of z.

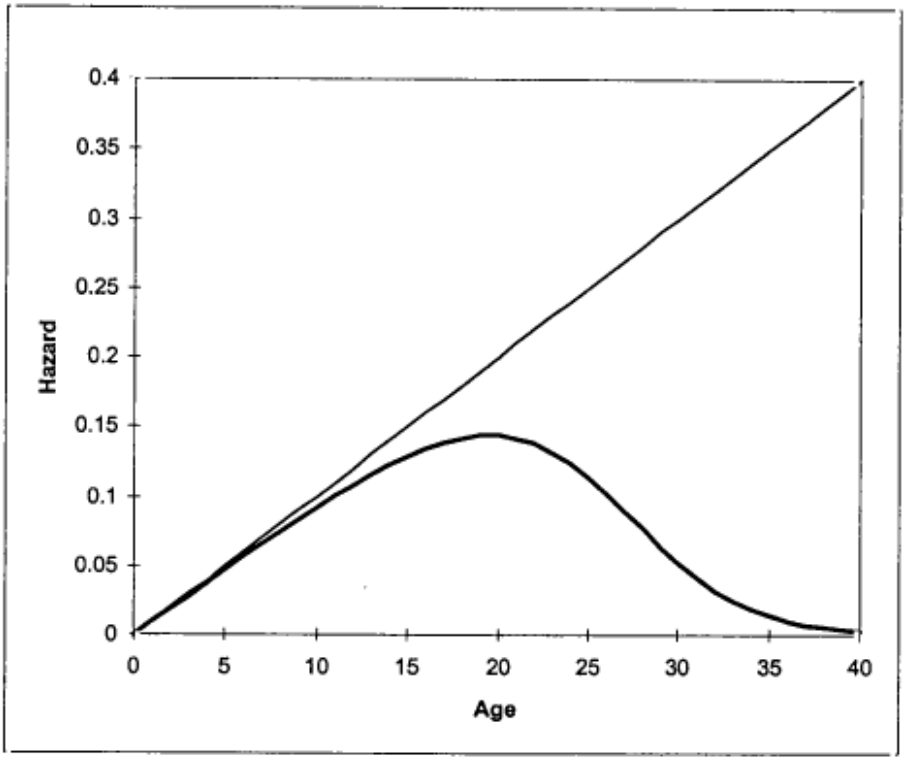
A simple example of discrete frailty models is the mover-stayer model (Blumen et al. 1955) in which one group in the population is susceptible to emigration, marriage, divorce, or some disease and the other group is immune. Let π be the proportion of the population that is susceptible. Then it follows from (15) that

$$\bar{\mu}(x) = \pi s(x) \mu(x) / (\pi s(x) + 1 - \pi). \qquad (16)$$

Even if $\mu(x)$ steadily increases, $\bar{\mu}(x)$ will eventually decline as $s(x)$ approaches zero. An illustrative example is provided in Figure 2.

Divorce rates in some countries and periods follow the kind of rising-falling pattern shown in Figure 2. Does this imply that marriages are shakiest after a few years of marriage? Not necessarily, as the Figure illustrates. The same general effect could be produced if the second group were not immune but simply at low risk. Indeed the rising-falling pattern could be produced if the hazard steadily increases for the high-risk group but steadily decreases for the low-risk group. For one group marriages strengthen with duration, while for the other, marriages weaken—despite the appearance of the curve for the entire cohort, there is no divorce hump.

FIGURE 2: The population hazard may increase and then decline if the hazard rate for one group is increasing and the other group is immune. The population hazard is shown by the thick line. The hazard for the susceptible group, shown by the thin line, is $\mu(x) = .01x$. It was assumed that 95% of the population is susceptible.



d) Changing Frailty Models

As George Box asserted, all models are wrong, but some models of useful (Box 19[illegible]). It is often useful to define an individual's frailty as fixed, at least after some age, and to classify individuals into groups depending on their frailty at that age. Alternatively, it may sometimes be useful to develop models in which an individual's frailty can change with time or age.

In one simple model of this kind, all individuals start out with frailty one. They suffer a hazard of death of $\mu_1(x)$ at age x. They also are subject to the hazard $\lambda(x)$ that their frailty will change from one to two, in which case their hazard of death changes to $\mu_2(x)$. The second state might be associated with some morbid event, such as having a heart attack or losing the ability to walk. Alternatively, the hazard of "death" could be the

hazard of divorce and the event could be having a baby. Let $s_1(x)$ denote the proportion of the cohort that is alive with frailty one at age x and let $s_2(x)$ similarly denote the proportion of the cohort that is alive with frailty two at age x. In the simplest case when the three hazards functions are constant, it is not difficult to show that

$$s_1(x) = e^{-(\mu_1+\lambda)x} \tag{17a}$$

and

$$s_1(x) = \frac{\lambda}{\mu_1 + \lambda - \mu_2}(e^{-\mu_2 x} - e^{-(\mu_1+\lambda)x}). \tag{17b}$$

The population hazard is given by

$$\overline{\mu}(x) = \frac{\lambda\mu_2 e^{(\mu_1+\lambda-\mu 2)x} + (\mu_1 - \mu_2)(\mu_1 + \lambda)}{\lambda e^{(\mu_1+\lambda-\mu_2)x} + (\mu_1 - \mu_2)}. \tag{18}$$

At age zero, $\overline{\mu}(0) = \mu_1$ and as x approaches infinity, $\overline{\mu}(x) = \mu_2$ if $\mu_2 \leq \mu_1 + \lambda$ and $\overline{\mu}(x) = \mu_1 + \lambda$ otherwise.

Le Bras (1976) and Gavrilov and Gavrilova (1991) proposed generalizations of this model. Instead of two states of frailty, suppose that frailty z can equal any non-negative integer. Initially everyone has frailty zero. People with frailty z face a hazard of death of $\mu_o + z\mu$, as well as a hazard of $\lambda_o + z\lambda$ that their frailty will change to z+1. Although μ_o, μ, λ_o and λ are constants and do not vary with age or time, the population hazard $\overline{\mu}(x)$ follows a logistic trajectory. Yashin, Vaupel, and Iachine (1994) show that this trajectory is identical to the trajectory obtained if frailty is fixed and gamma distributed and the baseline hazard is of the Gompertz form $\mu(x) = ae^{bx}$, with $\mu(x,z) = z\mu(x) + c$, where c is the constant Makeham term. Without ancillary information it is impossible to tell whether frailty is fixed or frailty is changing.

Instead of only taking on discrete values, frailty can be modeled to vary continuously. Vaupel, Yashin, and Manton (1988), for instance, develop a changing-frailty model based on a stochastic differential equation. They apply the model to clarify the interaction of debilitation, recuperation, selection, and aging. The model yields various insights about lingering mortality consequences of disasters such as wars, famines, and epidemics that may weaken the survivors. A key result is that debilitation

and selection are interdependent: debilitation that increases population heterogeneity will result in subsequent mortality selection; selection, by altering the distribution of frailty, will influence the impact of debilitating events. The basic equation of the model is

$$\mu(x,z) = \mu_o(x) + z(x)\mu^*(x), \tag{19}$$

where $\mu_o(x)$ is the baseline hazard, $\mu^*(x)$ determines the additional hazard, and $z(x)$ is the frailty of the individual at age x as given by

$$z(x) = Y^2(x), \tag{20}$$

where $Y(0)$ is normally distributed and

$$dY(x) = [a_o(x) + (a_1(x) - a_2(x))Y(x)]dx + b(x)dW(x), \tag{21}$$

where W is a Wiener process with $W(0) = 0$. The functions a_o and a_1 represent the effects of debilitation whereas a_2 represents homeostatic healing and recuperation; the function b determines the importance of the Wiener-process term.

e) Correlated-Frailty Models

Because of shared genes and a shared childhood environment, two twins may have similar frailties. More generally, relatives or people who live in the same environment may have similar frailties. As discussed by Vaupel (1991a and b), shared-frailty models can be used to analyze such situations, but a more appropriate and powerful approach involves the correlated-frailty models developed by Yashin and colleagues and explained in Yashin, Vaupel, and Iachine (1995) and Yashin and Iachine (1997).

A simple variant of this kind of model involves pairs of twins, with one twin having fixed frailty z_1 and the other twin having fixed frailty z_2 and with the hazard of mortality given by $\mu(x,z_i) = z_i\mu_o(x)$, $i = 1,2$. The correlation between the two frailties is modeled as follows. Let

$$z_1 = y_o + y_1 \tag{22a}$$

and let

$$z_2 = y_o + y_2, \tag{22b}$$

where the y_i, $i = 0,1,2$, are three independent random variables that are gamma distributed with the same scale parameter. The gamma distributions of y_1 and y_2 have the same shape parameter, but this parameter may differ for y_o. The frailties z_1 and z_2 are constrained to have a mean of one and they have the same variance σ^2. The values of σ^2 and ρ, the correlation coefficient between the two frailties, are simple functions of the scale and shape parameters.

As shown by Yashin, Vaupel and Iachine (1995), the bivariate survival function for the population of twins is given by

$$\bar{s}(x_1, x_2) = \bar{s}(x_1)^{1-\rho}\bar{s}(x_2)^{1-\rho}(\bar{s}(x_1)^{-\sigma^2} + \bar{s}(x_2)^{-\sigma^2} - 1)^{-\rho/\sigma^2}. \quad (23)$$

Because the survival of adult twins is very similar to the survival of adult singletons, in studies of adult mortality the function $\bar{s}(x)$ can be taken from demographic life tables for the general population. In this case the bivariate survival function depends only on σ^2 and ρ; no assumptions have to be made about the shape of a baseline hazard function. Using (23) values of σ^2 and ρ can be estimated using the kind of maximum-likelihood estimation described below. Then the baseline survival function can be calculated by rearranging (6) as follows:

$$s(x) = \exp((1 - \bar{s}(x)^{-\sigma^2}) / \sigma^2) \quad (24)$$

and $\mu(x)$ can be calculated from (1). Applying this model to survival data on Danish twins born between 1870 and 1900, Yashin and Iachine (1997) found that the baseline hazard of mortality increases faster than exponentially after age 30 even though the population hazard of mortality decelerates at advanced ages. That is, it is possible that the observed leveling off of mortality may be entirely accounted for by a level-2 explanation (compositional change due to mortality selection) and the actual trajectory of mortality for individuals may rise more rapidly than a Gompertz curve.

5. Empirical Data

The survival or duration data used in fitting frailty models is often of the following form. There are n individuals in some cohort, with observed ages at

death $X_i, i = 1,\ldots,n$. More generally, X_i stands for age at some event, such as marriage, or some duration, such as time from marriage to divorce. For simplicity, we will refer to X_i as age at death.

Age at death may not be known for all individuals: it may only be known that the individual survived at least until some age. These are called right-censored observations. They can arise if some individuals never "die" (e.g., some women never give birth, some people never marry, some married people never divorce). They can also arise if some individuals are still alive at the end of a study or if some individuals drop out of the study and are lost to follow-up.

It may also be the case that it is only known that individual i died between age x_1 and age x_2. And it may be the case that an individual is not followed from age zero but from some age x_o, so that the individual is only at risk of dying after age x_o. These are called left-censored or left-truncated observations.

Various covariates may be observed: we will let w_{ij} denote the value of the jth covariate for individual i.

6. Methods of Parameter Estimation

Various methods may be used to fit frailty models to empirical survival or duration data. In an important article on "deceleration in the age pattern of mortality at older ages", Horiuchi and Wilmoth (1998) estimate the parameters of a Gompertz-Makeham model with period effects and unobserved frailty by a weighted-least-squares procedure. They carefully document the method they use, so their article is of pedagogical value as well as being of substantive interest.

It is more common, however, to use maximum-likelihood methods in analyses of survival or duration data in general and in fitting frailty models in particular. For discussion of this approach, a textbook such as that by Cox and Oakes (1984) is recommended. Here we merely adumbrate a few points of particular relevance to the estimation of parameters of frailty models.

The likelihood of an observation X_i can be thought of as the probability of observing this value given a particular model with specific parameter values. More generally, the likelihood can be proportional to the probability instead of being equal to the probability, because any parameter values that maximize the probability of the data will also maximize any quantity that is proportional to the probability. Let $s_i(x)$ be the probability of surviving from age zero to age x, for some individual with a vector of covariates w_i and with some unobserved frailty z. Then if age at death X_i is observed and if the individual is followed from age zero, the likelihood of the observed age at death is $\mu_i(X_i)s_i(X_i)$. If it is only known that the individual survived at least to age X_i, then the likelihood of this observation is $s_i(X_i)$. If it is known that the individual died between ages X_{1i} and X_{2i}, then the likelihood of this datum is $s(X_{1i}) - s(X_{2i})$. If the individual is first observed at age X_{1i} and then dies at age X_{2i}, then the likelihood is $\mu(X_{2i})s(X_{2i}) / s(X_{1i})$.

It is customary in survival analysis to make calculations in terms of the logarithm of the likelihood, the log-likelihood, because the likelihood of a data set is often extremely small. Let $L(X_i)$ denote the log-likelihood of the observation. The log-likelihood of the entire data set is given by the sum of all the $L(X_i)$'s. The maximum-likelihood estimate of the parameter values in a model is the estimate that maximizes the likelihood or, equivalently, the log-likelihood of the data.

The effect of observed covariates on survival can be modeled in many ways. Because our focus here is on hidden heterogeneity and not on general methods of survival analysis, we restrict our attention to the simple case where the covariates are fixed (rather than changing over time). Furthermore, we will assume that the impact of the covariates on an individual's hazard is given by $W_i\mu_i(x,z)$, where W_i is the net relative-risk imposed by the vector of observed covariates. Often in survival analyses, W_i is modeled by $W_i = e^{\sum_j b_j w_j}$, where the b_j's are coefficients that are estimated.

For ease and conciseness of exposition we will consider only the relative-risk frailty model with gamma-distributed frailty. For the relative-risk model (cf. formula (3b)),

$$s_i(x) = s(x)^{W_i z}, \tag{25}$$

where $s(x)$ is the baseline or standard survival function for individuals of frailty 1 and with an estimated relative-risk W_i of 1. In the special case of gamma-distributed frailty, it follows from formula (6) that

$$\bar{s}_i(x) = (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma^2}, \tag{26}$$

where $\bar{s}_i(x)$ is the probability that an individual with estimated relative-risk W_i will survive to age x and where σ^2 is the variance of frailty. The bar over the s indicates that $\bar{s}_i$ is an average: unobserved frailty z is removed from the formula by taking the expected value of s with respect to z. The corresponding value of $\bar{\mu}_i(x)$ can be calculated by formula (1) and the log-likelihood L(x) can then be calculated as indicated above.

A remaining issue is how to estimate the baseline survival function $s(x)$. There are two approaches to this. First, a parametric form can be assumed. For instance, it might be assumed that $s(x)$ (and $\mu(x)$) are of the Gompertz or Weibull form. Manton, Stallard, and Vaupel (1985) provide an example of this kind of analysis.

Alternatively, $s(x)$ can be estimated nonparametrically. That is, values of $s(x)$ can be estimated for a sequence of ages over some age range without imposing any assumptions on the shape of the trajectory of $s(x)$. Several different methods of nonparametric estimation have recently been proposed and research in this area is rapidly developing. Here we sketch one method, to illustrate the general idea of nonparametric estimation.

Suppose that the survival data that are available for analysis are based on a large random survey of some population. Further suppose that survival in the population is known, perhaps from vital-statistics data. Let $\bar{\bar{s}}(x)$ be the survival curve for the population. As above, let $\bar{s}_i(x)$ be the survival function for the individual i in the survey. For a large random survey, the following equation might approximately hold:

$$\bar{\bar{s}}(x) = \sum_{i=1}^{n} s_i(x) / n. \tag{27}$$

If so, then the following method could be used. For the relative-risk gamma-frailty specification, formula (26) can be substituted in (27), yielding

$$\bar{\bar{s}}(x) = \sum_{i} (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma^2} / n. \tag{28}$$

For any specific set of estimated values for W_i and σ^2, one and only one value of $s(x)$ will satisfy this equation. Hence, σ^2 and the coefficients that determine W_i can be determined by maximum-likelihood estimation under the constraint that (28) holds.

The theoretical and practical properties of this algorithm still need to be investigated. Many other estimation procedures that do not require parametric estimation of $s(x)$ are being developed and various imputation methods, EM algorithms, and other concepts might be used (see, e.g., Andersen, Borgan, Gill, and Keiding 1992). The statistical estimation of frailty models is currently a hot topic of statistical research and the coming decade is likely to produce major advances in the development of powerful, practicable procedures.

8. Conclusion

The frailty models and parameter-estimation methods sketched above are not yet available as part of easy-to-use computer software packages. Several software packages, however, include convenient options for fitting other kinds of models to survival and duration data. In particular, Cox (partial likelihood) regression can be readily applied to empirical observations. Why bother with frailty models when Cox regression can be used to estimate the coefficients of covariates in hazards models? There are three main reasons.

First, Cox regression yields coefficient estimates that tend to be biased toward zero. As epidemiologists have often observed, most risk factors that raise the chance of death appear to become less important with age or duration. The reason generally is that high-risk individuals who survive often have unobserved strengths or advantages, whereas many of the apparently low-risk individuals who survive may be relatively weak

or unhealthy along unobserved dimensions. Consequently, at older ages or longer durations, the high-risk group differs in composition from the low-risk group: the high-risk group has lower unobserved frailty. If unobserved frailty is not included in the model, then this effect will result in a convergence with age of the hazard functions for the two groups, as discussed by Vaupel, Manton and Stallard (1979) and Vaupel and Yashin (1985). The proportional-hazards assumption used in Cox regression does not allow for such convergence: the estimated relative risk is a measure of the average relative risk over the entire age range. The implication of this is that Cox regression tends to result in under-estimates of risk factors: the estimates are biased toward zero. More generally, any method that ignores hidden heterogeneity will tend to under-estimate risk factors at older ages or longer durations.

Second, frailty models permit estimation of underlying (or baseline) hazards, i.e., the hazards that govern the trajectory of risks at the individual level. It may be of interest, for instance, to know whether or not the underlying hazard is monotonically increasing even though the observed population hazard first rises and then declines. More generally, demographers are concerned about whether observed trajectories of demographic rates over age or duration can be explained by level-1 accounts or level-2 accounts. Does the trajectory observed for a population also hold for the individuals who comprise the population—or is the trajectory attributable to compositional change? Frailty models are designed to address this question.

Third, frailty models permit the use of ancillary vital-statistics data in the analysis. For example, as briefly discussed above, it is possible to analyze detailed data on some subset of a population (e.g., twins or those who participate in a survey) together with the vital-statistics data on the survival of the general population. The combination of detailed data on a sub-population with survival data on the entire population leads to more accurate statistical estimates. This combination seems natural and highly appropriate for demographers.

References

Andersen, P. K., Borgan, O., Gill, R. D. and, Keyding, N. (1992) *Statistical methods based on counting processes.* London, New York, Springer Verlag.

Blumen, I., M. Kogan, and P.J. McCarthy (1955), *The Industrial Mobility of Labor as a Probability Process*, Ithaca, New York: Cornell University Press.

Box. G.E.P. (1976), "Science and Statistics", **Journal of the American Statistical Association** 71:791-802.

Cox, D.R. and D. Oakes (1984), *Analysis of Survival Data,* London: Chapman & Hall.

Gavrilov, L.A. and N.S. Gavrilova (1991), *The Biology of Life Span*, New York: Harwood.

Halley, E (1693) reprinted in D. Smith and N. Keyfitz (eds.) *Mathematical Demography*, New York: Springer 1977.

Heckman, J.J. and B. Singer (1984) "Econometric Duration Analysis", **Journal of Econometrics** 24:63-132.

Horiuchi, S. and J.R. Wilmoth (1998), "Deceleration in the Age Pattern of Mortality at Older Ages", **Demography** 35:391-412.

Jeune, B. and J.W. Vaupel, eds. (1999), *Validation of Exceptional Longevity*, Odense, Denmark: Odense University Press.

Le Bras, H. (1976), "Lois de Mortalite et Age Limite", **Population** 31:655-92.

Thatcher, A.R., V.Kannisto, and J.W. Vaupel (1998), *The Force of Mortality at Ages 80 to 120*, Odense, Denmark: Odense University Press.

Manton, K.G., E. Stallard, J.W. Vaupel (1986), "Alternative Models for the Heterogeneity of Mortality Risks among the Aged", **Journal of the American Statistical Association** 81:635-44.

Vaupel, J.W. (1991a), "Kindred Lifetimes: Frailty Models in Population Genetics" in J. Adams et al. (eds.) *Convergent Questions in Genetics and Demography*, London: Oxford University Press.

Vaupel, J.W. (1991b), "Relatives' Risks", *Theoretical Population Biology* 37:220-234.

Vaupel, J.W. and J.R. Carey (1993), "Compositional Interpretations of Medfly Mortality", **Science** 260:1666-67.

Vaupel, J.W. et al. (1998) "Biodemographic Trajectories of Longevity", **Science** 280:855-860.

Vaupel, J.W. K.G. Manton, and E. Stallard (1979), "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality", **Demography** 16:439-454.

Vaupel, J.W. and A.I. Yashin (1985), "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics", **American Statistician** 39:176-185.

Vaupel, J.W., A.I. Yashin, and K.G. Manton (1988), "Debilitation's Aftermath", **Mathematical Population Studies** 1:21-48.

Yashin, A.I. and I.A. Iachine (1997), "How Frailty Models Can Be Used for Evaluating Longevity Limits", **Demography** 34:31-48.

Yashin, A.I., J.W. Vaupel, and I.A. Iachine (1994), "A Duality in Aging: The Equivalence of Mortality Models Based on Radically Different Concepts", **Mechanisms of Ageing and Development** 74:1-14.

Yashin, A.I., J.W. Vaupel, and I.A. Iachine (1995), "Correlated Individual Frailty", **Mathematical Population Studies** 5:1-15.