Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics

JAMES W. VAUPEL and ANATOLI I. YASHIN*

As a cohort of people, animals, or machines ages, the individuals at highest risk tend to die or exit first. This differential selection can produce patterns of mortality for the population as a whole that are surprisingly different from the patterns for subpopulations or individuals. Naive acceptance of observed population patterns may lead to erroneous policy recommendations if an intervention depends on the response of individuals. Furthermore, because patterns at the individual level may be simpler than composite population patterns, both theoretical and empirical research may be unnecessarily complicated by failure to recognize the effects of heterogeneity.

KEY WORDS: Survival analysis; Unobserved heterogeneity; Mixed populations; Hazard rates; Mortality; Failure; Mixtures of distributions.

The members of many kinds of populations gradually die off or drop out. Animals and plants die, bachelors marry, machines break down, the childless give birth, the unemployed find jobs. A cohort's rate of death or exit is often measured by the so-called force of mortality or hazard rate, μ. At age x and time y,

$$\mu(x, y) = -[dp(x, y)/dx]/p(x, y), \qquad y = y_0 + x, \quad (1)$$

where $p(x, y)$ is the proportion of the cohort born x years ago that is surviving at time y and y_0 is the year the cohort was born. In a homogeneous population, all individuals of age x in year y face the same hazard rate $\mu(x, y)$. A heterogeneous population consists of various homogeneous subpopulations.

That the patterns of mortality (or exit) in a heterogeneous population can differ qualitatively from the patterns of mortality in the constituent subpopulations can be neatly illustrated in the simplest example of a heterogeneous population—namely, a composite population that consists of two homogeneous subpopulations. Indeed, almost all of the distinctive features of heterogeneous populations become apparent as soon as the transition is made from a homogeneous population to a mixed population with two major subpopulations. Important research on such mixed populations includes Blumen et al.'s (1955) pioneering work on mover–stayer models of labor mobility, Shepard and Zeckhauser's (1980) health-care research, and Keyfitz and Littman's (1980) analysis of mortality. Other applications abound in fields as disparate as reliability engineering and econometrics; a statistician might broadly classify this research as falling within the intersection of survival analysis and studies of mixtures of distributions.

THE DEVIOUS DYNAMICS OF AGING COHORTS

Consider first the dynamics of mortality among a cohort of aging individuals. Age here could represent time since marriage or since release from prison, and death could be interpreted metaphorically as divorce or recidivism. Let $\mu_1(x)$ and $\mu_2(x)$ be the hazard rates for the two subcohorts at age x and let $\bar{\mu}(x)$ be the observed hazard rate for the entire cohort. (Since age and time advance synchronously for a cohort, it is not necessary to explicitly consider time y in addition to age x; for simplicity, we suppress the argument y). The key question of interest is, How does the trajectory of $\bar{\mu}(x)$ compare with the trajectories of $\mu_1(x)$ and $\mu_2(x)$?

Let $p_1(x)$ and $p_2(x)$ be the survival functions of the two subcohorts:

$$p_i(x) = \exp\left[-\int_0^x \mu_i(t)\,dt\right], \qquad i = 1, 2. \quad (2)$$

Define $\pi(x)$ as the proportion of the surviving cohort at age x that is in the first subcohort:

$$\pi(x) = \frac{\pi(0)p_1(x)}{\{\pi(0)p_1(x) + [1 - \pi(0)]p_2(x)\}}. \quad (3)$$

Clearly,

$$\bar{\mu}(x) = \pi(x)\mu_1(x) + [1 - \pi(x)]\mu_2(x). \quad (4)$$

The dependency of the cohort hazard rate on the subcohorts' hazard rates is thus mediated by the changing proportion of the population that is in one or the other of the subcohorts. Over time, the observed hazard rate will approach the hazard rate of the more robust subcohort. Figures 1 through 5 illustrate some specific instances.

The recidivism rate for convicts released from prison declines with time since release (Harris et al. 1981). The recidivism rate for former smokers who are trying to stop smoking and for former alcoholics who are trying to stop drinking also declines with time. Does this imply that the hazard of recidivism for individual convicts, smokers, and alcoholics declines over time? Not necessarily. As illustrated in Figure 1 there might be two groups of individuals, the reformed and the incorrigible. For individuals in each group, the hazard of recidivism might be constant. The observed decline would be an artifact of heterogeneity, a ruse.

*James W. Vaupel is Research Scholar, International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria and Professor of Public Affairs and Planning, Humphrey Institute, University of Minnesota, Minneapolis, MN 55455. Anatoli I. Yashin is Research Scholar, International Institute for Applied Systems Analysis, and Senior Researcher, Institute for Control Science, 65 Profsojusnaja St., Moscow, USSR. The authors thank Philip Cook, Dianne Goodwin, Nathan Keyfitz, Eduard Loeser, Andrei Rogers, Thomas Schelling, Michael Stoto, and Bodil Vaupel for their insights and suggestions and Martina JoestlSegalla, Susan Stock, and Yasmin Taher for their help in preparing the manuscript.

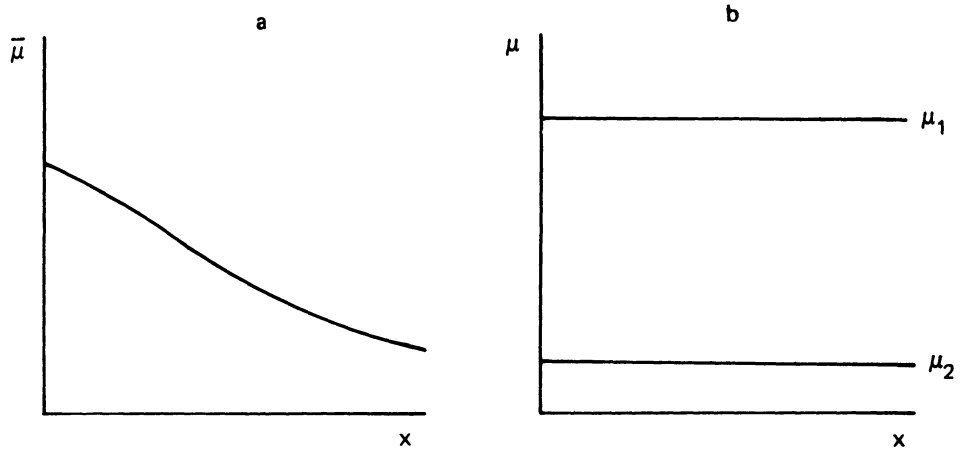

Figure 1. The observed hazard rate may decline even though the hazard rates for the two subcohorts are constant. The curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1 = .06$, $\mu_2 = .01$, and $\pi(0) = .8$. The curves are shown for values of x from 0 to 75.

As another example of the same kind of phenomenon, consider tooth decay. New caries tend to become less frequent with age. Does this mean that adults brush their teeth more carefully than children? Not necessarily. Various areas on the surface of teeth may simply differ in susceptibility to decay.

Over the course of the last century, it has taken more and more effort (as measured by cost or by feet drilled) to discover a specified amount of oil. Are geologists becoming more incompetent? Some insight can be gained by drawing an analogy between discovery rates per unit of effort and mortality rates per unit of time. It seems likely that the oil that is easiest to find and that is contained in the biggest fields tends to be found (i.e., "die") first. Even if geologists were steadily becoming more and more expert, this selection effect could outpace their growing knowledge and make it increasingly difficult for them to discover oil.

The cohort hazard rate shown in Figure 2a follows the "bathtub" shape familiar to reliability engineers (see, e.g., Gnedenko et al. 1969; Mann et al. 1974; or Barlow and Proschan 1975) and reminiscent of some human and animal mortality curves. Does this cohort curve imply that the failure rate for a specific device decreases during the infant mortality phase, is roughly constant during the useful life phase, and increases during the wear-out phase? Not necessarily. The high initial rate of breakdown could be due to a group of lemons. Note that if the population were only observed for a short time, then the cohort curve would be steadily decreasing even though none of the devices or individuals in the population would be experiencing a declining risk of mortality.

Figure 3 depicts another ruse: the observed hazard rate increases steadily, suddenly declines, and then starts increasing again, albeit at a slower rate. This trajectory is produced by two subcohorts that suffer constantly increasing hazard rates. The sudden decline in the observed hazard rate is produced by the rapid extinction of the frailer subcohort. Until the point of decline, the frailer subcohort experiences death rates that are relatively low. Then, due to the exponential increase in the force of mortality, the death rates become sufficiently large that within a few years almost all of the frailer subcohort dies. The observed hazard rate declines to the level of the hazard rate for the more robust subcohort. Since this hazard rate is increasing, the observed hazard rate then starts to increase as well: the observed hazard rate now equals the hazard rate for the more robust subcohort because only members of the more robust subcohort are still alive.

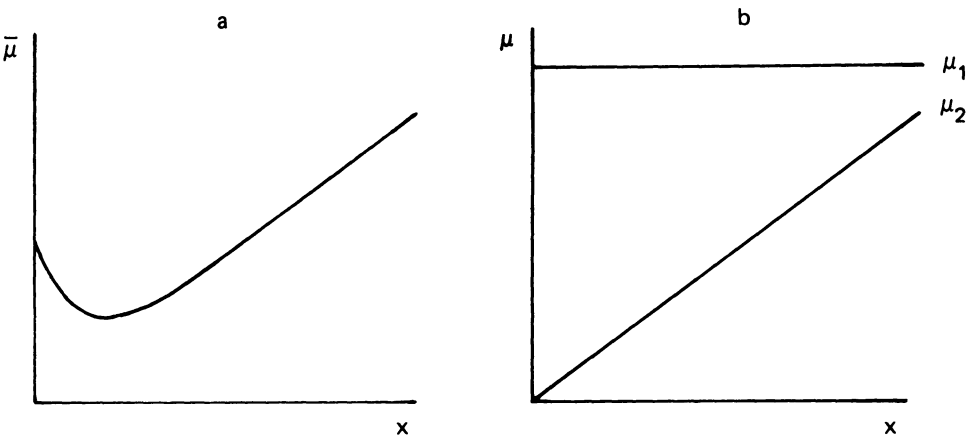

Figure 2. The observed hazard rate may decline and then rise even though the hazard rate for one subcohort is rising steadily and that of the other is constant. The curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1 = .14$, $\mu_2(x) = .001 + .0015x$, and $\pi(0) = .5$. The curves are shown for values of x from 0 to 75.

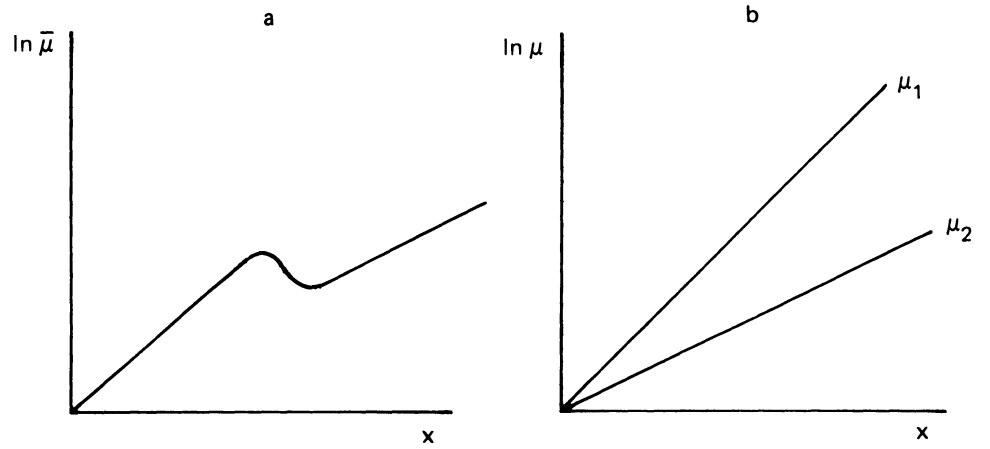

Figure 3. The observed hazard rate may rise steadily, then decline, and then rise again even though the hazard rates for the two subcohorts are steadily rising. The curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1(x) = .0001 \cdot \exp(.2x)$, $\mu_2(x) = .0001 \cdot \exp(.1x)$, and $\pi(0) = .5$. The curves are shown for values of x from 0 to 75. Note that $\bar{\mu}$ and μ_i are plotted on logarithmic scales.

Figure 4 depicts a somewhat subtler ruse: the observed cohort hazard rate increases more slowly than the hazard rates for individuals in either subcohort. Individuals are, in a sense, aging more rapidly than the cohort data show. Vaupel et al. (1979), Vaupel and Yashin (1983), and Horiuchi and Coale (1983) explored various demographic implications of this effect.

In the so-called mover–stayer model (Blumen et al. 1955), one group in the population is susceptible to emigration, marriage, divorce, some disease and so forth, and the other group is immune. If the hazard for the susceptible subcohort is steadily increasing, then as shown in Figure 5a the observed hazard for the entire population may rise and then fall. Divorce rates, for instance, follow this general rising–falling pattern (Rogers 1982). Does this imply that marriages are shakiest after a few years of marriage? Not necessarily, as Figure 5 illustrates. The same basic effect can be produced even if one group is not immune but simply at low risk. Indeed the rising–falling pattern can 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 cohort curve, there is no "seven-year itch."

In the five examples illustrated by Figures 1 through 5, the focus is on the deviation of the trajectory of the observed hazard rate from those of the hazard rates for individuals in the two subcohorts. Similar ruses may hold for any characteristic of an individual that is correlated with an individual's hazard rate.

For instance, suppose that individuals of some animal species (fluke, say, or perhaps red herring) are either lean or fat. Suppose that the fat individuals suffer a higher mortality rate. Observations indicate that the average weight of 3-year-olds is about the same as that of 4-year-olds. Does that mean that individual members of the species do not gain any weight between the age of 3 and 4? Not necessarily—each individual may be gaining weight, but selection of the fatter individuals may hold the average weight of the surviving individuals approximately constant.

As another example, imagine an anthropologist who is observing a food market where sellers bargain with potential customers. The anthropologist discovers the price of tomatoes to be steadily falling over the course of the day.

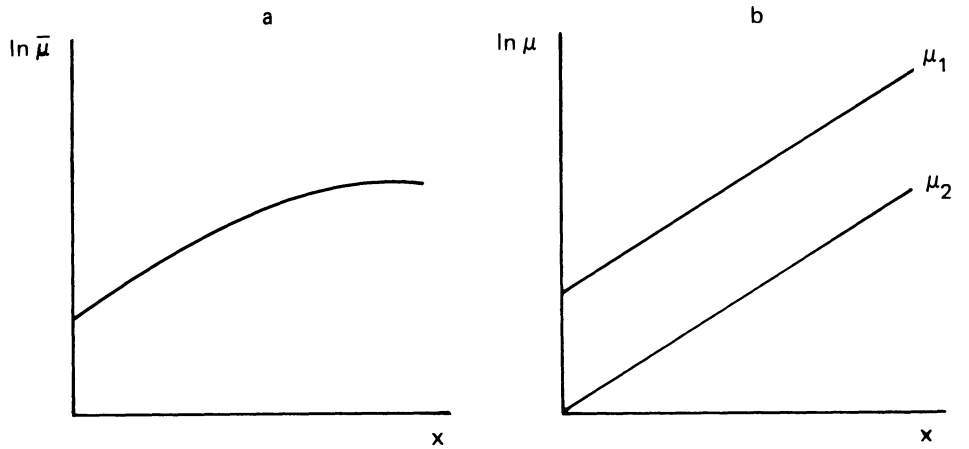

Figure 4. The observed hazard rate may increase more slowly than the hazard rates for the two subcohorts. The curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1(x) = .01 \cdot \exp(.04x)$, $\mu_2(x) = .002 \cdot \exp(.04x)$, and $\pi(0) = .8$. The curves are shown for values of x from 0 to 75; if the curve in a were continued for higher values of x, it would begin rising again and would asymptotically approach μ_2. Note that $\bar{\mu}$ and μ_i are plotted on logarithmic scales.

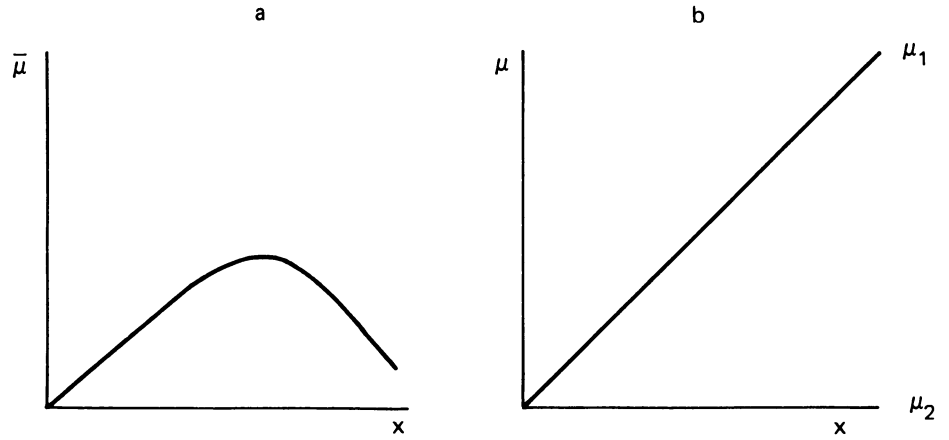

Figure 5. The observed hazard rate may increase and then decline if the hazard rate for one subcohort is increasing and the other subcohort is immune. The curve for μ was calculated from (2), (3), and (4) using $\mu_1(x) = .002x$, $\mu_2(x) = 0$, and $\pi(0) = .95$. The curves are shown for values of x from 0 to 75.

The initial hypothesis is that tomatoes deteriorate rapidly, but by studying a few selected tomatoes, the anthropologist discovers that tomatoes do not lose much flavor or texture from hour to hour or even from one day to the next. What is happening is that the best tomatoes get sold (i.e., "die") first; as the day goes on, the remaining tomatoes tend to be the most inferior ones.

MORTALITY CROSSOVERS

Figure 6a depicts a so-called mortality crossover. One subcohort's hazard rate is lower than the other subcohort's at younger ages, but higher at advanced ages. Numerous such crossovers have been discovered in comparisons of different national populations and of the same national population at different points in time (Nam et al. 1978; Manton et al. 1981); the effect also occurs for U.S. blacks vs. whites (Manton and Stallard 1981a). Some of these crossovers may be due to incorrect reporting of the age of death; others may be due to differences in life-style or other factors. Some of the crossovers may also be, at least in part, artifacts of heterogeneity.

In particular, the cohort curves in Figure 6a might be produced by the subcohort curves shown in Figure 6b. The robust subcohorts of each of the two populations face the same mortality chances. The frail subcohort of the disadvantaged population, however, faces higher mortality chances than the frail subcohort of the advantaged population. Consequently, the frailer members of the disadvantaged population die off relatively quickly, leaving a surviving population that largely consists of the robust subcohort. If this selection effect is strong enough, a crossover may be observed for the two populations (Vaupel et al. 1979; Vaupel and Yashin 1983). A crossover can also be produced if the frail and robust subcohorts of both populations experience the same death rates, but the disadvantaged population has, at birth, a larger proportion of frail individuals.

The relative prevalence of various diseases changes with age. Cancer, for example, is more common than heart failure at younger ages but less common at older ages. Does this imply that any particular individual is more likely to die from cancer in youth and from heart disease in old age? Not necessarily, as illustrated by Figure 7. A simple model (that can readily be made more realistic) might assume that everyone faces the same hazard of heart failure, but that people differ in their susceptibility to cancer. In Figure 7b, the top line gives the hazard rate for individuals at high risk of getting cancer, the bottom line gives the corresponding

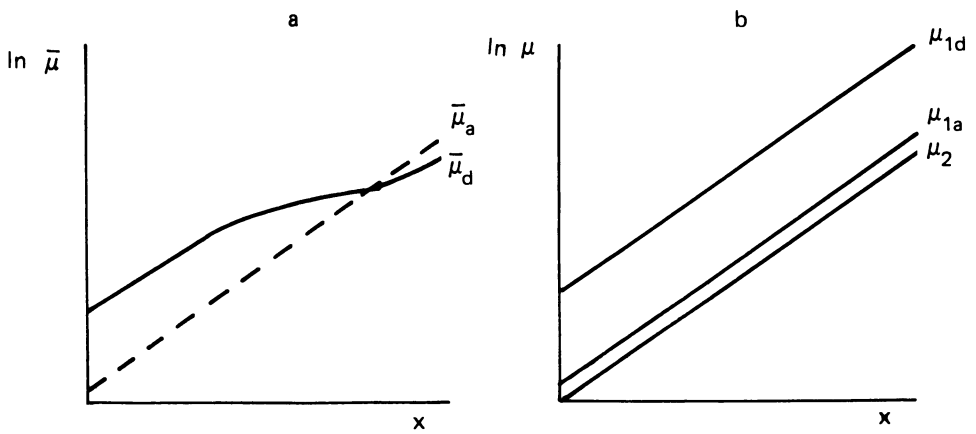

Figure 6. A disadvantaged cohort may appear to suffer lower mortality rates than an advantaged cohort at older ages. The curves for μ_a and μ_d were calculated from (2), (3), and (4) using $\mu_{1a}(x) = .0025 \exp(.04x)$, $\mu_{1d}(x) = .01 \exp(.04x)$, $\mu_2(x) = .002 \exp(.04x)$, and $\pi(0) = .7$. The curves are shown for values for x from 0 to 96. Note that the curves are plotted on logarithmic scales.

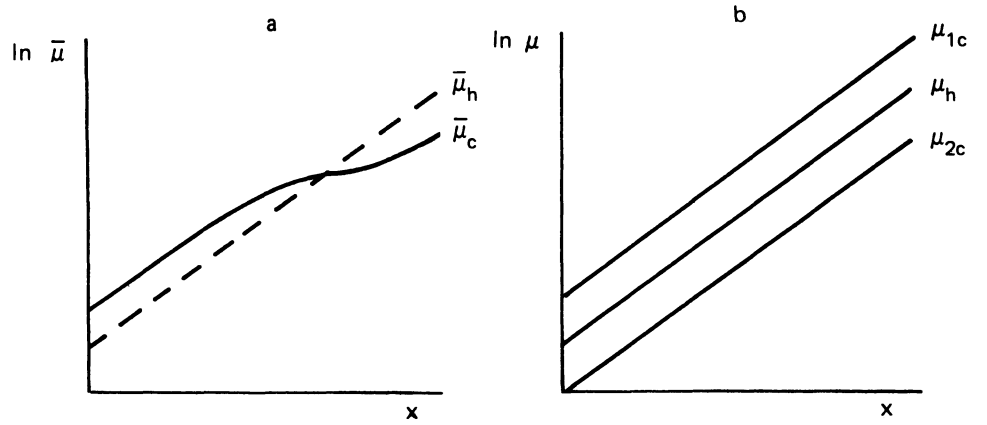

Figure 7. Observed mortality rates for two causes of death may appear to intersect. The curve for μ_c was calculated from (2), (3), and (4) using $\mu_{1c} = .01 \exp(.04x)$, $\mu_{2c} = .0025 \exp(.04x)$, and $\pi(0) = .8$. The curve for $\bar{\mu}_h$ is given by $\bar{\mu}_h(x) = \mu_h(x) = .005 \exp(.04x)$. The curves are shown for values of x from 0 to 96. Note that the curves are plotted on logarithmic scales.

hazard rate for individuals at low risk of getting cancer, and the middle line gives the hazard rate for heart failure. These hazard lines produce the apparent crossover in mortality rates shown in Figure 7a. Essentially, the incidence of cancer declines relative to the incidence of heart failure because the individuals most susceptible to cancer have died.

REDUNDANCY AND THE DEATH OF FAMILIES

Suppose that a machine or device will fail if some specific component fails. To guard against this, a component is installed in parallel to the original component so that the machine will run if either component is operating; the failure rates of the two components are independent. Will the failure rate of the machine be reduced at all ages? Not necessarily. If the two components are heterogeneous in that the second component is somewhat less reliable than the original component, then, as Barlow and Proschan (1975) have shown, the failure rate of the redundant system will, after some age, exceed the failure rate of the original, single-component system. Furthermore, as shown by the solid curve in Figure 8,

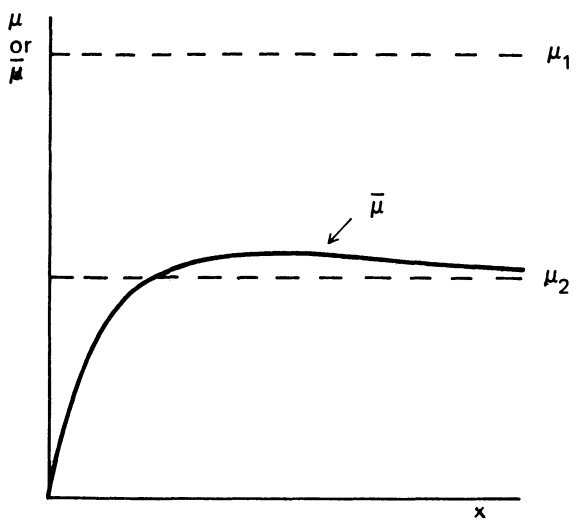

Figure 8. The hazard rate for a redundant system may exceed the hazard rate of its more reliable component. The curve for $\bar{\mu}$ was calculated from (5) using $\mu_1 = .1$ and $\mu_2 = .05$. The curves are shown for x from 0 to 64.

a system consisting of two components with constant failure rates will have a failure rate that first increases and then decreases; the levels of the failure rates for the two components are shown by the dotted lines in the figure.

At first thought, it may seem rather mystifying that a redundant system can be less reliable than a single-component system. A common sense explanation runs as follows. The functioning system can be in three possible states: both components are working, only the more reliable one is working, or only the less reliable one is working. As time passes, it becomes more likely that only one of the components is still working. If the probability that both components are still working is low enough, then the failure rate for the system is roughly equal to a weighted average of the failure rates of the two components. Thus the failure rate of the system can rise to a level between the failure rates of the two components. As more time passes, it becomes increasingly likely that if the machine is still working, it is working using the more reliable component. Consequently, the failure rate approaches the failure rate of the more reliable component.

Although Barlow and Proschan's example concerns two components with constant failure rates, the same effect can be shown in more elaborate examples with several components with changing failure rates. Consider a system with i independent components in parallel: the system fails when all i components fail. Let $p_i(x)$ be the probability that component i is functioning, at time x, as given by (2). Because the system can only fail when its last functioning component fails, the failure rate for the system is

$$\bar{\mu}(x) = \sum_i \mu_i(x) \frac{p_i(x) \prod_{j \neq i} [1 - p_j(x)]}{1 - \prod_j [1 - p_j(x)]}. \qquad (5)$$

Some models of human disease processes are based on the hypothesis that the body has several lines of defense and that some diseases occur only after all of these lines of defense have failed. Thus the formula might also be applied to the study of human mortality and morbidity.

Another application might be in studies of the extinction of families: a family unit might be defined as extinct when the last member of the unit dies. For example, evolutionary

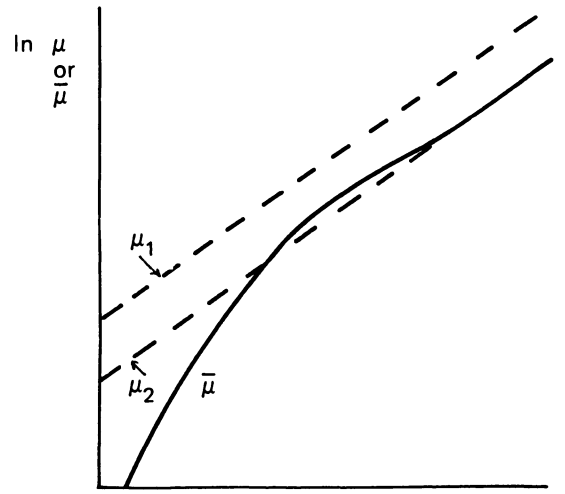

Figure 9. The hazard rate for a family may exceed the hazard rate for the more robust member of the family. The curve for $\bar{\mu}$ was calculated from (5) using $\mu_1(x) = .00333 \exp(.2x)$ and $\mu_2(x) = .01 \exp(.2x)$. The curves are shown for x from 0 to 30. Note that the curves are plotted on a logarithmic scale.

biologists study the extinction not only of species but also of higher taxonomic levels such as genera, families, and orders: A taxon dies when all the species in the taxon become extinct (Simpson 1983). As another, simpler example, consider a husband and wife who own an annuity that guarantees some monthly payment as long as either of them is living. If the husband's and wife's forces of mortality are independent of each other and are given by the dotted lines in Figure 9, then the hazard rate for the annuity is given by the solid curve in Figure 9. As the figure shows, at advanced durations the hazard rate for the annuity exceeds the wife's force of mortality. Furthermore, the hazard rate for the annuity follows a winding curve that initially rises at a much more rapid rate, but eventually at a somewhat slower rate, than the force of mortality curves. (The assumption that forces of mortality for members of a family are independent may be unrealistic; it is not difficult to adjust the calculations for a common cause of death.)

APPARENT FAILURES OF SUCCESS

In heterogeneous populations progress sometimes comes out looking like failure. Seven such ruses are adumbrated below.

The apparent gerontological failures that can be produced by pediatric success are illustrated by Figure 10. As shown in Figure 10b, a cohort consists of a frail and a robust subcohort. Health progress reduces mortality rates, at younger ages, from the solid lines to the dotted lines. As shown in Figure 10a, this does indeed lower mortality rates for the entire cohort at younger ages. At later ages, however, the observed cohort death rate is higher than it would have been. The frail individuals saved in childhood are dying at older ages. Every individual's life chances are improved at younger ages and are as good as ever at later ages, but observed cohort mortality makes it look as if pediatricians are making progress, whereas gerontologists are losing ground.

Consider now another kind of progress, namely, steady progress over time in reducing mortality at all ages:

$$\mu_i(x, y) = \mu_i(x, 0) \exp(-ry), \qquad i = 1, 2, \qquad (6)$$

where r is the rate of progress. [As before, μ is defined by (1); we now explicitly indicate that μ is a function of time y because we are no longer following a single cohort but are interested in an entire population over age and time.] The observed mortality rate will then steadily decline at age zero, but at older ages the pattern may be more complex. Observed mortality rates may decline at an increasing rate; they may rise and then fall; or, as shown on the left side of Figure 11, they may decline, increase, and then decline again.

An intuitive explanation of the pattern of the curve in Figure 11 runs as follows. Reductions in mortality rates at younger ages permit more individuals from the frailer subpopulation to survive to older ages. This influx of frailer individuals serves as a brake or counter-current on reductions in mortality rates at older ages. If the influx is small enough, progress may still be observed; but if the influx is large enough, observed mortality rates may actually increase. The size of the influx depends on the absolute magnitude of the reduction in mortality rates at younger ages (i.e., on the number of lives being saved in the frailer subpopulation) and on the chance a frailer individual has of reaching older ages. For the curve in Figure 11, the influx is small initially because so few frail individuals live to age 75; the influx becomes small again later because then so

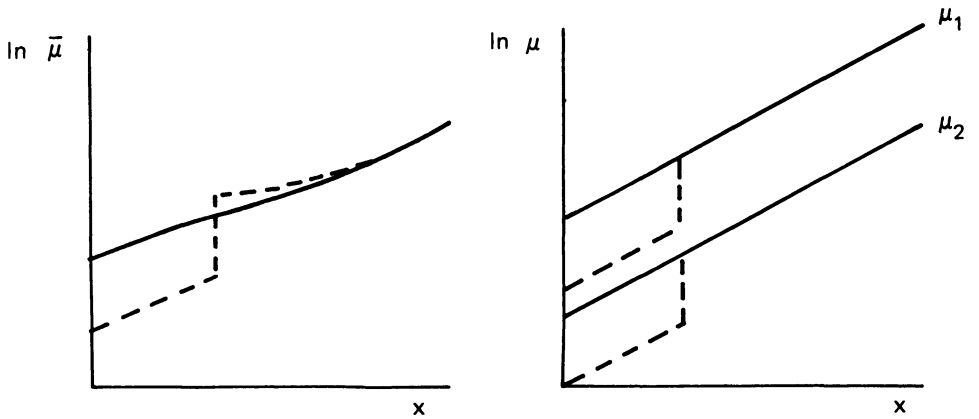

Figure 10. Lowering mortality rates before some age may increase observed mortality rates after that age. The solid curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1(x) = .05 \exp(.025x)$, $\mu_2(x) = .02 \exp(.025x)$, and $\pi(0) = .5$. For the dotted curves, mortality rates before age 24 were cut in half. The curves are shown for values of x from 0 to 72. Note that the curves are plotted on logarithmic scales.

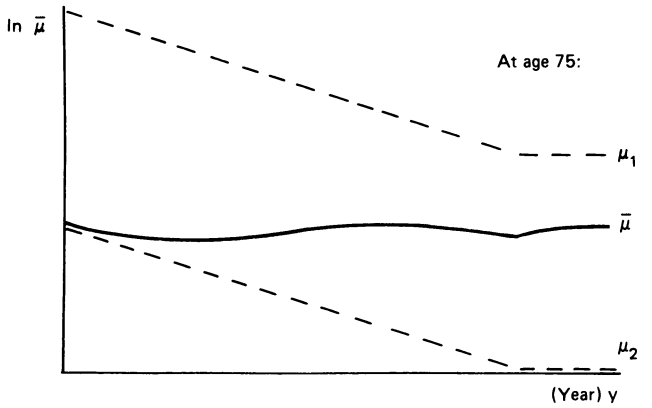

Figure 11. Observed mortality rates may follow complex patterns over time even though individual mortality rates are steadily declining (or become constant) at all ages. The curve for $\bar{\mu}$ was calculated from (6) and a simple generalization of (2), (3), and (4) using $\mu_1(x, 0) = .002 \exp(.07x)$, $\mu_2(x, 0) = .0001 \exp(.07x)$, $\pi_0 = .5$, and $r = .02$ until $y = 100$ and $r = 0$ afterwards. The curves are shown for values of y from 0 to 120.

few deaths occur before age 75.

Now suppose that progress against mortality ceases:

$$\mu_i(x, y) = \mu_i(x, y_0), \qquad i = 1, 2, y \geq y_0. \tag{7}$$

The observed mortality rate at age zero will then stay constant, but as shown on the right side of Figure 11, observed mortality rates at older ages will increase before leveling off. To understand this phenomenon, consider the cohorts aged 50 and 70 in the year progress ceases. Because the 50-year-olds have benefited from 20 more years of mortality progress than the 70-year-olds, there will be more frail individuals among the 50-year-olds than there were among the 70-year-olds twenty years ago (when they were 50 years old). Furthermore, because of the additional 20 years of mortality progress, more of these frail 50-year-olds will survive to age 70. Thus 20 years hence, when the 50-year-olds are 70 years old, more of them will be from the frailer subpopulation than is currently the case. Consequently, the observed mortality rate among those future 70-year-olds will be higher than it currently is.

This implies that when progress is being made against mortality, then currently observed mortality rates are lower than the mortality rates that would be observed if the current rates for individuals persisted or, indeed, merely declined at a slower rate of progress than before. Vaupel et al. (1979) indicated how to calculate the values of mortality rates under current health conditions, adjusted for heterogeneity and past health progress.

As explained by Keyfitz and Beekman (1984), another kind of ruse occurs in growing populations. In a population that consists of subpopulations with differing mortality rates, reductions in mortality rates for all of the subpopulations may lead to an increase in the observed mortality rates for the entire population. This ruse will occur if the reduction in mortality leads to more rapid growth in the size of the subpopulations that have high mortality rates.

As a simple example, consider a population with crude death rate $\bar{d}$ that consists of two stable subpopulations with equal crude growth rates and with death rates d_1 and d_2, d_1 being substantially greater than d_2. (Because the crude growth rates are the same, the birth rates also differ.) If the first subpopulation constitutes a proportion π of the total population, then

$$\bar{d} = \pi d_1 + (1 - \pi)d_2. \tag{8}$$

If the crude death rates of the two subpopulations are reduced by δ_1 and δ_2, such that δ_1 is greater than δ_2, then the crude growth rate for the first subpopulation will start exceeding the crude growth rate for the second subpopulation. The first subpopulation will thus constitute a greater and greater share of the total population: π will approach one. Hence the crude death rate will approach $d_1 - \delta_1$. As long as this value is greater than $\bar{d}$, the crude death rate will increase.

Since d_1 exceeds d_2, an equal percentage reduction yields δ_1 greater than δ_2, so equal percentage reductions are a special case of the situation just described. It is not difficult to generalize to n subpopulations or to the case in which the crude growth rates of the subpopulations are different. Under a variety of conditions, lowering individual or subpopulation death rates in a growing population can result in increases in the observed population mortality rate.

Yet another of heterogeneity's sleights of hand can be illustrated by a simple stochastic discrete-state model. Each of the members of some population are in one of two states. The hazard rates in these two states, μ_1 and μ_2, are constant over time. Moreover, μ_2 exceeds μ_1: the individuals in the second state are the frail or debilitated individuals. There is a constant transition intensity from state 1 to state 2, denoted by λ, but there is no transition from state 2 to state 1. Let $\pi(t)$ denote the proportion of individuals in state 2 at time t.

If λ is smaller than $\Delta\mu = \mu_2 - \mu_1$, then as shown by Petrovski et al. (1984) $\pi(t)$ will approach $\lambda/\Delta\mu$ and the observed force of mortality for the population as a whole will approach

$$\bar{\mu} = \lambda + \mu_1. \tag{9}$$

This is a surprising result because the observed force of mortality does not depend on the force of mortality in the second state. Any attempt to reduce $\bar{\mu}$ by reducing μ_2 will fail unless μ_2 can be sufficiently reduced so that $\Delta\mu$ is less than λ. Although this is an asymptotic result, to the extent that $\Delta\mu$ exceeds λ the asymptote will be approached in a fraction of the life span of the cohort. In an illustrative example given by Petrovski et al. (1984) cutting μ_2 in half hardly alters $\bar{\mu}$, but cutting λ in half reduces $\bar{\mu}$ by 30%: prevention can be much more effective than mitigation.

Consider now a generalization of the model such that λ depends on μ_2. In particular, suppose that $\lambda(\mu_2)$ increases as μ_2 decreases. This effect may occur widely: if cigarette smoking were made safer (if, say, a cure were developed for lung cancer), more people might smoke even though they would still face an elevated rate of heart disease; similarly, if automobiles were made safer, more people might drive recklessly (Peltzman 1975; Wilde 1982). If, as before, λ exceeds $\Delta\mu$, then

$$\bar{\mu} = \lambda(\mu_2) + \mu_1. \tag{10}$$

Consequently, if μ_2 decreases, $\bar{\mu}$ will increase. Making an activity safer can increase mortality.

Consider again a two-state model with constant transition rate λ from state 1 to state 2. As before, suppose that the hazard rate in the first state is given by a constant μ_1. But now suppose that there are two causes of death in the second state, with constant hazard rates μ_2 and μ_3. Let $\pi(t)$ represent the proportion of the surviving population in the second state at time t. The observed hazard rates from the three causes of death for the entire population are given by

$$\bar{\mu}_1(t) = [1 - \pi(t)]\mu_1, \tag{11}$$

$$\bar{\mu}_2(t) = \pi(t)\mu_2, \tag{12}$$

$$\bar{\mu}_3(t) = \pi(t)\mu_3. \tag{13}$$

It follows from Petrovski et al. (1984) that if μ_3 is decreased, then formula (10) implies that $\pi(t)$ will increase: more of the surviving individuals will be in the debilitated state if debilitated individuals are not dying as rapidly as before. Hence the observed hazard rate from the second cause of death, $\bar{\mu}_2$, will increase. In addition, the observed hazard rate from the first cause of death, $\bar{\mu}_1$, will decrease. Even if the three causes of death are independent at the individual level, at the population level they are linked.

This result can be generalized to more complex situations in which mortality rates increase with age, there are several causes of death, and there are several different states. In particular, it seems likely that in a wide variety of situations, reducing one cause of death will result in an increase in the observed mortality rate from some other causes and, perhaps, a decrease in the observed mortality rate from some remaining causes. Because everyone has to die of something, it is obvious that reducing one cause of death will increase the number of people dying from another. The point here is deeper: contrary to the commonly made assumption of independence among competing causes of death, reducing one cause of death may change the observed force of mortality from another cause of death—even if, on the individual level, it is true that the two causes of death are independent. In a heterogeneous population, a cure for cancer might raise the mortality rate from heart disease and lower it from automobile accidents.

STATISTICAL INFERENCE

A variety of methods exist for drawing statistical inferences using survival data. Books by Cox and Oakes (1984), Elandt-Johnson and Johnson (1980), Gnedenko et al. (1969), Kalbfleisch and Prentice (1980), and Lawless (1982) survey different approaches. Recently some attention has been focused on statistical inference when there is hidden heterogeneity. Manton and Stallard (1981b) reviewed applications in mortality and morbidity analyses; Heckman and Singer (1984) discussed economic applications. Empirical studies include Manton et al. (1981), Heckman and Singer (1982), Tuma (1983), and Trussel and Richards (1985).

As a simple illustration of this area of research, consider the following problem. Individuals fall into two unobserved subcohorts with constant hazard rates μ_1 and μ_2. As a result of ancillary studies, the values of μ_1 and μ_2 are known. Observations are available on age at death for every individual who died during the observation period. For all other individuals, the age at which the individual ceased to be observed is known. What is unknown and is to be estimated is the proportion, π_0, of individuals who are in the first subcohort at time zero.

To motivate this inference problem, consider a disease against which an imperfect vaccine has been developed. The disease might be encephalitis carried by ticks, and the exposed population of concern might be people who work outdoors in infested areas, or, perhaps, a population of animals being used to study the efficiency of a vaccine that is being tested. Vaccinated individuals may remain unprotected and at risk level μ_2 if their immunization response is inadequate. Successfully vaccinated individuals are at risk level μ_1, which might be zero. The statistical task is to estimate the proportion of individuals who were successfully vaccinated, based on data about the incidence of the disease in the vaccinated cohort over some period of observation.

As a second kind of application, suppose that some piece of equipment can be manufactured (or perhaps used) in two ways: the correct way and a shortcut way. If the equipment is made correctly, the hazard of failure is some constant μ_1; otherwise it is $\mu_2 > \mu_1$. Time to failure is observed. The statistical task is to estimate the proportion manufactured correctly.

The likelihood of an observed death time, x_i, is simply $\bar{\mu}(x_i)\bar{p}(x_i)$, and the likelihood that an individual who does not die in the observation period ceases to be observed at age x is simply $\bar{p}(x_i)$, where $\bar{\mu}$ can be calculated from (2), (3), and (4) and $\bar{p}$ is given by a formula like (2). Thus the likelihood of the data is

$$L(\pi_0) = \prod_{i=1}^{n} \bar{\mu}(x_i)^{\delta_i}\bar{p}(x_i), \tag{14}$$

where n is the size of the cohort, x_i is age at death or cessation of observation, and δ_i equals 1 if the individual dies (from the cause being studied) in the observation period and equals 0 otherwise. The maximum likelihood approach estimates the value of π_0 by the value $\hat{\pi}_0$ that maximizes L. Standard computer algorithms are available for this optimization task.

This simple example might be generalized in several ways. The values of μ_1 and μ_2 might not be known and have to be estimated. The hazard rates might depend on time such that, say,

$$\mu_i(x) = a_i e^{b_i x}, \qquad i = 1, 2, \tag{15}$$

and it might be necessary to estimate a_i and b_i. A vector of covariates, ω, might be observed for each individual, and a proportional hazards model might be assumed of the form

$$\mu_i(x, \omega) = a_i e^{\beta\omega} e^{b_i x}, \qquad i = 1, 2, \tag{16}$$

where β is a vector of coefficients to be estimated. The references cited at the beginning of this section suggest some ways of tackling these more complicated problems of statistical estimation, but more research is needed on how to take into account hidden heterogeneity.

Such methodological research is especially important be-

cause even randomization, the statistician's traditional source of comfort and security, provides no immunity in survival analyses from heterogeneity's ruses. Randomized experiments are frequently used to control for hidden factors that influence the efficacy of some treatment or intervention; in the simplest case, the population is divided randomly into two groups, one of which receives the treatment while the other does not. If, however, the intervention affects survival chances, randomization does not preclude systematic bias caused by hidden heterogeneity.

As an illustration, suppose that a population consists of frail and robust individuals with hazard rates $\mu_1(x)$ and $\mu_2(x)$, where $\mu_1(x)$ is greater than $\mu_2(x)$ at all ages x; some proportion $\pi(0)$ of the initial population is frail. Further suppose that the treatment only helps the frail, reducing $\mu_1(x)$ at all ages by some constant δ that is less than the difference between $\mu_1(x)$ and $\mu_2(x)$. Let the expected values of the hazard rates for the treated and untreated groups be denoted by $\overline{\mu}'(x)$ and $\overline{\mu}(x)$, and let $\overline{\delta}(x)$ be the difference, $\overline{\mu}(x)$ minus $\overline{\mu}'(x)$. Formulas (2), (3), and (4) imply that $\overline{\delta}(0)$ will equal $\pi(0)\delta$, but that because both treated and untreated groups will increasingly consist of robust individuals, $\overline{\delta}(x)$ will approach zero as x increases. A naive experimenter with too much faith in the power of randomized, controlled trials, might erroneously conclude that the effectiveness of the treatment declines with time.

As a second illustration, suppose that the population consists of two subpopulations with identical hazard rates $\mu(x)$. The treatment, however, is only effective for the first subpopulation, which makes up some proportion $\pi(0)$ of the initial population; the treatment reduces the hazard rate for this responsive subpopulation to $\mu(x)$ minus some constant δ. As in the situation just described, let $\overline{\delta}(x)$ give the difference between the expected hazard rates in the untreated and treated groups. Formulas (2), (3), and (4) imply that $\overline{\delta}(0)$ will equal $\pi(0)\delta$, but that because the treated group will increasingly consist of responsive individuals, $\overline{\delta}(x)$ will approach δ as x increases. Consequently, in this example, a naive experimenter might erroneously conclude that the effectiveness of the treatment increases with time.

Thus even if a treated group and a control group at the start of some trial are absolutely identical in their composition—either as the result of some lucky randomization or some lucky nonrandom method of assignment—hidden heterogeneity may result in hazard trajectories for the treated and control groups that are misleadingly different from the underlying trajectories at the individual level.

DISCUSSION

Regardless of how many different attributes are considered, individuals who are grouped together will differ along various neglected dimensions. Some of these differences will almost certainly affect the individuals' chances of death, marriage, unemployment, or other transition. Because of this heterogeneity, selection will occur: the surviving population will differ from the original population. This in turn means that observations of the surviving population cannot be directly translated into conclusions about the behavior or characteristics of the individuals who made up the original population. The observed dynamics at the population level will deviate from the underlying dynamics at the individual level.

Sometimes this is not important. Perhaps the population, when classified along various observed factors, is more or less homogeneous, so that effects of unobserved heterogeneity are unsubstantial.

Sometimes, however, selection is important; and when it is, the patterns observed may be surprisingly different from the underlying patterns on the individual level. Researchers interested in uncovering these individual patterns, perhaps to help develop or test theories or to make predictions, might benefit from an understanding of heterogeneity's ruses. Because the impact of a policy intervention can sometimes only be correctly predicted if the varying responses of different kinds of individuals are taken into account, awareness of the effects of selection may also help policymakers.

When should a researcher suspect substantial heterogeneity? A useful clue occurs when theory and evidence pertaining to individuals suggest a trajectory of mortality that diverges from the observed trajectory for the population. For instance, human mortality may increase exponentially, but observed mortality curves appear to level off at advanced ages: the discrepancy suggests heterogeneity (Horiuchi and Coale 1983; Manton et al. 1984). Similarly, recidivism rates for individuals may be roughly constant, but observed rates decline, again suggesting heterogeneity (Harris et al. 1981). Another kind of clue may be offered by mortality convergences and crossovers: if one cohort is disadvantaged at earlier ages, then heterogeneity implies that it will appear to be less disadvantaged or even advantaged at later ages (Vaupel et al. 1979). Apparent waves in mortality rates over time may be a third kind of clue, if theory and ancillary evidence suggest a steadier pattern of change. More research needs to be done on this key question of how to tell when a population is sufficiently heterogeneous that selection matters.

[*Received February 1984. Revised February 1985.*]

REFERENCES

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

Barlow, R., and Proschan, F. (1975), *Statistical Theory of Reliability and Lifetesting*, New York: Holt, Rinehart & Winston.

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

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

Gnedenko, B. V., Belyaev, Yu. K., and Solovyev, A. D. (1969), *Mathematical Methods of Reliability Theory*, New York: Academic Press.

Harris, C. M., Kaylan, A. R., and Maltz, M. D. (1981), "Recent Advances in the Statistics of Recidivism Measurement," in *Models in Quantitative Criminology*, ed. J. A. Fox, New York: Academic Press, pp. 61–79.

Heckman, J. J., and Singer, B. (1982), "Population Heterogeneity in Demographic Models," in *Multidimensional Mathematical Demography*, eds. K. Land and A. Rogers, New York: Academic Press.

——— (1984), "Econometric Duration Analysis," *Journal of Econometrics*, 24, 63–132.

Horiuchi, S., and Coale, A. J. (1983), "Age Patterns of Mortality for Older Women: Analysis Using the Age-Specific Rate of Mortality Change

With Age," paper presented at the 1983 Annual Meeting of the Population Association of America, Pittsburgh.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley.

Keyfitz, N., and Littman, G. (1980), "Mortality in a Heterogeneous Population," *Population Studies*, 33, 333–343.

Keyfitz, N., and Beekman, J. (1984), *Demography Through Problems*, New York: Springer-Verlag.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York: John Wiley.

Manton, K. G., and Stallard, E. (1981a), "Methods for Evaluating the Heterogeneity of Aging Processes in Human Populations Using Vital Statistics Data: Explaining the Black/White Mortality Crossover by a Model of Mortality Selection," *Human Biology*, 53, 47–67.

——— (1981b), "Heterogeneity and Its Effect on Mortality Measurement," in *Methodologies for the Collection and Analysis of Mortality Data*, eds. J. Vallin, J. H. Pollard, and L. Heligman, Liege, Belgium: International Union for the Scientific Study of Population, pp. 265–299.

Manton, K. G., Stallard, E., and Vaupel, J. W. (1981), "Methods for Comparing the Mortality Experience of Heterogeneous Populations," *Demography*, 18, 389–410.

——— (1984), "Alternative Estimates of the Heterogeneity of Mortality Risks Among the Aged," submitted for publication.

Nam, C. B., Weatherby, N. L., and Ockay, K. A. (1978), "Causes of Death Which Contribute to the Mortality Crossover Effect," *Social Biology*, 25, 306–314.

Peltzman, S. (1975), "The Effects of Automobile Safety Regulations," *Journal of Political Economy*, 83, 677–725.

Petrovski, A. M., Vaupel, J. W., and Yashin, A. I. (1984), "Models of Risk Group Dynamics," in *System Science in Health Care*, eds. W. van Eimeren et al., Berlin: Springer-Verlag, pp. 936–937.

Rogers, A. (1982), "Parametrized Multistate Population Dynamics," Working Paper WP-82-125, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Shepard, D. S., and Zeckhauser, R. J. (1980), "Long-Term Effects of Interventions to Improve Survival in Mixed Populations," *Journal of Chronic Diseases*, 33, 413–433.

Simpson, G. G. (1983), *Fossils and the History of Life*, New York: W. H. Freeman.

Trussel, J., and Richards, T. (1985), "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman–Singer Procedure," in *Sociological Methodology 1985*, ed. N. B. Tuma, San Francisco: Jossey Bass.

Tuma, N. B. (1983), "Effects of Labor Market Structure on Job-Shift Patterns," Working Paper WP-83-11, Laxenburg, Austria: International Institute for Applied Systems Analysis.

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

Vaupel, J. W., and Yashin, A. I. (1983), "The Deviant Dynamics of Death in Heterogeneous Populations," Research Report RR-83-1, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Wilde, G. J. S. (1982), "The Theory of Risk Homeostasis: Implications for Safety and Health," *Risk Analysis*, 2, 209–225.