HETEROGENEITY'S RUSES: SOME SURPRISING EFFECTS OF SELECTION ON POPULATION DYNAMICS

James W. Vaupel
Anatoli I. Yashin

November 1983
CP-83-56

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

This is a working draft. Your comments and suggestions, especially concerning additional ruses, applications, or references, would be sincerely appreciated. Please write: James W. Vaupel, Institute of Policy Sciences and Public Affairs, 4875 Duke University Station, Durham, North Carolina, USA 27706 or Anatoli I. Yashin, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

ACKNOWLEDGEMENTS

We thank Philip Cook, Dianne Goodwin, Nathan Keyfitz, Andrei Rogers, Thomas Schelling, and Bodil Vaupel for their insights and suggestions. In addition, we would like to thank Martina Joestl-Segalla and Yasmin Taher for their contribution to making possible the preparation of the manuscript.

HETEROGENEITY'S RUSES: SOME SURPRISING EFFECTS OF SELECTION ON POPULATION DYNAMICS

James W. Vaupel* and Anatoli I. Yashin**

INTRODUCTION

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.

$$\mu(x,y) = \frac{\frac{dp(x,y)}{dy}}{p(x,y)} , \quad y = y_0 + x \quad , \tag{1}$$

where $p(x,y)$ is the proportion of the cohort born x years ago that is surviving at time y and where y_0 is the year the cohort was born. Note that x and y are continuous and that the derivative is a full, and not a partial, derivative. In a homogeneous population, all individuals age x in year y face the same hazard

*Duke University, USA and International Institute for Applied Systems Analysis, Austria.
**International Institute for Applied Systems Analysis, Austria, and Institute for Control Sciences, USSR.

rate $\mu(x,y)$. A heterogeneous population consists of various homogeneous sub-populations. In the extreme case, each homogeneous sub-population consists of a single individual.

Heterogeneity is sometimes used as a synonymn for variability or diversity; here it has a narrower meaning of variability with respect to mortality (or with respect to attributes of individuals that affect their mortality). This concept of heterogeneity is closely linked with the concept of selection: a heterogeneous population is one in which there is differential mortality and hence one in which selection is occurring.

Because of the effects of selection, the patterns of mortality (or exit) in a heterogeneous population can differ qualitatively from the patterns of mortality in the constituent sub-populations. These qualitative differences can be surprising; unsuspecting researchers who are not wary of heterogeneity's ruses may fallaciously assume that observed patterns for the population as a whole also hold on the sub-population or individual level. Such incorrect inferences may produce erroneous policy recommendations, because the effect of an intervention usually depends on the behavior and response of individuals. In addition, because rates for homogeneous groups often follow simpler patterns than composite population rates, both theoretical and empirical research may be unnecessarily complicated by failure to recognize the effects of heterogeneity.

The multiplicity of heterogeneity's ruses can be neatly illustrated in the simplest example of a heterogeneous population — namely, a composite population that consists of two homogeneous sub-populations. It is not difficult to develop models of heterogeneous sub-populations that consist of a very large or infinite number of sub-populations, as shown, for example, in some of our earlier work (Vaupel, Manton, and Stallard 1979; Vaupel and Yashin 1983). The ruses illustrated here could have been described in the context of such a model, but for purposes of simplicity and clarity, a focus on the most elementary kind of heterogeneous population seems appropriate. Moreover, it turns out that almost all 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 sub-populations. Important research on such mixed populations includes Blumen, Kogan, and McCarthy's pioneering work on mover/stayer models of labor mobility (1955), Shepard and Zeckhauser's health-care studies (1975; 1980), and Keyfitz and Littwin's (1980) and Keyfitz's (1983) analyses of mortality.

1. 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 sub-cohorts at age x and let $\overline{\mu}(x)$ 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 given in the definition of μ in (1). The key question of interest is: how does the trajectory of $\overline{\mu}(x)$ compare with the trajectories of $\mu_1(x)$ and $\mu_2(x)$?

Let $p_1(x)$ and $p_2(x)$ be the proportions of the two cohorts that survive to age x. And let n_1, n_2, and n be the sizes of the two sub-cohorts and the entire population at birth. Clearly, n is the sum of n_1 and n_2. Define $\pi(x)$ as the proportion of the surviving cohort at age x that is in the first sub-cohort. By definition,

$$\pi(0) = \frac{n_1}{n_1 + n_2} \tag{2a}$$

and

$$\pi(x) = n_1 \frac{p_1(x)}{n_1 p_1(x) + n_2 p_2(x)} \ , \quad x > 0 \quad . \tag{2b}$$

An alternative equation for $\pi(x)$ that follows from (2a) and (2b) is

$$\pi(x) = \pi(0) \frac{p_1(x)}{\pi(0) p_1(x) + [1 - \pi(0)] p_2(x)} \quad . \tag{2c}$$

Clearly,

$$\bar{\mu}(x) = \pi(x)\mu_1(x) + [1 - \pi(x)]\mu_2(x) \quad . \tag{3}$$

The dependency of the cohort hazard rate on the sub-cohorts' hazard rates is thus mediated by the changing proportion of the population that is in one or the other of the sub-cohorts.

Suppose the first sub-cohort is the weaker or frailer:

$$\mu_1(x) > \mu_2(x), \quad \text{all } x \quad .$$

since (1) implies

$$p_1(x) = \exp\left[- \int_0^x \mu_1(t) \mathrm{d}t \right] \ , \quad i = 1,2 \quad , \tag{4}$$

the proportion of the surviving population that is in the frailer sub-cohort, $\pi(x)$, will steadily decline. Consequently, the observed hazard rate will approach the hazard rate of the more robust sub-cohort. Figures 1 through 5 illustrate some specific instances. Figures 1-5a plot the observed hazard rate for the entire cohort, whereas Figures 1-5b plot the hazard rates for the two sub-cohorts.

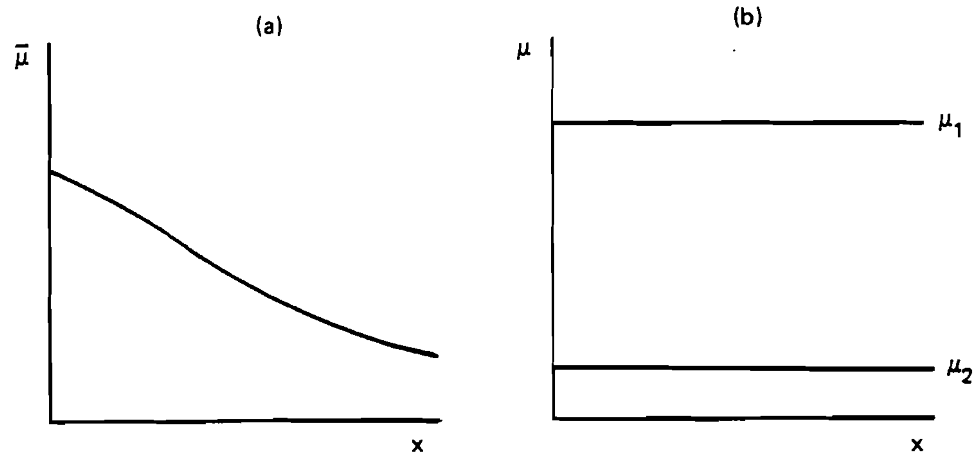

FIGURE 1 The observed hazard rate may decline even though the hazard rates for the two cohorts are constant.
NOTE The curve for $\bar{\mu}$ was calculated from (2c), (3), and (4) using μ_1=.06, μ_2=.01, and $\pi(0)$=.8. The curves are shown for values of x from 0 to 75.

1.1. Decreasing Recidivism

The recidivism rate for convicts released from prison declines with time since release (Harris, Kaylan, and Maltz, 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 recidivism rate (i.e., the instantaneous "force" or hazard of recidivism) for individual convicts, smokers, and alcoholics declines over time? Not necessarily. As illustrated in Figures 1a and b, 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.

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. The surface area of teeth may simply differ in susceptibility to decay.

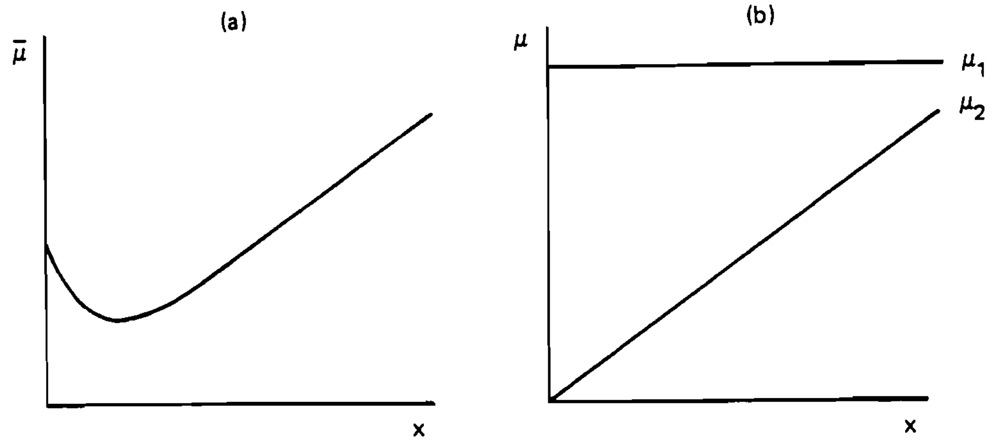

FIGURE 2 The observed hazard rate may decline and then rise even though the hazard rate for one cohort is rising steadily and the hazard rate for the other cohort is constant.
NOTE The curve for $\bar{\mu}$ was calculated from (2c), (3) and (4) using μ_1=.14, $\mu_2(x)$=.001+.015x, and $\pi(0)$=.5. The curves are shown for values of x from 0 to

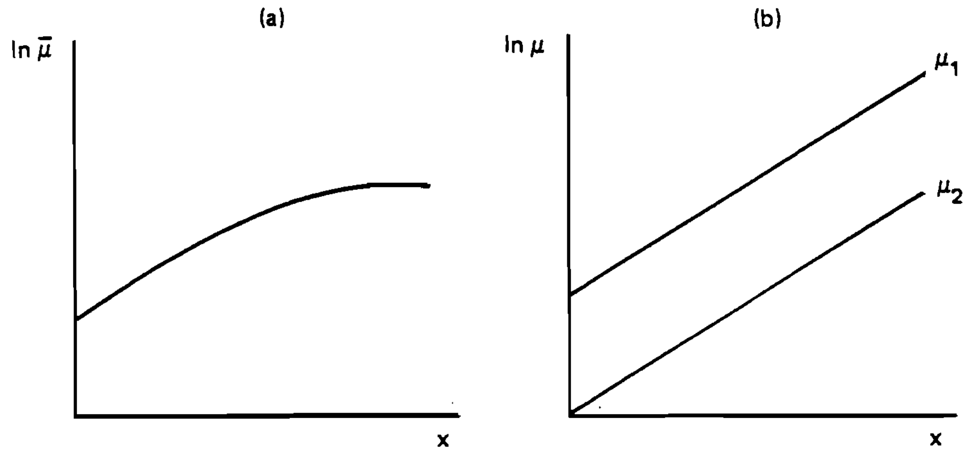

FIGURE 3 The observed hazard rate may rise steadily, then decline, and then rise again even though the hazard rates for the two cohorts are steadily rising.
NOTE The curve for $\bar{\mu}$ was calculated from (2c), (3), and (4) using $\mu_1(x)$=.0001·exp(.2x), $\mu_2(x)$=.0001·exp(.1x) and $\pi(0)$=.5.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

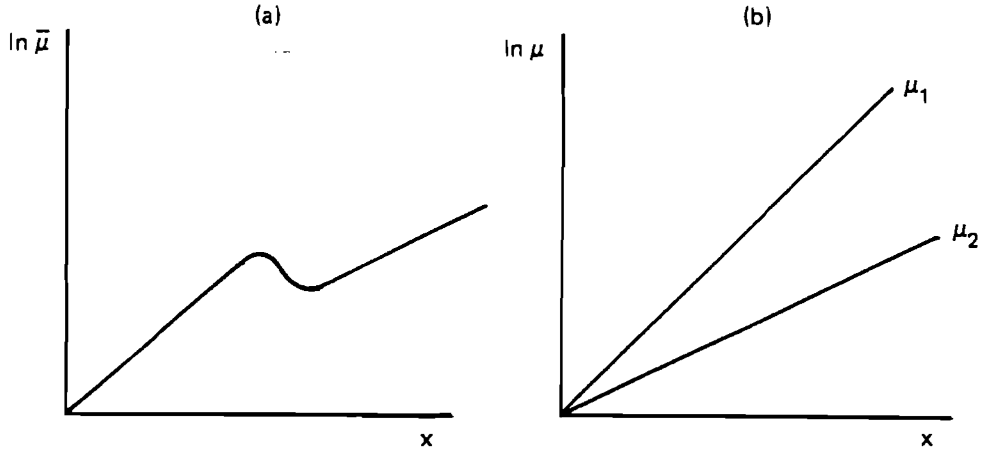

FIGURE 4 The observed hazard rate may increase more slowly than the hazard rates for the two cohorts.
NOTE The curve for $\bar{\mu}$ was calculated from (2c), (3), and (4) using $\mu_1(x)$=.01·exp(.04x), $\mu_2(x)$=.002·exp(.04x) and $\pi(0)$=.8. The curves are shown for values of x from 0 to 75. Note that $\bar{\mu}$ and μ_i are plotted on logarithmic scales.

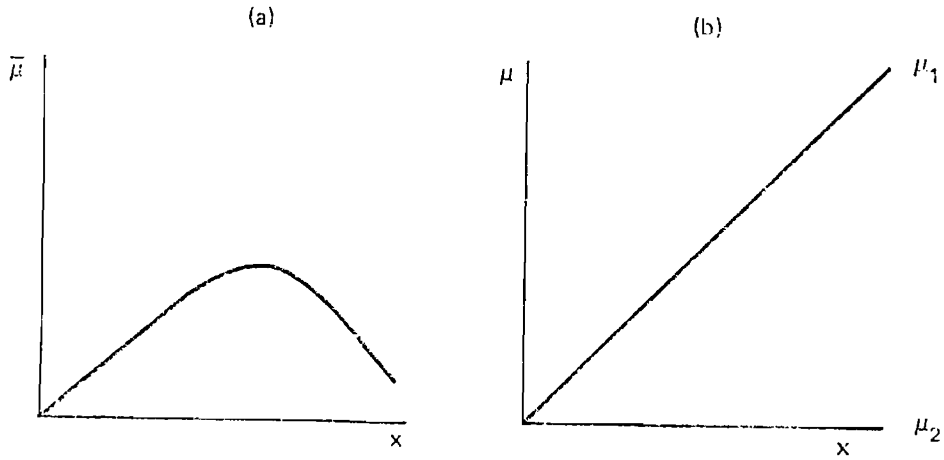

FIGURE 5 The observed hazard rate may increase and then decline if the hazard rate for one cohort is increasing and the other cohort is immune.
NOTE The curve for μ was calculated from (2c), (3), and (4) using $\mu_1(x)$=.002x, $\mu_2(x)$=0 and $\pi(0)$=.95. The curves are shown for values of x from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

A similar phenomenon is familiar to atomic physicists. Various elements have different isotopes: uranium, for example, has the common isotope U238 and the rarer isotope, used in atomic bombs, U235. Both isotopes are radioactive and thus decay: ultimately uranium is transformed into lead. In a sense, then, uranium can be considered to be dying and the population of uranium can be considered to consist of the subcohorts U235 and U238 that are dying at different rates: U235 decays faster than U238. Over time the observed radioactivity of a sample of uranium will decline, because the sample will consist increasingly of the "less frail" U238. "In 1910 Frederick Soddy called the varieties of the same element isotopes, as they were in the same place (*iso* means same; *tope* means place) in the list of chemical elements" (Weinberg 1983). By the way of analogy, members of the subcohorts of a population might be called "isotypes".

As a final example, consider the rate of innovation among some population of business firms. For instance, as the result of a banking reform, the savings banks in some states of the United States have been permitted to offer commercial loans. An economist keeps track of the rate at which savings banks start to offer this service. Over time, the rate declines. Does this imply that it is becoming more difficult or less profitable to break into the commercial market? Not necessarily. Some savings banks, perhaps the larger or more aggressive ones, may have a high probability of adopting the innovation whereas other savings banks may have a much lower probability.

1.2. Lemons

The cohort hazard rate shown in Figure 2a follows the "bathtub" shape familiar to reliability engineers (see, e.g., Mann, Shafer and Singpurwalla 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 "wearout phase"? Not necessarily. The high initial rate of breakdown could be due to a group of "lemons". Although in Figure 2b the hazard rate for the "lemons" is constant and for the other cohort it is increasing linearly, the bathtub curve can be produced by a variety of patterns in a mixed population. As long as one group is frail enough initially, its hazard rate can be increasing, constant, or even decreasing. The hazard rate for the more reliable group could be increasing linearly, quadratically, exponentially, etc. Note that if the population were only observed for a short time, about a quarter of the time displayed along the x-axis in Figures 2a and b, then the cohort curve would be steadily decreasing even though the curves for every device or individual in the population would be steadily increasing.

1.3 Waves

Figures 3a and b depict 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 cohorts that suffer constantly increasing hazard rates. The sudden decline in the observed hazard rate is produced by the rapid extinction of the frailer cohort. Up until the point of decline, the frailer cohort experiences death rates that are relatively low. Then, due to the exponential increase in the force of mortality, the death rates become sufficiently large so that within a few years almost all the frailer cohort dies. The observed hazard rate declines to the level of the hazard rate for the more robust cohort. 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 cohort because only members of the more robust cohort are still alive.

1.4. People Are Older Than They Look

Figures 4a and b depict a somewhat subtler ruse: the observed cohort hazard rate increases more slowly than the hazard rates for individuals in either sub-cohort. Individuals are, in a sense, aging more rapidly than the cohort data shows. Vaupel, Manton, and Stallard (1979) Vaupel and Yashin (1983), and Horiuchi and Coale (1983) explore various demographic implications of this effect.

1.5. The Extinction of the Vulnerable

In the so-called mover/stayer model, one group in the population is susceptible to emigration, marriage, divorce, some disease, etc., and the other group is immune. If the hazard for the susceptible cohort 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 Figures 5a and b illustrate. 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, for the other group 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 how the trajectory of the observed hazard rate deviates from the trajectories of the hazard rates for individuals in the two sub-cohorts. Similar ruses may hold for any characteristic of an individual that is correlated with an individual's hazard rate.

1.6. The Weight of Herring

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 the average weight of 4-year-olds. Does that mean that individual members of the species do not gain any weight between age 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.

1.7. A Vegetable Market

As another example, imagine an anthropologist who is observing a food market where sellers bargain with potential customers. She discovers the price of tomatoes steadily falling over the course of the day. Her initial hypothesis is that tomatoes deteriorate rapidly, but by studying a few selected tomatoes she 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.

1.8. Should Geologists Be Paid More?

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? It seems more 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 outspace their growing knowledge and make it increasingly difficult for them to discover oil.

2. Mortality Crossovers

2.1. The Advantages of Being Disadvantaged

Figure 6a depicts a so-called mortality crossover. One cohort's hazard rate is lower than the other cohort'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, Weatherby, and Ockay 1978); the effect also occurs for US Blacks vs. Whites (Manton and Stallard 1981). Some of these crossovers may be due to incorrect reporting of 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.

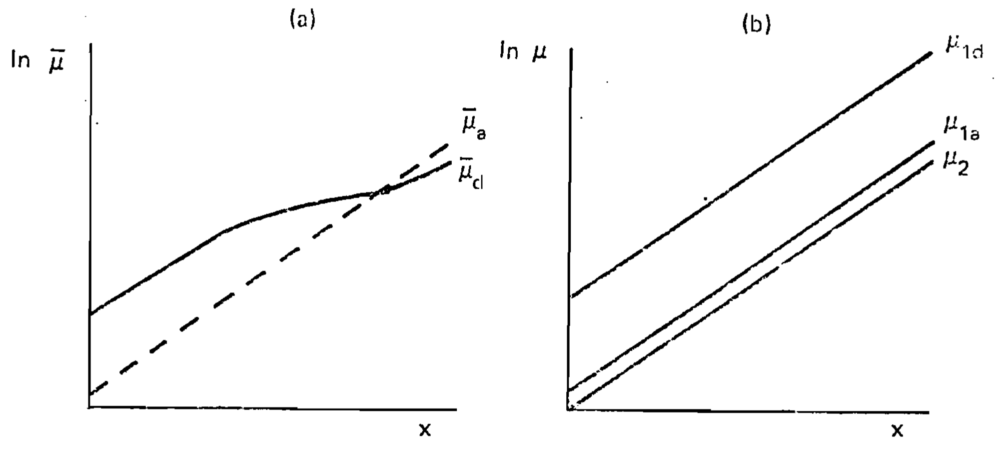

FIGURE 6 A disadvantaged cohort may appear to suffer lower mortality rates than an advantaged cohort at older ages.
NOTE The curves for μ_a and μ_d were calculated from (2c), (3), and (4) using $\mu_{1a}(x)$=.0025exp(0.4x), $\mu_{1d}(x)$=.01exp(.04x), $\mu_2(x)$=.002exp(.04x), and $\pi(0)$=.7. The curves are shown for values for x from 0 to 96. Note that the curves are plotted on logarithmic scales.

In particular, the cohort curves in Figure 6a can be produced, using (2), (3), and (4), from the sub-cohort curves shown in Figure 6b. The robust subcohorts of each of the two populations face the same mortality chances. the frail sub-cohort of the disadvantaged population, however, faces higher mortality chances than the frail cohort 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, Manton, Stallard 1979; Manton and Stallard 1981). 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.

2.2. Heart Failure

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 Figures 7a and b. A simple model (that readily can 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 cancer, the bottom line gives the corresponding hazard rate for individuals at low-risk of 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: the calculations are based on (2), (3), and (4). Essentially, the incidence of cancer declines relative to the incidence of heart failure because the individuals most susceptible to cancer have died.

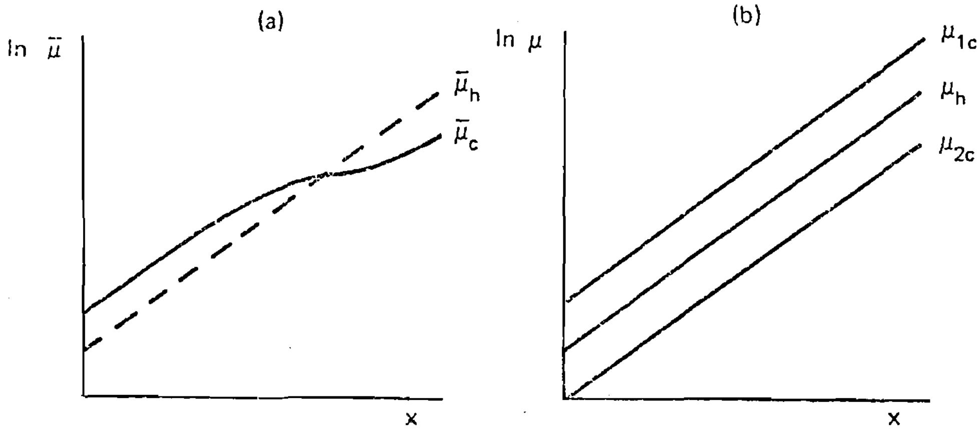

FIGURE 7 Observed mortality rates for two cause of death may appear to intersect.
NOTE The curve for μ_c was calculated from (2c), (3), and (4) using μ_{1c}=.01exp(.04x), μ_{2c}=.0025exp(.04x), and $\pi(0)$=.8. The curve for $\bar{\mu}_h$ is given by $\bar{\mu}_h(x)=\mu_h(x)$=.005exp(.04x). The curves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

3. Redundancy and the Death of Families

3.1. Does Redundancy Help?

Suppose a machine or device will fail if some specific component fails. To guard against this, a backup 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 backup component is somewhat less reliable than the original component, then Barlow and Proschan (1975) have shown that 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, 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. (The equations used to calculate this failure rate curve are presented below.)

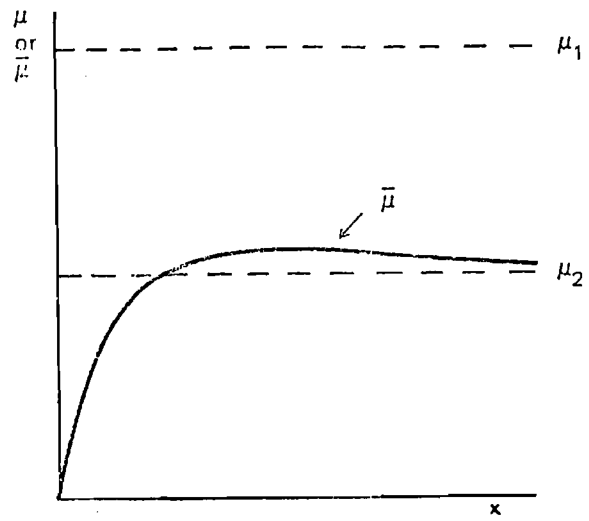

FIGURE 8 The hazard rate for a redundant system may exceed the hazard rate of its more reliable component.
NOTE The curve for $\bar{\mu}$ was calculated from (5), (6), and (7) using μ_1=0.1 and μ_2=0.05. The curves are shown for x from 0 to 64.

At first thought, it may seem rather mystifying that a redundant system can be less reliable than a single component. 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 components in parallel: the system fails when all i components fail. As before, let $\bar{\mu}(x)$ be the hazard or failure rate for the system. The failure rates for the various components are independent of each other. Let $p_1(x)$ be the probability that component i is functioning, given by

$$p_i(x) = \exp\left[-\int_0^x \mu_1(t)\mathrm{d}t\right] \quad . \tag{5}$$

Let $q_i(x)$ be the probability that only component i is working: all the other components have failed. Clearly,

$$q_i(x) = p_i(x)\left\{1 - \prod_{j \neq i} p_j(x)\right\} \quad . \tag{6}$$

Since the system can only fail when its last functioning component fails.

$$\bar{\mu}(x) = \sum_i \mu_i(x) \frac{q_i(x)}{1 - \prod_i \left[1 - p_i(x)\right]} \quad . \tag{7}$$

These equations, simplified to the special case where there are only two components, were used to calculate Figure 8.

The equations could be applied to the study of human mortality and morbidity rather than equipment failure. Some models of human disease processes are based on the hypothesis that the body has several lines of defense and that some diseases only occur after all of these lines of defense have failed. The multiple-hit model of carcinogenesis, for example, assumes that malignant tumor growth results only after several independent "hits", perhaps from radiation or exposure to a chemical, have altered the nature of the affected cell. If the different kinds of hits occur at different hazard rates, then the mathematics sketched above may be useful.

3.2. The Life-Span of a Family

Another application might be in actuarial studies of the death of families: a family unit might be defined as dead when the last member of the unit dies. For example, evolutionary 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 ages 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).

4. Apparent Failures of Success

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

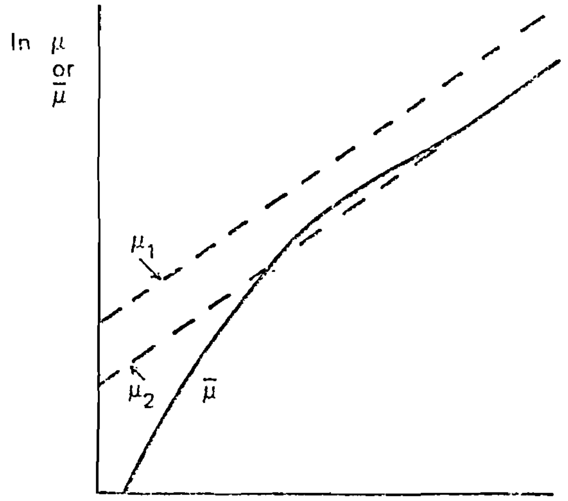

FIGURE 9 The hazard rate for a family may exceed the hazard rate for the more robust member of the family.
NOTE The curve for $\bar{\mu}$ was calculated from (5), (6), and (7) using $\mu_1(x)=.00333\exp(.2x)$ and $\mu_2(x)=.01\exp(.2x)$. The curves are shown for x from 0 to 30. Note that the curves are plotted on a logarithmic scale.

4.1. The Future of Gerontology

The apparent gerontological failures that can be produced by pediatric success are illustrated by Figures 10a and b. As shown in Figure 10b, a cohort consists of a frail and a robust sub-cohort. 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.

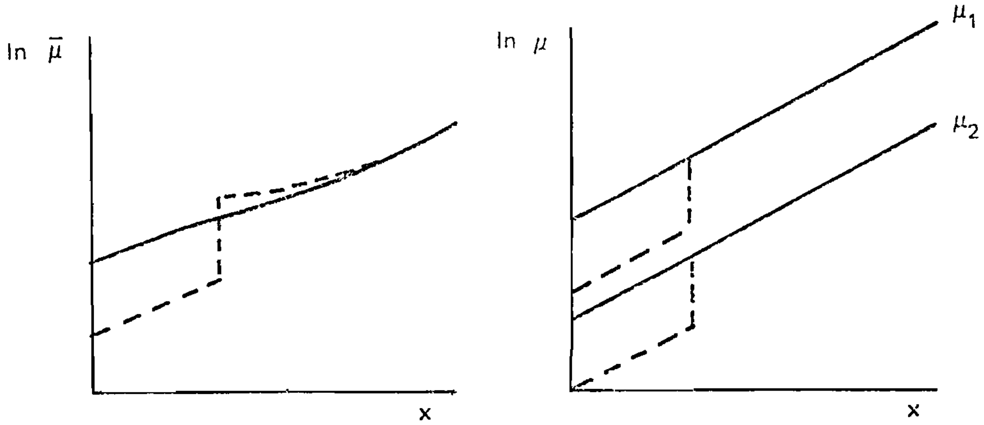

FIGURE 10 Lowering mortality rates before some age may increase observed mortality rates after that age.
NOTE The solid curve for $\bar{\mu}$ was calculated from (2c), (3), and (4) using $\mu_1(x)$=.05exp(.025x), $\mu_2(x)$=.02exp(.025x) and $\pi(0)$=.5. For the dotted curves, mortality rates before 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.

4.2. More Puzzles for Demographers

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

$$\mu_1(x,y) = \mu_1(x,0)\ \exp(-ry), \quad i = 1,2 \quad . \tag{8}$$

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.) Then, the observed mortality rate will 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-hand side of Figure 11, they may decline, increase, and then decline again. The curve in Figure 11 is based on exponentially increasing mortality with age, but similarly complex patterns can be generated using constant mortality, linearly increasing mortality, and so on.

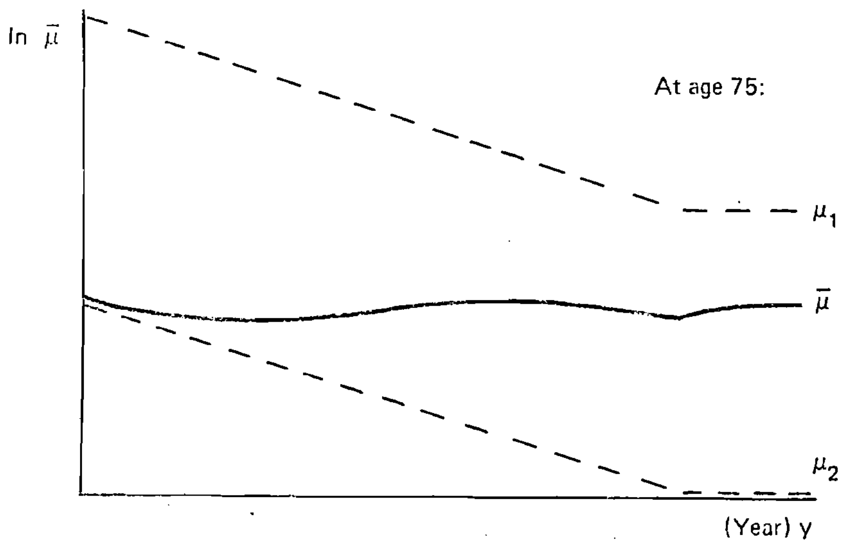

FIGURE 11 Observed mortality rates may follow complex patterns over time even though individual mortality rates are steadily declining (or become constant) at all ages.
NOTE The curve for $\bar{\mu}$ was calculated from (8), (9), (10), and (11) using $\mu_1(x,0)$=.002exp(.07x), $\mu_2(x,0)$=.0001exp(.07x), π_0=.5, and r=.02 up until y=100 and r=0 afterwards. The curves are shown for values of y from 0 to 120. Note that the curves are plotted on a logarithmic scale.

The formulas used to calculate such curves are simple generalizations of (2c), (3) and (4). For each of the two sub-populations, the proportion of a cohort born in year $y-x$ that is alive at age x is given by:

$$p_i(x,y) = \exp\left[-\int_0^x \mu_i(t,y-x+t)\mathrm{d}t\right], \quad i = 1,2 \quad . \tag{9}$$

The proportion of the entire surviving population at age x in year y that is in the first sub-population is given by:

$$\pi(x,y) = \pi_o \frac{pi(x,y)}{\pi_o p_1(x,y) + (1-\pi_o)p_2(x,y)} \quad , \tag{10}$$

where πo is the proportion of the entire population at birth that is in the first sub-population; this proportion is assumed to be constant over time. Finally, the observed hazard rate is given by

$$\bar{\mu}(x,y) = \pi(x,y)\mu_1(x,y) + [1 - \pi(x)]\mu_2(x,y) \quad . \tag{11}$$

To generate the curves in Figure 11, it was assumed that

$$\mu_i(x,0) = a_1 \exp(bx) \quad , \tag{12}$$

where a_1 was greater than a_2. The values of $\mu_i(x,y)$ were calculated using (8).

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 sub-population 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 sub-population) 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 on because so few deaths occur before age 75.

4.3. When Progress Stops

Now suppose that progress against mortality ceases: after declining for many years, mortality rates henceforth remain constant:

$$\mu_i(x,y > y0) = \mu_i(x,y_0), \quad i = 1,2 \quad . \tag{13}$$

Then, the observed mortality rate at age zero will stay constant, but as shown on the right-hand 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 have 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 twenty years of mortality progress, more of these frail 50 year-olds will survive to age 70. Thus, twenty years hence, when the 50 year-olds are 70 years old, more of them will be from the frailer sub-population 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, Manton, and Stallard (1979) indicate how to calculate the values of mortality rates under current health conditions, adjusted for heterogeneity and past health progress.

4.4. The Growth of Failure Out of Success

As explained by Keyfitz (1983), another kind of ruse occurs in growing populations. In a population that consists of a number of sub-populations with differing mortality rates, reductions in mortality rates for all the sub-populations 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 sub-populations that have high mortality rates.

As a simple example, consider a population with crude death rate $\bar{d}$ that consists of two stable sub-populations with the same crude growth rate and with death rates d_1 and d_2, d_1 substantially greater than d_2. (Because the crude growth rates are the same, the birth rates will also differ.) Then reducing both d_1 and d_2 by the same amount, say .01, or by the same percentage, say 10 percent, will eventually result in an increase in $\bar{d}$. If the first sub-population constitutes a proportion π of the total population, then

$$\bar{d} = \pi d_1 + (1 - \pi) d_2 \quad . \tag{14}$$

If the crude death rates of the two populations 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 sub-population will start exceeding the crude growth rate for the second sub-population. The first sub-population 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 above. It is not difficult to generalize to n sub-populations or to the case where the crude growth rates of the sub-populations are different. Under a variety of conditions, lowering individual or sub-population death rates in a growing population can result in increases in the observed population mortality rate.

4.5. When Death Does Not Influence Mortality

Yet another of heterogeneity's sleights of hand can be illustrated by a simple type of stochastic discrete-state (or compartment) model, depicted in Figure 12. Each of the members of some population are in one of two states. The forces of mortality from 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 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. Assume that $\pi(0)$ is given.

As shown by Yashin, Vaupel, and Manton (1983),

$$\pi(t) = \frac{1 - \frac{\lambda}{\Delta\mu} \frac{\pi(0)-1}{\pi(0)-\lambda/\Delta\mu} e^{\frac{\Delta\mu-\lambda}{\Delta\mu}t}}{1 - \frac{\pi(0)-1}{\pi(0)-\lambda/\Delta\mu} e^{\frac{\Delta\mu-\lambda}{\Delta\mu}t}} , \tag{15}$$

where $\Delta\mu = \mu_2 - \mu_1$ exceeds zero. If λ is smaller than $\Delta\mu$, then $\pi(t)$ will approach $\lambda/\Delta\mu$ and the observed force of mortality for the population as a whole will approach

$$\bar{\mu} = \lambda + \mu_1 \quad . \tag{16}$$

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 $\Delta\mu$ exceeds λ the asymptote will be approached in a fraction of the life-span of the cohort.

4.6. Prevention vs. Cure

As a concrete illustration, consider a population of individuals who are exposed to a condition or disease that in most cases quickly results in death. Specifically, assume that μ_1 = .03 , u_2 = .80, and λ = .02. As shown in Table 1, the overall death rate $\bar{\mu}$ quickly approaches .05 even in the extreme case where all the individuals in a cohort start off without the condition. If an astonishing breakthrough is made such that μ_2 is cut from .80 to .10 or even

.50, then $\bar{\mu}$ changes relatively little. However, halving the incidence of the condition — i.e., reducing λ from .02 to .01 — reduces $\bar{\mu}$ by about 20 percent. And cutting λ from .02 to .001 produces a close to 40 percent reduction in $\bar{\mu}$.

TABLE 1 The impact over time on $\bar{\mu}$ of various changes in μ_2 and λ.

μ_1=	.03	.03	.03	.03	.03
μ_2=	.80	.10	.05	.80	.80
λ=	.02	.02	.02	.1	.01
Time					
0	.030	.030	.030	.030	.030
1	.043	.042	.040	.036	.031
2	.047	.046	.043	.039	.031
3	.049	.048	.045	.039	.031
4	.050	.049	.046	.040	.031
5	.050	.050	.047	.040	.031
∞	.050	.050	.050	.040	.031

4.7. Trying to Help Smokers

Consider now a generalization of the model shown in Figure 12 such that λ depends on μ_2. In particular, suppose that λ 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; if automobiles are made safer, more people might drive recklessly (Peltzman, 1975; Wilde, 1982). If, as before, λ exceeds $\Delta\mu$, then

$$\bar{\mu} = \lambda(\mu_2) + \mu_2 \quad . \tag{16'}$$

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

4.8. Debilitation and Death

Another simple kind of stochastic discrete-state model is shown in Figure 13. There are two states and a single, constant transition rate λ. The force of mortality in the first state is given by a constant μ_1. There are two causes of death in the second state, with constant forces of mortality μ_2 and μ_3. Let $\pi(t)$ represent the proportion of the surviving population in the second state at time t. State 1 might be imagined as being the "healthy" state, and state 2 the "debilitated" state, perhaps the state of old age or of high blood pressure.

The observed forces of mortality from the three causes of death for the entire population are given by

$$\bar{\mu}_1(t) = [1 - \pi(t)]\mu_1 \quad , \tag{17a}$$

$$\bar{\mu}_2(t) = \pi(t)\mu_2 \quad , \tag{17b}$$

and

$$\bar{\mu}_3(t) = \pi(t)\mu_3 \quad . \tag{17c}$$

If μ_3 is decreased, then equation (15) 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 force of mortality from the second cause of death, $\bar{\mu}_2$, will increase. And the observed force of mortality from the first cause of death, μ_1, will decrease. Even if the three causes of death are independent on the individual level, on the population level they are linked.

This result can be generalized to more complex situations where mortality rates increase with age, where there are several causes of death, and where there are several different states (see Yashin, Vaupel, and Manton 1983 for some of the mathematics). 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.

DISCUSSION

Every individual (or thing) differs from every other individual (or thing) in countless ways. It is impossible to take all these differences into account: in all research and in all policy analysis individuals are classified along a few dimensions or in a few categories. In the analysis of human mortality, the salient dimensions have usually been taken to be age, sex, race, and nationality. Sometimes other factors have also been considered, such as educational achievement or cholesterol level, but most of the characteristics of individuals have been ignored. Studies of phenomenon other than mortality, such as marriage or employment, have taken other kinds of factors into account. Hair color may not be significant in a study of mortality but it might be significant in a study of marriage. Hence, depending on a researcher's or policymaker's interests, individuals should be classified in different ways.

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 on the population level will deviate from the underlying dynamics on 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's 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.

REFERENCES

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

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

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

Horiuchi, S. and A.J. Coale (1983) Age Patterns of Mortality for Older Women: Analysis Using the Age-Specific Rate of Mortality Change with Age. Unpublished paper presented at the 1983 Annual Meeting of the Population Association of America.

Keyfitz, N., and G. Littman (1980) Mortality in a Heterogeneous Population. Population Studies 33: 333-343.

Keyfitz, N. (1983) Composite of Stable Subpopulations. Theoretical Population

Biology (forthcoming).

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

Manton, K.G. and E. Stallard (1981) 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.

Nam, C.B., N.L. Weatherby, and K.A. Ockay (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.

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

Shepard, D.S. and R.J. Zeckhauser (1975) The Assessment of Programs to Prolong Life, Recognizing Their Interaction with Risk Factors. Discussion Paper 32D. Cambridge, Massachusetts: Kennedy School of Government, Harvard University.

Shepard, D.S. and R.J. Zeckhauser (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.

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 (1983) The Deviant Dynamics of Death in Heterogeneous Populations. Research Report 83-1. International Institute for Applied Systems Analysis, Laxenburg, Austria. Also published in N.B. Tuma (ed.) Sociological Methodology 1984. San Francisco: Jossey Boss.

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

Yashin, A.I., J.W. Vaupel, and K.G. Manton (1983) Stochastic Discrete-State Models of the Dynamics of Mortality and Morbidity in Heterogeneous Populations. Unpublished Working Paper. International Institute for Applied Systems Analysis, Laxenburg, Austria.