J Chron Dis Vol. 33, pp. 413 to 433

0021-9681/80/0701-0413$02.00/0

LONG-TERM EFFECTS OF INTERVENTIONS TO IMPROVE SURVIVAL IN MIXED POPULATIONS*

DONALD S. SHEPARD and RICHARD J. ZECKHAUSER

(*Received in revised form* 10 *September* 1979)

Abstract—This analysis uses the concept of mixed populations—that is, ones among whose members the probability of loss varies, either in its baseline value or in its response to an intervention. Losses from such heterogeneous populations will be skewed in a systematic manner; members with high loss probabilities will be disproportionately represented among early dropouts. Similarly, such populations will respond to an intervention designed to postpone losses in a manner that reflects any differential in benefits afforded members at varying risk levels.

Traditional assessments of interventions are systematically biased, for they fail to take adequate account of variability in risk among members of a population. A general methodology is developed here for inferring the structure of a mixed population to the extent possible, for predicting accurately its response to an intervention and for extending existing models of mortality. The methodology is applied to data drawn from a number of health-related examples. For hernia recurrence, an outstanding fit is achieved when individuals at the 10th percentile are assigned 100 times the 5-yr recurrence risk of those at the 90th percentile. A blood pressure control example shows that traditional assessments overstate mortality reductions at age 75 by 16%. Cross-population comparisons of life expectancy (by nation and by race) exhibit crossovers in remaining life expectancy with increasing age.

The mixed-population approach can be useful whenever there are heterogeneous populations with dropouts. Whether the populations at issue consist of college students, satisfactory housing units, reformed criminals or patients receiving hernia repairs, attention to the variability of risk among individuals will strengthen our understanding of the dynamic processes at work, thereby enhancing our predictive powers.

THE PRIMARY objective of many health-promoting interventions, whether based in medicine, lifestyle or environment, is to postpone a morbid event. Most commonly, the critical event is death, but it may equally well be the onset or progression of any illness. Such an event can be defined as a loss from a population—that is, a cohort of people, usually but not necessarily of the same age and sex, for whose members the specified event has not yet occurred. Although this discussion focuses on issues related to health, the underlying analysis applies equally well to problems in such diverse fields as education (where the critical event might be dropping out of school), production (machine malfunction), rehabilitation of criminals (recidivism), demography (migration) or unemployment (getting a job). In formulating policy, it is often important to assess the benefits returning to alternative programs of health intervention. Our approach provides an objective and systematic method of predicting the way a population will respond to such programs.

Numerous examples are developed here. Their purpose is not to report definitive medical findings—our restricted data sets and one- or two-factor hypotheses are too limited for that. Rather, our goal is to illustrate the widespread applicability of the mixed-population concept in medical contexts.

*Dr. Shepard is Adjunct Research Associate, Kennedy School of Government and Lecturer, School of Public Health, Harvard University. Dr. Zeckhauser is Professor of Political Economy, Kennedy School of Government.

1. MIXED POPULATIONS

If a population were homogeneous, assessments of alternative interventions should be rather straightforward, at least in principle. Homogeneity, however, is unlikely. Even individuals of the same age and sex differ on many dimensions that virtually affect health status. Our interest here, therefore, is in *mixed populations*, ones in which the probability of loss from the population varies among members of a cohort in either its initial value, its evolution over time or its response to an intervention. This analysis models losses from a mixed population, hoping thereby to provide improved predictive tools for policy analysts and decisionmakers, and an enhanced interpretative capability for those who wish to draw inferences from longitudinal data.

In recent years, mixed-population models have been applied in a number of disciplines. According to a 'mover–stayer' model for labor market mobility, only a portion of the workers (the movers) are subject to job changes [1]. Extensions of this model allow a continuous distribution of workers' rates of job change [2], and changes over time in an individual's rate of mobility [3–5]. In the biostatistics literature, regression models have been proposed for estimating times to failures of medical treatments according to characteristics of the subjects and their treatments [6, 7]. We previously found mixed-population concepts useful for studying cohorts as diverse as recipients of kidney transplants or hernia repairs (who might suffer rejection or relapse) and automobile drivers (who might have accidents) [8, 9]. This paper sets forth a systematic approach for examining interventions in mixed populations. Thus, it provides a means to assess the effects of policies designed to alter probabilities of dropping out of such populations. It describes the survival pattern with and without the intervention, suggests how the concept can be applied even though the underlying risk factors may be unobservable, indicates the biases if heterogeneity is not recognized, illustrates the diversity of situations for which the concept is useful and presents policy implications for a sample of health programs.

A. *Competing risk models*

Predicting the response of a mixed population to an intervention is closely related to the classic demographic problem of analyzing competing causes of death. There the objective is to assess the impact of adding, deleting or altering one of a number of causes of death. Most efforts to model this problem start with an admittedly naïve assumption: within a given age–sex group, the competing causes of death act independently. This assumption permits ready computation of the gains in life expectancy to be reaped, say, by eliminating cardiovascular death [10] or different types of cancer [11]. Post-intervention mortality rates are estimated by removing the age-specific mortalities due to the causes met by the intervention.

Empirical evidence shows the independence assumption to be invalid for many applications. Examining mortality rates in Massachusetts, Jenkins *et al.* [12] find that regions where mortality rates for one cause of death are higher than expected on the basis of age and sex tend to have higher standardized rates for a number of other causes as well. Jenkins *et al.* [12] suggest that common factors—environment, lifestyle or inadequate medical care—simultaneously raise a number of risks. Influenza epidemics seem to raise cancer death-rates, as they weaken cancer victims [13]. States and countries with high rates of breast cancer tend to have high rates of colorectal cancer. Cultural and dietary influences (such as high intakes of beef) may be common to both diseases [14]. Individuals who have contracted one type of cancer appear to face one-and-a-half to two times the expected risk of certain other primary cancers [15, 16]. Part of this risk is attributable to genetic or environmental predispositions, although often the specific factor(s) has not been identified.

Although the need to recognize interdependence among competing death causes has long been recognized [17], models to include such dependencies have been limited [18]. These dependencies have been formalized, for example, by assuming that potential sur-

vival times until death due to competing causes (also termed multiple decrements) follow some joint probability function of a specified parametric form [19, 20], or derive from some common factor [21]. Because these models seek to estimate a survival distribution from known data on numbers, times and causes of death, they are not well adapted to use epidemiological findings to estimate the effect of changes in risk factors or exposures or partial elimination of certain causes of death. Furthermore, models that integrate increasing mortality with age in a realistic pattern tend to be cumbersome.

Although they are not often identified as such, two types of interdependencies, age and sex, are commonly incorporated into competing risk models. Epidemiologists know that most risks of death vary significantly with age, and many vary with sex as well. An elevated risk of death due to cancer is associated with an elevated risk from stroke because both are associated with age. Similarly, death due to coronary heart disease is associated with death due to motor vehicle accidents because both causes represent greater risks in males than in females.

Estimates of the potential benefit of an intervention that could eliminate cardiovascular–renal disease provide a dramatic illustration of the importance of modeling the dependence of risk on age. If one assumed that age (and other possible interdependencies) were unimportant, then in the absence of cardiovascular–renal disease, the long-run crude death-rate would be simply the rate from all other causes. Using 1960 data for white males as an example, this rate was 67/10,000 man-years. Life expectancy would be the reciprocal of this death-rate, or 149 yr. This estimate, though impressive, is absurd: it ignores the fact that persons dying of cardiovascular disease tend to be older than those still alive, and thus faced an increased risk of death from other causes. Eliminating one cause of death would increase the *crude* death-rates from other causes, because the population as a whole would be older. (This effect is termed the Taeuber paradox [22].) Using the same data, and appropriately considering the factor of age, Chiang [23] calculated that eliminating cardiovascular disease would increase life expectancy from 67.3 to 79.0 yr.

Heterogeneity of risk has been shown to explain the shape of certain relative survival curves [24]. Relative survival at some number (x) years is the ratio of the proportion of persons with a given disease surviving for x yr to the proportion of a comparison group (generally matched for age and sex) surviving for the same number of years. After an expected steep decline, relative survival for women with cancer of the colon and small intestine flattens and begins to rise around the fifth year, indicating that the current mortality rate among these cancer cases has dropped below that of the comparison group. The rise in relative survival can be explained by the variation in risk among colon and intestinal cancer cases. The patients who die within 5 yr are likely to be those at highest risk (in terms of age or other factors). Survivors after 5 yr tend to be predominantly low-risk patients. This selection effect explains why the overall prognosis for 5-yr survivors of colon and intestinal cancer may (and, in fact, does) turn out to be better than that of a comparison group.

An innovative paper by Park and Lees [25] in 1951 indicated that consideration of heterogeneity can lead to an appraisal of clinical therapies very different from the conventional wisdom. They contended that early treatment for breast cancer at that time was virtually ineffective, because the apparently better prognosis of cases identified when the tumor was small was only a spurious association. They reasoned that there was heterogeneity in the rate of growth of malignant tumors of the breast. Those diagnosed early were more likely to be slower growing and inherently less prone to metastasize. More recently, Fox [26] has shown that heterogeneity in the case fatality rate of breast cancer cases combined with a greater proportion of less virulent cases receiving treatment could explain why the case fatality rate for breast cancer has declined while the population mortality rate has remained constant over recent decades.

B. *Mixed population models, a disaggregation approach*

Our model of mixed populations extends the principle of disaggregation inherent in

TABLE 1. FIRST RECURRENCE AFTER HERNIORRHAPHY

Interval after surgery (yr)	Proportion of survivors free of recurrence at end of interval*	Rate of first recurrence during interval†
0–1	$1-0.1 \times 0.30 = 0.970$	30.0
1–5	$1-0.1 \times 0.65 = 0.935$	9.1
5–10	$1-0.1 \times 0.81 = 0.919$	3.4

*This value is computed in the table as 1 − (proportion of patients eventually suffering recurrences) × (cumulative proportion of recurrences up to the end of this interval). Data derived from Neuhauser [27].
†Recurrences per 1000 person-years among persons still free of recurrence.

sex and age classification to stratification on other variables such as smoking habits or blood pressure characteristics. The principle is the same: among persons *with a given set of characteristics*, the risks of death due to competing causes can be modelled on the assumption of independence. If these risks prove not to be independent, generally the characteristics can be stratified further until independence is achieved, or until dependencies fall to levels that are below practical importance. In general, the risks of death will differ among the strata of this mixed population. If only one type of risk is being examined, it is only necessary to continue stratifying until that particular risk achieves independence from other risks of death; the other risks may continue to be interdependent. (Computational complexity could interfere with this process, as could hard-to-separate dependencies between highly correlated risks.)

Mixed-population models are mathematically tractable. Results for a homogeneous population within which risks are independent can be applied directly to each stratum. Corresponding results for the aggregate population can be obtained by summing or integrating, in the continuous case, over the strata.

C. *Inferring the presence of a mixed population—hernia recurrence example*

Losses from a mixed population will be skewed in a systematic manner: members with continuing high probabilities of loss will obviously be included disproportionately among early losses. An example from the field of surgery provides an intuitive feel for the change over time in the composition of a mixed population. A hernia can be repaired through a surgical operation called herniorrhaphy. If the abdominal tissues rupture and allow the intestine to protrude again, the hernia is said to recur. Neuhauser [27] has summarized data from several studies of recurrences after first herniorrhaphy representing roughly 600 cases; freedom from recurrence is measured from the herniorrhaphy. From his summary, we computed the proportion of patients free of recurrence at the end of 1-, 5-, and 10-yr intervals, as shown in Table 1. The population under study was all individuals still free of recurrence. This is a first-failure analysis, with hernia recurrence analogous to death in more traditional demographic studies.

Heterogeneity of patients with respect to their per-period probability of recurrence offers a simple and consistent explanation for the observed pattern. To estimate the first-recurrence rate in each time interval, we look at the ratio of survivor proportions at the beginning and end. Thus, for the interval years one to five, the calculation is:

$$\text{Annual rate} = 1 - (0.935/0.970)^{1/4} = 0.0091.$$

These rates decline with increasing length of time after surgery—the result that would have been predicted by a model with heterogeneous recurrence probabilities.*

*Chance is unlikely to be the explanation for the decline in recurrence rates; given constant recurrence probabilities, there is less than one chance in 1000 that departures of this magnitude would be observed. We might think that the probability of recurrence for each patient falls over time as tissues heal and strengthen after surgery. Our surgical colleagues suggest that this explanation is unlikely after the initial weeks of healing. At first, sutures provide protection; then scar tissue forms and the wound gains maximum strength within a few weeks, not a few years.

TABLE 2. DEMONSTRATION OF DECLINING RECURRENCE RATES IN A POPULATION WITH TWO RISK GROUPS

Interval after surgery (yr)	Number of patients free of recurrences at end of interval			Rate of first recurrence during interval*
	Low-rate group (0.01/yr)	High-rate group (0.25/yr)	Total	
0	100	100	200	Not applicable
0–1	99	75	174	130
1–5	95	24	119	91
5–10	90	6	96	42

*Recurrences per 1000 person-years among persons still free of recurrence.

The most basic model—deliberately simplified to highlight the qualitative characteristics of heterogeneity—posits two types of patients receiving herniorrhaphy, distinguished by their probability of a first recurrence in any given year. (This structure is equivalent to a discrete form of the 'mover–stayer' model [2].) Patients in the low-rate group have a 0.01 chance of having a recurrence in any given year. High-rate group patients have a 0.25 chance. (For this analysis we will ignore the possibility of death; as noted earlier, it could be incorporated using models of competing risk [23].) Consider a population consisting of 100 individuals of each type. After 1 yr, the number of patients in the low-rate group still free of recurrences is $100 \times (1 - 0.01) = 99$. In the high-rate group, $100 \times (1 - 0.25) = 75$ have escaped a recurrence. The recurrence rate during the initial year is $1 - (99 + 75)/200 = 130/1000/\text{year}$. For the 4-yr period, 1–5 yr after surgery, the comparable annual rate is $1 - (119/174)^{1/4} = 91/1000$.

The data and rates in Table 2 have the qualitative characteristics that Neuhauser observes.

In our applications section, we extend this analysis using a gamma distribution to define a continuous spectrum of annual recurrence rates across the population.* This recurrence example reveals the potential magnitude of heterogeneity within a population in one instance, and shows the way its existence could be inferred and estimated.

If this model is construed more broadly as one in which individuals have varying risks of dropping out of the population, then it offers a qualitative result that has widespread application. Any such population will 'improve' itself as it differentially eliminates individuals at high risk. The greater the absolute level of risk and the more substantial the variation in probabilities among risk groups, the more noticeable will be this selection process, hence the more rapid the 'improvement' of the population. This property has made mixed-population models useful in analyzing migration [4], labor mobility [28] and reliability [29].

D. *Interventions in mixed populations*

One of our primary interests here is to estimate the benefits of interventions intended to lower the dropout rate, when applied to mixed populations. Phenomena related to the selective survival pattern just noted play a significant role whenever there is a relationship between the benefits an individual receives from the intervention and his preintervention risk. Commonly, such an intervention offers its greatest benefits to those who are at highest risk.† If a tutorial program instituted to prevent freshmen from flunking out from college is, as we would expect, most beneficial to those whose probabilities of flunking were greatest, the sophomores next year will have a greater proportion of

*The distributions must be written in terms of instantaneous recurrence rates, sometimes called 'hazard rates'. The observed annual rate of recurrences, q, will be related to the hazard rate, μ, by $q = 1 - e^{-\mu}$. The more familiar term 'force of mortality' is the corresponding instantaneous mortality rate.

†The opposite situation could apply. Airline safety measures, for example, provide benefits to passengers who are substantially wealthier and somewhat healthier than the average citizen, and probably at lower risk of death than the average for their age.

'previously endangered' students. If there is a correlation between risks in freshmen and sophomore years, the survivors will be on average a weaker group. Hence, there will be a greater proportion of sophomores flunking. The intervention will prove to have offered less benefit than might have been inferred from a naïve headcount of successful passing freshmen.

We have observed the phenomenon of an intervention weakening a population in a simulation of the effects of various motor vehicle safety programs [8, 30]. We found that air bags would have a qualitatively similar effect on the composition of the driving population. Since air bags reduce the fatality rate of motor vehicle accidents, they differentially benefit heavy drinkers. If heavy drinkers have an accident rate 10 times normal, air bags would increase their share of the male population from a presumed 2.6 to 2.8%. The change is slight, since motor vehicle accidents account for less than 3% of deaths.

The benefits of an intervention that lowers risk in the short run may in part be offset by a worsening of the population at risk in the future. Traditional assessments, generally overlooking this factor, will overestimate the benefits of many health-promoting interventions. The more an intervention concentrates benefits in a subgroup at continuing high risk, the more distorted will be a traditional assessment. At an extreme, we constructed a hypothetical example in which all motor vehicle fatalities are attributed to a tiny group of extremely reckless drivers. We assumed that air bags would cut the fatality rate of accidents in half. A traditional assessment would suggest that air bags could increase life expectancy by 0.3 yr; but if the increasing prevalence of reckless drivers is recognized, the gain would be calculated as only one tenth as large.

2. PROCEDURES FOR ASSESSING SURVIVAL CHANGES FROM INTERVENTIONS

We wish to examine the performance of interventions, such as medical procedures or safety measures, in reducing probabilities for the event, 'loss from a population'. Usually, we shall focus on an intervention first applied at a particular chronological time to individuals of a specified age and sex. The loss from the population may represent an actual death, or merely the onset of some illness or condition.

We employ a *risk model* to provide a mathematical statement of the probability of experiencing some event in an interval of time. Risk will be a function of background variables such as age, smoking habits and time, as well as the policy intervention. An *intervention* is a program intended to affect risk. The baseline situation is also referred to as an intervention. A *risk group* is a homogeneous stratum within a population; its surviving members of the same age bear identical risks. Risk group categories are mutually exclusive and exhaustive.

A. *The standardized assessment—the recommended procedure*

Classification of the population into homogeneous risk groups will enable us to compute unbiased assessments of the performance of an intervention. The recommended procedure comprises five steps:

(i) Divide the population into homogeneous strata. In principle, any number of strata are possible on any number of stratification variables (e.g. age, sex, blood pressure.) Stratification variables may be qualitative or quantitative; the latter may be discrete or continuous.* In practice, two discrete strata often suffice to capture much of the variability in a mixed population.

(ii) Indicate the prevalence of each stratum in the population at the age the intervention would begin, i.e. its proportion of the initial age cohort.

*Thus, with continuous variables, the number of strata is infinite. In the analysis that follows, however, only one continuous variable besides age will be considered at a time. Stratification in our model bears an analogy to the subdivision of disease states in a Markov model. This subdivision, based on differences in prognosis, has been applied to distinguish dialysis patients in their first year from those in subsequent years [31], and heart attack patients according to the presence of a previous attack [32].

(iii) Estimate the age-specific loss rate (mortality) from the population in each stratum with or without each intervention at each age.

(iv) Compute the outcome measure(s) of interest (such as life expectancy, survival to particular ages or duration of freedom from disease) separately for individual strata.

(v) Compute the overall outcome measure for the population by averaging the measures for the strata according to the relative sizes of the strata. (Within a homogenous population there is only one stratum; steps (3) and (4) are sufficient to compute outcome measures.)

Classification into risk strata can be a sophisticated process. The crux is judging when classification has proceeded far enough to achieve sufficient homogeneity within risk groups, in the sense that within a single risk stratum and age, all members derive nearly the same expected benefits from a proposed intervention. A refined stratification may prove difficult if, beyond obvious variables such as blood pressure and smoking status, there is insufficient evidence to indicate which classifying variables are important.

Stratification variables might be single medical variables or multivariate risk scores which modify the effect of a proposed intervention [33]. They might also include lifestyle characteristics, such as smoking and drinking habits. Socio-economic variables are often helpful proxies for both environmental and lifestyle variables. For example, cervical cancer has a higher incidence among women living in less hygienic conditions [34]. These are all observable variables.

Many factors that determine future mortality risks may be unobserved [35, 36]. Some may become evident as further events unfold; others may never come to view, though their presence might be inferred through statistical experiments. In assessing the gains from an intervention over time, unobserved factors are as important as those that can be identified and must be taken into account when individuals are classified into risk categories. Frequently this will be done on an *ad hoc* basis. For example, if it is known that among individuals with common background variables, some confront substantially higher cardiac risk than others, one might use an additional classification variable with two categories: high cardiac risk and normal cardiac risk [36]. When a population is characterized by an unobserved risk factor, we cannot estimate the risk for an individual. Often we can, however, infer the statistical distribution and evolution of risks for the population as a whole. This distribution is sufficient for a standardized assessment.

Although our proposed procedure of classifying the population into risk groups may prove difficult in any particular circumstance, it has the redeeming virtue, as we shall shortly show, of avoiding systematic biases inherent in traditional modes of assessment. Three paradigms—the simple extrapolation, the traditional assessment and the standardized assessment—will help us make our points precisely.

B. *The simple extrapolation—without longitudinal data*

The simplest estimate of the benefit of an intervention—what we refer to as the simple extrapolation—assumes the population is homogeneous and the loss rate for each individual is constant over time. Under this assumption, the aggregate population will have a constant loss rate equal to the initially observed loss rate. In the example of hernia recurrences discussed above, the proportion of the population free of recurrence at 10 yr, given a first-year rate of 30/1000 persons, would be calculated as $(1 - 0.030)^{10} = 0.74$. Because the actual rates of recurrence decline, as shown in Table 1, the true proportion is 0.919. Even in the absence of an intervention, the simple extrapolation may be a poor predictor of survival over time.

C. *The traditional assessment—with longitudinal data available*

Epidemiological data like those in Table 1 often reveal that rates of death or other losses vary over time. In such cases, the simple extrapolations described above are clearly inadequate to describe survival even under the baseline intervention. Many careful inves-

tigators then use what we call a 'traditional assessment', which posits a pattern of losses for the baseline intervention that varies with age or time according to some empirically observed pattern (e.g. mortality rates from some standard life table) or according to some statistical model. Losses or deaths under the active intervention are assumed to follow the same age and time patterns, but to reflect an additional intervention effect as well. While risk groups are not recognized explicitly, the adjustments for age and time implicitly allow for the changes in the mixture of risk groups over time in a manner comparable with the standardized assessment. We define these procedures more rigorously in the next section.

A traditional assessment of the benefits of an intervention starts by computing age-specific mortality (loss rates from the population) in the presence of the intervention, assuming that the current population mix at each age is maintained. This assessment computes survival, life expectancy or some other output measure employing these new mortality rates. *A traditional assessment, in other words, implicitly assumes that the prevalence of different risk groups at future times with the active intervention will be the same as it would have been under the baseline intervention.*

To our knowledge, all the published calculations on the benefits of removing some particular cause of death are traditional assessments: they assume the risks of other causes remain unchanged, which is equivalent to assuming that the persons who benefited from the removal of one cause of death were at average risk for their age and sex from other causes.

The assumption that an intervention will not change prevalence rates strikes us as highly unrealistic for most circumstances. In an example we have developed at length elsewhere, we showed that the continued availability of a mobile cardiac care unit in a community will increase from 12 to 15% the proportion of men alive at age 75 who have had a previous heart attack [8]. This represents an increase in the prevalence of a group at high risk for future cardiac events, and no doubt other conditions as well. A beneficial intervention thus provides the greatest absolute benefit to those groups at highest risk, and may have the long-term effect of increasing the prevalence of high-risk persons. Traditional assessments, overlooking this factor which offsets short-term benefits, will overestimate the value of such an intervention.

D. *Formal concepts for assessing interventions*

Consider a mixed population where the subscript i denotes interventions and j indexes risk groups. The initial prevalence or proportion of risk group j at the initial age is r_j. The *hazard rate*, μ_{ij}, also referred to as the incidence density or loss rate, is the instantaneous probability of the dropout event per unit time. When the event is death, μ_{ij} is called the force of mortality.

The *survival function* gives the probability of surviving from the initial age, at which the model begins, to some specified age, under a given intervention. If we let $\mu_{ij}(x)$ denote the hazard rate, then survival to age x will be:

$$l_{ij}(x) = \exp\left[-\int_0^x \mu_{ij}(t)\mathrm{d}t\right] \text{[37]}. \tag{1}$$

Notice that at $x = 0$, $l_{ij} = 1$. The mixed-population's survival at age x, denoted by $l_{i\cdot}(x)$, is a weighted average of $l_{ij}(x)$:

$$l_{i\cdot}(x) = \sum_j r_j l_{ij}(x). \tag{2}$$

The *prevalence*, or proportion of survivors, of risk category j under intervention i at age x, $r_{ij}(x)$, is:

$$r_{ij}(x) = r_j \frac{l_{ij}(x)}{l_{i\cdot}(x)}.$$

The overall hazard rate under intervention i at age x, $m_i(x)$, is a weighted average of the

hazard rates for the individual risk groups weighted by their respective prevalences; that is:

$$m_i(x) = \sum_j r_{ij}(x)\mu_{ij}(x). \tag{3}$$

If this rate is small, it will closely approximate the ratio of expected events (e.g. death or recurrences) at age x to the population attaining that age.

The three *outcome measures* of a health intervention that are most commonly employed are change in survival for a specified interval, say 5 yr; change in the mortality rate at a given age; and change in life expectancy. In some circumstances, alternative measures may be appropriate, such as period of freedom from a disease, expected time until relapse or 'quality-adjusted life years' (QALYs). In what follows, let $i = 1$ denote the baseline intervention of carrying on as usual, and let $i = 2$ denote the new program or intervention whose benefits are to be computed.

The three outcome measures can be written formally using the terminology defined above. For survival over a specified interval, suppose that the interval begins at age 0 and ends at age x.* Then the increase in survival is:

$$l_{2\cdot}(x) - l_{1\cdot}(x).$$

The change in mortality at some arbitrary age x is:

$$m_2(x) - m_1(x).$$

(A beneficial intervention will reduce mortality and yield a negative difference above.) The life expectancy of the mixed population under intervention can be represented as a weighted average of the life expectancies of the risk groups under that intervention, or as:

$$\mathring{e}_{i\cdot} = \int_0^\infty l_{i\cdot}(x)dx.\dagger$$

The increase in life expectancy from an intervention is thus:

$$\mathring{e}_{2\cdot} - \mathring{e}_{i\cdot}.$$

E. *Comparison of traditional and standardized assessments*

Our standardized assessment stratifies the population into subgroups homogeneous with respect to risk of the event under study. A 'traditional assessment', by contrast, treats a population aas a single group, all of whose members are presumed subject to the same risks. This section formalizes the definition of a traditional assessment in terms of the concepts outlined above, and compares traditional and standardized assessments.

1. *Formalized definition of the traditional assessment.* In a traditional assessment, unrecognized substrata are implicitly weighted by their prevalence under the baseline intervention. The prevalence of risk stratum j at age x under the baseline intervention is $r_{1j}(x)$ defined by equation (3), so the aggregate risk under the traditional assessment for the baseline intervention is:

$$m_1(x) = \sum_j r_{1j}(x).$$

2. *Bias in a traditional assessment of an intervention.* For the baseline intervention, the weighting under the traditional assessment is the same as for the standardized assessment. Therefore, the estimate of risk, and all measures that derive from it, will be identical for both assessments. The difference arises under the innovative treatment.

*We may interpret age zero as the initial age for our analysis, the age from which the proposed intervention is being modeled. This origin need not be birth; its designation as age zero is simply a matter of convenience.

†Remaining life expectancy is usually defined as the expected number of years until death:

$$\int_0^\infty x\mu_{ij}(x)l_{ij}(x)dx.$$

Integrating by parts and using the fact that $l_i(\infty) = 0$ gives the more convenient formula in the text.

Failing to recognize the existence of substrata, the traditional assessment continues to weight by the same relative prevalences as under the baseline intervention. Thus, it estimates risk at age x under the active treatment (treatment 2) as:

$$m'_2(x) = \sum_j r_{1j}(x)\mu_{2j}(x).$$

This expression differs from that for the standardized assessment because different—and incorrect—prevalence rates have been used. The inaccurate survival function derived from the traditional assessment is:

$$l'_{2\cdot}(x) = \exp\left[-\int_0^x m'_2(t)\,dt\right].$$

We are interested in the difference between a traditional and a standardized assessment. We term this difference $\Delta m(x)$ at age x and define it by:

$$\Delta m(x) = m_2(x) - m'_2(x). \tag{4}$$

A positive difference means that true mortality is higher than the traditional estimate, and that the benefit of an intervention has been overstated.

3. *Direction and magnitude of bias.* Only under special circumstances, when one of two particular conditions is satisfied, will the traditional and standardized assessments be identical, i.e. will $\Delta m(x)$ be zero. Those conditions are:

(i) The innovative treatment lowers risk by the same constant amount in each risk group. In other words, $u_{2j}(x) - u_{1j}(x)$ is constant for all risk groups j at all ages x. This means that the division into risk groups is irrelevant with respect to the intervention. Conceivably this condition might apply for reductions in death risks related to nuclear wars.

(ii) The intervention reduces all risks to the same level. That is, $u_{2j}(x)$ is constant for all risk groups j and for all ages x less than some designated X. This condition would apply if existing differences in risk were due to a single factor which was eliminated by the intervention. Such an intervention, for instance, might be a vaccine against an infectious disease for which susceptibility and fatality are uncorrelated with other risks.

Except in these special cases, the traditional assessment will give a biased mortality estimate. The bias will be indicated by a nonzero value for $\Delta m(x)$. To help define its sign and magnitude, we define comparative survival gain as:

$$g_j(x) = \frac{l_{2j}}{l_{2\cdot}} - \frac{l_{1j}}{l_{1\cdot}}.$$

Thus, g_j measures the gain in survival to group j relative to the overall gain in survival from treatment. It is highest (and positive) for the risk group that gains most, and lowest (and negative) for the group that gains the least from treatment. An important result, which we prove elsewhere [30, 36], is that this mortality difference has the simple expression:

$$\Delta m(x) = \text{covariance}\,[\mu_{2j}(x), g_j(x)]. \tag{5}$$

Here the products and means in the covariance are weighted according to r_j, the prevalence at the starting age. An equivalent expression that writes $\Delta m(x)$ as a weighted crossproduct of $\mu_{2j}(x)$ and $g_j(x)$ is:*

$$\Delta m(x) = \sum_j r_j \mu_{2j}(x) g_j(x).$$

If an intervention benefits groups at high risk most, then the covariance pairs larger values of g_j with larger values of $\mu(x)$, so the covariance is positive. (The covariance always has the same sign as a correlation coefficient.) Thus, the bias is positive—the

*These expressions are equivalent because $g_j(x)$ has a mean of zero when weighted by r_j.

standardized mortality is higher than the mortality computed using the traditional assessment. The direction of bias can be summarized as follows:

*If an intervention is more beneficial to the high-risk group, in the sense that the amount that it subtracts from the mortality force of high-risk persons is more than it subtracts from the force for low-risk persons, then the traditional method will overestimate the benefit of the intervention.**

In other words, the traditional method will compute mortality rates under the intervention that are too low, and estimates of life expectancy or effective life expectancy that are too high.

The simple extrapolation suffers from the same direction of bias as the traditional assessment, and the magnitude of bias may be much greater. Starting with the baseline data of our hernia-recurrence example, assume that there was an intervention that cut each individual's per period loss probability by 50%, and employ as a benefits measure the increase in the number of patients not having a recurrence within 10 yr. The simple extrapolation, mistakenly assuming that loss rates would stay constant into the future, would estimate that with the intervention 86.0% would survive in contrast to 73.7% without it, or a gain of 12.3 percentage points. For the traditional assessment, the figures would be 96.7, 91.9 and 4.8%. The correctly computed standardized assessment would give 93.5, 91.9 and 1.6%. The simple extrapolation produces by far the greater bias, as it always will with constant loss rates. (Keyfitz and Littman [38] analyze further the bias of the simple extrapolation in mortality studies.) In what follows, we shall restrict our attention to traditional and standardized assessments.

Most familiar interventions confer the greatest benefit on high-risk persons. Automobile safety measures do more for the accident-prone than for others. The elimination of cardiovascular–renal disease would lower mortality more among older persons than younger ones.

4. *Bias over time.* Note from the expressions for $\Delta m(x)$ that the magnitude of the bias will be time-dependent. Interestingly enough, the bias is greatest at intermediate time intervals. At time zero, there is no bias, for there has been no opportunity for selective mortality to affect the composition of the population. That is, $g_j(x)$ is zero since all survival functions, l, are unity. Naïve extrapolations of benefits from an intervention after a long period of time would not be biased, since, with or without the intervention, virtually all survivors are by then low-risk individuals. For intermediate time intervals, where selective survival bias operates, but low-risk individuals are not overwhelmingly predominant, the bias will be greatest. The next section presents a numerical example where, as expected, bias is greatest for intermediate time periods.†

5. *Bias under two widely used models of mortality.* The *multiplicative model* is sometimes termed a model of preventive or etiological independence, since the effects of the intervention are assumed to be independent of other determinants of risk. The model states that for every risk category and at every age, risk under the active intervention is a constant multiple of baseline risk. In epidemiology and biostatistics, the constant multiple is termed the risk ratio; in life insurance, it is called the loading factor. Cox's [6] widely used regression model for analysis of failure times is a generalization of the multiplicative model. As the multiplicative model can also be written as:

$$\mu_{ij}(x) = h(x)\exp(\alpha + \beta_i + \gamma_j),$$

where $h(x)$ is a function of age, and α, β_i and γ_j are parameters, it is also termed the exponential hazard model.

The *logistic model*, applied extensively in cardiovascular epidemiology, assumes that

*The formal proof of the sign of this bias requires that there be no age for which the low-risk group benefits more. Specific numerical computations would be required to determine the direction of bias when sometimes those at high risk benefit more, other times those at low risk.

†This discussion was based on discrete risk groups. With continuous risk groups, other patterns are possible; for example, the relative bias may increase asymptotically with time. Bias will always be zero at time zero.

the probability of an event during a unit interval of time, q_{ij}, is a logistic function of intervention, risk group and age [33]. In contrast to the multiplicative model, which applies a constant factor to the instantaneous probability of death, the logistic model applies a constant factor, called the odds ratio, to the odds of death in a discrete time interval. The two models are practically identical except when risks during the interval are very high (above 0.1). For each of these two models of risk, the covariance in (5) above is positive [36]. Whichever model is used, traditional assessments of mortality under other than baseline conditions will be biased downwards. Benefits of useful interventions will be overstated.

F. *Numerical examples of bias*

1. *Compliance with medical therapies.* A simple example shows the bias inherent in traditional estimation methods. The higher the rate of dropping out (analogous to rates of mortality in traditional population studies), the more pronounced the effect. Problems involving compliance with medical therapies provide compelling examples in which loss rates are substantial within the time frame of most studies.

Haynes and Taylor's [39] review of the compliance literature found that compliance rates for medical treatments initiated by providers are of the order of 50% after 1 yr. The review identified only a few factors consistently predictive of compliance, such as a patient's belief about the seriousness of the condition to be avoided and the effectiveness of the recommended preventive action. Assume that an index of these factors is employed to divide the population into two groups. Under baseline conditions, the 'believers' ($j = 1$) have an incidence density (or hazard rate) of 0.2/person-year. The 'non-believers' ($j = 2$) are much more likely to drop out of the population; indeed their hazard rate is 1.0/person-year, five times as high. Initially, half the population is in each category.

An intervention designed to reduce drop out rates is brought to the population. It might consist of telephone contacts, or payments for compliance. The intervention is expected to cut drop-out rates in half to 0.10 and 0.50/person-year, respectively. Rates for survival in treatment at 1 yr are lowered from 0.59 under baseline conditions to 0.76 with the intervention. These magnitudes are consistent with the literature.

We now use our equations to calculate survival, prevalence and loss rates at 3 yr. The upper part of Table 3 shows the prevalence of the two risk groups under alternative treatment conditions. Under baseline conditions, the annual loss rate at 3 yr, $m_1(3)$, is 0.26 ($0.92 \times 0.2 + 0.08 \times 1.0$). The intervention lowers this mortality to $m_2(3)$, 0.19, according to the standardized assessment ($0.77 \times 0.1 + 0.23 \times 0.5$). However, the traditional assessment would use the old weights, and compute a mortality of $m'_2(3)$, 0.13 ($0.92 \times 0.1 + 0.08 \times 0.5$). *The traditional assessment is 31% too optimistic.* The traditional assessment fails to note that the intervention enriches the mix of nonbelievers among the survivors. A good indication of the way the selection effect dilutes the benefit of the intervention is the fact that the aggregate loss rate at 3 yr, appropriately measured, is only 28% below the initial loss rate, despite the fact that the intervention initially cut the loss rate by one half, and the loss rate for a surviving member of either group has been cut by one half.

TABLE 3. EFFECTS OF HYPOTHETICAL INTERVENTION TO BOOST COMPLIANCE

Characteristic	Baseline intervention	Active intervention: Traditional assessment	Active intervention: Standardized assessment
Prevalence at end of third year			
Believers, $r_{i1}(3)$	0.92	0.92	0.77
Non-believers, $r_{i2}(3)$	0.08	0.08	0.23
Rates for combined group			
Dropout rate after 3 yr, $m_i(3)$	0.26	0.13	0.19
Survival to third year, l_i (3)	0.30	0.55	0.48
Mean survival $\mathring{e}_i$	3.00	7.42	6.00

The time pattern of bias exhibits the properties discussed in the previous section. For extremely short periods, selection will not yet have operated. For very long periods, selection will no longer be relevant because virtually all survivors will be believers. At these extremes, there will be no bias. For intermediate periods, the selection effect strongly offsets the benefits of the intervention. Computations with equation (4) reveal that the bias as a percentage of the true value reaches a maximum at 3.5 yr. Traditionally estimated losses will be 0.123/yr; true losses will be 0.179/yr, an error of 31.4%.

TABLE 4. GAIN IN LIFE EXPECTANCY IN YEARS FROM HYPOTHETICAL TREATMENT CONSIDERING SINGLE RISK FACTOR, AND PERCENTAGE OVERESTIMATION IF RISK FACTOR IS IGNORED

Odds ratio for treatment	Odds ratio for risk indicator 1.2	1.6	2.0	4.0	10.0
1.2	2.73	2.69	2.36	2.36	1.90
%	0.4	2.4	5.1	18.3	42.5
1.6	7.02	6.93	6.81	6.14	4.95
%	0.4	2.7	5.7	19.8	44.8
2.0	10.25	10.14	9.98	9.10	7.34
%	0.5	3.0	6.3	20.8	45.9
4.0	18.99	18.92	18.79	17.82	14.92
%	0.7	4.3	8.2	23.1	45.8
10.0	29.98	27.02	27.06	26.81	24.45
%	1.3	6.1	10.6	20.4	41.5

2. *Effects of risk factors.* Our second example concerns loss rates that increase with time, a pattern observed for mortality rates above age 30. We employ the logistic model in one of its traditional areas of application to model the survival of a population with multiple risk indicators. The presence or absence of the intervention enters the mortality equation in a manner parallel to other risk factors and age.

Since any number of potential risk factors could function as confounding variables, a researcher must inevitably exercise judgment in choosing which factors to include and in estimating the magnitude of possible bias from variables excluded. To assist in these decisions, Table 4 shows the correctly computed gain in life expectancy for various hypothetical treatments with one additional binary risk factor. The initial prevalences of the high- and low-risk categories are 50%. The calculations assume that there are constant odds ratios relating to differences in risk category and treatment. An odds ratio of two for treatment, for example, would suggest that an individual in any risk category who did not receive the treatment would have a probability of dying within a year two times greater than if he did. Table 4 assumes that mortality rates can be represented by a three-variable logistic model similar to the multivariate logistic models estimated in the Framingham Study [40] and the Pooling Project [41]. The three-variable model gives the probability of death in a year as:

$$q_{ij} = 1/(1 + \exp\{-[A + Bx - C(i-1) + D(j-1)]\}), \tag{6}$$

where x denotes age, i treatment group (with $i = 1$ being untreated) and j risk factor (with $j = 2$ being high risk). The coefficients, C, is a measure of the strength of the treatment, in terms of the natural logarithm of its odds ratio. Similarly, the coefficient, D, describes the risk factor in terms of the logarithm of its odds ratio. The only effects of age on the intervention or risk factors are represented by an additive term in years of age. The age coefficient, B, was set at 0.037931, its value as determined from a Framingham multiple-risk-factor equation. The constant, A, was set so that the annual mortality in the untreated group at age 50, $m_1(x_0)$, would equal 0.00894, the rate published in a life table best representing the Framingham population (white males in the New England census division) [42]. Thus, A is constant within each column, but decreases as we move to the

right. The percentage below each entry in Table 4 shows the amount by which the gain in life expectancy would be overstated if the risk factor were not considered.*

Elsewhere, we have extended the techniques for modeling mixed populations to deal with changing loss rates over time (using the Gompertz function to capture increasing mortality with age) and to allow for continuous distribution of risk groups. Many risk factors are continuous—physiological factors, such as blood pressure; environmental exposures, such as pollution concentrations; and unhealthy lifestyles, such as physical inactivity. A rich and manageable family of mathematical models based on gamma distributions enables us to break down observed loss rates in a population into heterogeneity effects, which lower observed rates for a population over time, and individual loss rates, which evolve with time in a variety of systematic manners. In these more complex models, the relative error in mortality from traditional assessments is roughly proportional to: the number of years since the intervention was first applied, the relative benefit of treatment and the variance in risk divided by the initial risk assuming no intervention [43].

Still further generalizations may be appropriate. The assumption that an individual always remains in the same risk group could be relaxed [3], allowing for changes in underlying risk based on random disturbances or past experience. Alternative functional forms may be utilized to represent dependence on age or time. With actual data sets, likelihood ratio tests may permit comparisons of the goodness of fit of alternative specifications.

3. APPLICATIONS

We consider four applications of our methodology. The first estimates the magnitude of the traditional assessment bias in estimating the benefits of a hypertension control program; the second illustrates the use of continuous risk groups to model hernias and other problems with recurrences; the third applies the principles of mixed populations to a central problem of demography, assessing mortality and constructing life tables; the fourth examines interpretations of relative survival data given a heterogeneous population.

A. *Control of hypertension*

We wish to estimate the improvement in survival and reduction in mortality from the control of hypertension (high blood pressure). For the sake of illustration, we define 'hypertension' to mean a diastolic blood pressure of 110 mmHg and 'controlled' to mean 90 mmHg. Although randomized trials have established that antihypertensive therapy is effective, the populations in these experiments are too small and specialized to provide a basis for quantitative estimates of the long-term effects of blood pressure control in the general population [44]. Therefore, at least at present, one must rely on existing epidemiological data as a guide. A traditional assessment would calculate mortality and life expectancy for controlled (new intervention) and uncontrolled (baseline intervention) hypertensives as if hypertensives of the same age and sex were a homogeneous risk group. (This approach formed the basis of a recent major policy study of hypertension [44].)

Studies of cardiovascular disease support the hypothesis that the effects of risk factors are roughly multiplicative and are well represented by the logistic model [38]. Thus, blood pressure control would reduce risk most in persons already at high risk. Our analysis begins with a population of 50-yr-old males who would be hypertensive unless controlled. To disaggregate this population, we classify the men by cigarette-smoking habits, the strongest cardiovascular risk factor after age and hypertension itself. For simplicity, we consider this factor in binary form in a manner that divides the initial

*The percentage overestimation is a single-humped function of the initial prevalence. When the prevalence is 0 or 100% (i.e. the population is homogeneous) the percentage error is zero. The maximum error occurs when the prevalence is between 30 and 70% for the range of odds ratios considered above. The gain in life expectancy is practically invariant with respect to changes in the prevalence.

population into two equal-sized risk categories of 'heavy smokers' (here defined as individuals who smoke a pack or more a day*) and 'others' (light and non-smokers). In future years, the surviving cohort will continue to be categorized by its smoking status at age 50.† An equal proportion of each risk group receives the intervention 'blood pressure control'. Together, these men are termed the controlled group. The remainder of each risk group remains hypertensive. Figure 1 shows the age-specific annual mortality rates for the controlled group estimated from the Framingham risk coefficients and vital statistics for New England. As smoking is hazardous to survival, the proportion of heavies declines in both the hypertensive and controlled groups. This selection is

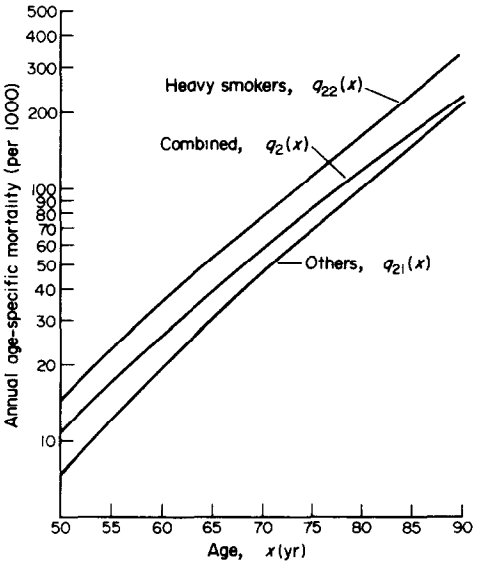

FIG. 1. Mortality ratios of controlled group.

reflected in Fig. 1; with increasing age, the combined line approaches the others line. Under the multiple-risk-factor hypothesis, smoking is more lethal for hypertensives. Thus at any age beyond 50, the proportion of smokers among the survivors will be lower in the hypertensive group than the controlled group, as is shown in Table 5.

TABLE 5. COMPUTED PROPORTION OF HEAVY SMOKERS AMONG SURVIVORS, $r_{i2}(x)$

Treatment	Age, x 50	70	90
Hypertensive, $i = 1$	0.500	0.383	0.105
Controlled, $i = 2$	0.500	0.415	0.138

Table 6 reveals that blood pressure control increases life expectancy in the combined group from 18.65 to 21.17 yr, a gain of 2.52 yr, making appropriate classification for smoking status. A less sophisticated approach that did not stratify by smoking habits would neglect the increased proportion of smokers among hypertensives surviving due to blood pressure control. The gain would then be estimated at 2.63 yr, some 4.4% too high. A more sensitive outcome measure is mortality at advanced ages. At age 75, for example, blood pressure control is predicted to reduce annual mortality by 80/10,000. The traditional assessment would predict a decrease of 93/10,000, some 16% larger. At older ages, the percentage error is even greater.

*The cutoff level of one pack per day was chosen to divide the population in half equally and to be consistent with the Framingham population over the period these data were gathered (late 1950s to early 1970s).

†This simplifying assumption is consistent with the finding that, at least as far as respiratory physiology is concerned, stopping smoking does not yield immediate benefits [45].

TABLE 6. PREDICTED IMPROVEMENTS IN LIFE EXPECTANCY AND MORTALITY FROM TREATMENT OF HYPERTENSION

Outcome measure	Baseline level	Assessment of change: Standardized	Assessment of change: Traditional	Bias	Percentage bias*
Expectation of remaining life (yr)	$\mathring{e}_1$	$\mathring{e}_2 - \mathring{e}_1$	$\mathring{e}'_2 - \mathring{e}_1$	$\mathring{e}_2 - \mathring{e}'_2$	
At age 50	18.65	2.52	2.63	−0.011	−4.4
Annual mortality	m_1	$m_2 - m_1$	$m'_2 - m_1$	Δm	
At age 65	0.0450	−0.0041	−0.0046	0.0005	−13
At age 75	0.1066	−0.0080	−0.0093	0.0013	−16
At age 85	0.1850	−0.0161	−0.0194	0.0033	−21

*Bias as a per cent of change according to standardized assessment. For expectation of remaining life, it is $(\mathring{e}_2 - \mathring{e}'_2)/(\mathring{e}_2 - \mathring{e}_1)$; for mortality, it is $\Delta m/(m_2 - m_1)$.

These modest differences are the effect of only a single, observable risk factor, categorized into only two levels. If smoking behavior had been divided more finely, i.e. by separating those who don't smoke at all from light smokers, the estimated benefits of the intervention would be lowered further.

More importantly, a complete analysis should include other risk factors or indicators. These other factors should include possibly unobservable variables as well, such as coronary-prone behavior patterns, willingness to adhere to dietary changes and the like. The effects of such other risk factors would be roughly additive. If there were a number of such factors, or even one or two important ones, the traditional assessment's overstatement of gains, even correcting for smoking behavior, could be even more significant than in the example given above.

B. *Hernias and other problems with recurrences*

We return to the issue of recurrence of hernias which we considered qualitatively at the beginning of this paper. We assume that for a single individual the rate of recurrence is fixed over time, but the rates among individuals in the population are gamma-distributed. The gamma distribution that best fits the summary of Neuhauser's data reproduces the survival rates in Table 1 to three decimal places.* The individuals in the population at time zero, right after the completed herniorrhaphy, have a mean instantaneous recurrence rate of 0.0552/yr; the standard deviation among their rates is 0.3296. The distribution of recurrence rates is highly skewed to the right; that is, the bulk of the patients have low rates, but a few have quite high rates. Indeed, the risk at the 90th percentile is 100 times as great as at the 10th percentile. The analysis suggests that substantial variability in risk of recurrence offers a satisfactory explanation of the observed pattern of declining rates over time. The strong clinical implication is that study of possible correlates of risk such as weak abdominal tissue, chronic cough or imperfect surgical repair could prove highly beneficial. Investigations may suggest ways to lower these risks or, alternatively, may recommend that high-risk patients be given different or supplementary treatments.

The problem of recurrence or recidivism has widespread application outside the medical area. A continuous-risk-factor model yields results that accord well with data on the return to the use of alcohol, cigarettes and heroin [30]. Relapse rates decline over time as the high-risk individuals separate themselves from the population of abstainers. The commonly heard statement that the first few months are the hardest, say for those who have given up smoking, may miss the mixed-population aspect of the problem. The importance of getting a patient through the first few months may not be that his personal relapse probability will decline with time. Rather, if he gets over the first few months it is substantially less likely that he was an individual who had a high monthly relapse rate. Indeed, it is conceivable that each individual's relapse rate rises over time, but that the

*Since the gamma distribution offers two free parameters, α and β, any two of the three survival probabilities could be fit exactly. This spectacularly good fit is partly serendipitous. The fact that we are fitting a cumulative distribution (survival) rather than a density function makes it easier to secure a good fit.

heterogeneity is great enough that the population's relapse rate declines over time. Naive extrapolation assuming that the population is homogeneous would give an impression that was wrong qualitatively as well as quantitatively.

C. *Expected remaining life at selected ages*

An examination of life expectancy data at various ages by race and nationality provides further evidence confirming our hypothesis of variability in risk and selective attrition.

1. *Comparison by race: white and non-white.* The number of expected remaining years of life at selected ages by race, sex and year are given in Table 7. Non-whites of both

TABLE 7. EXPECTED REMAINING YEARS OF LIFE AT SELECTED AGES, BY SEX, RACE AND YEAR*

Year and age	Male		Female		Difference	
	White	Non-white	White	Non-white	Male	Female
1973						
Birth	68.4	61.9	76.1	70.1	6.5	6.0
70	10.4	10.7	13.7	13.2	−0.3	0.5
85	4.7	6.3	5.7	7.3	−1.6	−1.6
1930						
Birth	59.1	47.6	62.7	49.5	11.5	13.2
70	9.2	8.8	10.0	10.3	0.4	−0.3
85	4.0	4.3	4.2	5.5	−0.3	−1.3

*From National Center for Health Statistics [46].

sexes have a lower life expectancy at birth than whites, presumably because of poorer nutrition, medical care and education. With increasing age, the life expectancy of non-whites equals and eventually surpasses the life expectancy of whites for both sexes. (Although these patterns are confounded by possible misreporting of age of death among non-white persons, adjustment for these inaccuracies by reference to census records does not eliminate the crossover [47].)

We can frame the comparison between whites and non-whites using our previous terminology. The 'treatment' is advantages in medical care, nutrition and environment which have been more regularly available to white than to non-white persons. The hypothesized risk indicator is 'constitutional weakness'. The data suggest that 'treatment' may have interacted positively with this unobservable risk indicator. That is, those with constitutional weakness are hurt much more severely by lack of access to health-promoting advantages. With two groups that are otherwise identical, the one with less access to these advantages will initially experience much higher mortality. The gap will narrow over time, and may well reverse itself at higher ages. The scope for selection is ample: out of a cohort of 1000 male births, using 1973 mortality rates, 555 whites but only 406 non-whites would reach age 70, when the two groups of survivors have essentially the same prospects.

There is no need to make an argument as strong as the one that whites and non-whites would have had identical mortality experience had they had identical access to health-promoting advantages. Genetically connected racial differences undoubtedly play a role in determining mortality patterns, though even the direction of possible genetic differences is difficult to assess. What is clear is that the selective-risk-factor model outlined above is consistent with data on racial differences in mortality. We know of no equivalently simple hypothesis that can explain the observed patterns.

2. *A comparison by nation: Sweden and the United States.* Sweden, renowned for its advanced health care system and emphasis on healthy lifestyles, enjoys a greater life expectancy at birth than the United States for both males and females. As Table 8 shows, however, this difference narrows with increasing age, becoming zero at age 76 in males and age 64 in females, and then reverses. Since both countries have highly developed medical care systems and similar major health problems (e.g. coronary heart disease,

TABLE 8. EXPECTED REMAINING YEARS OF LIFE AT SELECTED AGES IN SWEDEN AND THE U.S.A., 1972

Sex and age	Sweden*	U.S.A.†	Difference
Males			
Birth	71.97	67.38	4.59
60	17.68	16.06	1.62
70	10.87	10.39	0.48
80	6.04	6.38	−0.34
90	4.27	4.86	−0.59
Females			
Birth	77.41	75.08	2.33
60	21.06	20.82	0.24
70	13.13	13.54	−0.41
80	7.07	7.92	−0.85
90	4.81	5.83	−1.02

*From United Nations [48].
†Calculated as the ratio of T_x (total person-years remaining) to l_x (survivors) [49].

cancer), it is hard to identify isolated medical factors that would explain this pattern. The hypothesis of heterogeneity again offers a simple and consistent explanation which is compellingly parallel to the one comparing white and non-white patterns of mortality in the United States. Similar patterns are widespread. A comparison of mortality rates in 46 national and racial populations found that 37% of the possible pairs exhibited at least one crossover in age-specific mortality at age 60 and above [50].

3. *Construction of life tables.* An understanding of mixed-population models highlights a conceptual error in traditional means of constructing updated life tables. The usual procedure is to examine present mortality on a cross-sectional basis and assume that without any further medical advance it will persist over time. Thus, a newborn baby today is expected to face the mortality at age 50 that today's 50-yr-olds face at present. But if there has been health progress, reflected by the fact that today's observed cross-sectional mortality rates are lower than those the present 50-yr-old faced at the equivalent ages, then there will be a systematic bias. Assuming, as we have in most of this analysis, a positive interaction between benefits from these improvements and initial mortality, the 50-yr-olds of today have a higher proportion of low-risk individuals than will the 50-yr-olds of 50 yr in the future. Gauging future mortality by looking at today's cross-sectional rates will give an underestimate. This bias will be hard to notice, of course, if we continue to make progress in reducing mortality. Such progress would counteract the bias. The net effect would be that we would overlook the role of mixed populations, and underestimate the extent of progress in reducing age-specific mortality for individuals in particular risk groups. As the population mix effect is more important at higher ages, we would expect the least (or possibly no) reduction in mortality at high ages even if medical and environmental advances benefitted all ages equally.

D. *Relative survival*

The mixed-population model facilitates a proper interpretation of relative survival curves. As noted previously, selective survival can explain the rise in relative survival in colon and intestinal cancer after 5 yr [24]. Heterogeneity in risk has an even broader importance, however. It is commonly thought that the time at which the relative-survival curve between those with and without a disease becomes flat indicates the point at which excess mortality from the disease has ceased. Actually, the slope of the relative-survival curve at any time depends on two effects—the excess mortality in each individual due to the disease being studied and the difference in selection effects due to heterogeneity in risk between the diseased and comparison groups. The two effects would probably counteract each other. A flat or rising relative-survival curve beyond some follow-up time does not necessarily indicate that the excess mortality has ceased, but only that it is offset

by the selection effect. Although the selection effect or improvement in mortality due to variation in risk operates in both the diseased and comparison groups, that effect is usually greater in the former because of its higher absolute levels of risk. Thus, the excess risk imposed by a disease on an individual tends to be greater than would be inferred from the relative survival curve if one incorrectly assumed homogeneity.

Stratification by known risk factors, such as age and race, can reduce the confounding of individual risk levels and the changing composition of the population. But unless the analyzed group and its control are completely homogeneous in risk, the problem cannot be eliminated. Matching to obtain a comparison group that is similar to the diseased group gives an accurate measure of the excess mortality from a disease at first, but cannot standardize for the evolving mixture of risk factors among the survivors as opposed to the control group over time. The problems are analogous to those of the traditional assessment. By stratifying on important risk factors, our standardized assessment for interventions could be adapted to measure the excess mortality of a disease. In place of an intervention that lowers risk differentially according to risk groups, the model would include a disease that increases risk in each stratum. The resulting predictions could be tested, like our previous examples, for accuracy and parsimony in explaining observed relative survival curves.

4. CONCLUSIONS

Users of health statistics are familiar with the fact that risks of death and disease—both total and specific to given causes—vary markedly by age and sex. Accordingly, analyses of the effects of proposed interventions or historical changes commonly disaggregate by those two factors. We contend here that age and sex are only the two most familiar examples of a larger and very common analytical problem—that risk varies from person to person in a consistent pattern over time according to characteristics that may or may not be observable.

Mixed-population models enable us to develop consistent hypotheses that explain otherwise perplexing patterns in such diverse areas as hernia recurrences and international comparisons of age-specific mortality. The composition of a cohort in a mixed population will change over time as selection differentially removes individuals at high risk. This finding has wide applicability. For example, it explains why recidivism or relapse rates in a great variety of policy areas are observed to decline over time.

This observation has important implications for treatment policies. If we accepted the homogeneous-population model, and concluded that individuals' relapse rates declined rapidly over time, we would want to concentrate antirecidivism treatment strongly on the first few months. On the other hand, if we understood that the population was heterogeneous, it might even prove beneficial to delay expensive long-term treatment for a few months, understanding that the missed early lapsers, being differentially high-risk individuals, would have been more likely to relapse ultimately anyway, in spite of treatment.

Frequently for policy purposes we wish to make prospective estimates of the benefits of interventions designed to reduce drop-out rates from populations. Traditional assessment procedures do not consider the influence that an intervention will have on the mix of risk levels within the population at future dates. Yet if, as we would usually expect, the intervention increases most the survival chances of those at highest risk, the mix of the population will change. At any time after some losses have occurred, those who have survived with the intervention will be a weaker group than those who survived without it. A traditional assessment will overstate benefits. An accurate estimate of benefits can be provided only by what is here called a standardized assessment, an assessment that tracks over time the evolving risk level mix within a cohort. The difference between the two types of assessments, i.e. the bias inherent in most actual assessments, can be expressed in terms of the covariance between benefits secured from the intervention and future risk levels.

The concept of mixed populations, as we hope we have demonstrated, provides useful *qualitative* insights into the interpretation of longitudinal studies. In cases where a sub-

stantial portion of at least one risk stratum is expected to experience a loss within the period of analysis, mixed-population models are likely to offer a worthwhile *quantitative* refinement as well. This occurs when the levels of risk are high, the differences in risk among strata are high and/or the period of analysis is long. As we have shown, these conditions can be satisfied by some situations with a period of analysis of only a few years, such as dropping out of medical therapy, or death from a disease involving a high case fatality rate. If the period of analysis is several decades, the conditions are met by most models of adult mortality. Attention to heterogeneity in such circumstances can help us to construct more informative and accurate models.

The application of mixed-population models goes beyond the health-oriented issues discussed in this paper. Wherever there are populations with drop-outs, whether composed of college students, satisfactory inner-city housing units, reformed criminals or individuals receiving hernia repairs, attention to the heterogeneity of risk rates among individuals will yield dividends in prediction and understanding.

Acknowledgements—This work was supported in part by NSF grant SOC77-16602 to the Kennedy School, and by a grant from the Robert Wood Johnson Foundation through the Center for the Analysis of Health Practices. We would like to thank colleagues at the School of Public Health and the Kennedy School for helpful comments, particularly Dr. Benjamin Barnes, Dr. William Stason and John Pratt. Nathan Keyfitz provided continuing encouragement and stimulation.

REFERENCES

1. Blumen I, Kogan M, McCarthy PJ: **The Industrial Mobility of Labor as a Probability Process.** Cornell Studies of Industrial Relations, Vol. 6, Ithaca: 1955
2. Spilerman S: Extensions of the mover–stayer model. **Am J Sociol** 78: 599–627, 1972
3. Heckman JJ, Willis R: A beta logistic model for the analysis of sequential labor force participation by married women. **J Pol Econ** 85: 27–58, 1977
4. Singer B, Spilerman S: Some methodological issues in the analysis of longitudinal surveys. **Ann Econ Social Measmt** 5: 447–474, 1976
5. Tuma NB: Rewards, resources, and the rate of mobility: a non-stationary multivariate stochastic model. **Am Sociol Rev** 41: 338–360, 1976
6. Cox DR: Regression models and life-tables. **J R Stat Soc B** 34: 187–202, 1972
7. Kalbfleisch JD, Prentice RL: Marginal likelihoods based on Cox's regression and life model. **Biometrika** 60: 267–278, 1973
8. Zeckhauser RJ, Shepard DS: Where now for saving lives. **Law Contemp Prob** 40(4): 5–45, 1976
9. Shepard DS, Zeckhauser RJ: Heterogeneity among patients as a factor in surgical decision making. In: **Costs, Risks and Benefits of Surgery.** Bunker JP, Barnes BA, Mosteller F (Eds.). New York: Oxford University Press, 1977, pp. 56–69
10. Preston SH, Keyfitz N, Schoen R: **Causes of Death: Life Tables for National Populations.** New York: Seminar Press, 1972
11. Murray JL, Axtell LM: Impact of cancer: years of life lost due to cancer mortality. **J Natn Cancer Inst** 52: 3–7, 1974
12. Jenkins CD, Tuthill RW, Tannenbaum SI *et al.*: Zones of excess mortality in Massachusetts. **New Eng J Med** 296: 1354–1356, 1977
13. Schmeck HM Jr: US death rate in '74 lowest on record. **New York Times** February 4 1976, p. 1
14. Howell MA: The association between colorectal cancer and breast cancer. **J Chron Dis** 29: 243–261, 1976
15. Schottenfeld D: Concluding commentary for the international workshop on multiple primary cancers. **Cancer** 40 (Suppl.): 1982–1985, 1977
16. Schoenberg B, Greenberg RA, Eisenberg H: Occurrence of certain primary cancers in females. **J Natn Cancer Inst** 43: 15–32, 1969
17. Makeham WM: On an application of the theory of the composition of decremental forces. **J Inst Actuaries** 18: 317–322, 1874
18. Cornfield J: The estimation of the probability of developing a disease in the presence of competing risks. **Am J Pub Hlth** 47: 601–607, 1957
19. Berkson J, Elveback L: Competing exponential risks, with particular reference to the study of smoking and lung cancer. **J Am Stat Soc** 55: 415–428, 1960
20. Gail M: A review and critique of some models used in competing risk analysis. **Biometrics** 31: 209–222, 1975
21. David HA: **Parametric Approaches to the Theory of Competing Risks, Reliability and Biometry.** Proschan F, Serfling RJ (Eds.). Philadelphia: Society for Industrial and Applied Mathematics, 1974, pp. 275–290
22. Keyfitz N: What difference would it make if cancer were eradicated? an examination of the Taeuber paradox. **Demography** 14: 411–418, 1977
23. Chiang CL: **Introduction to Stochastic Processes in Biostatistics.** New York: John Wiley, 1968
24. Hakulinen T: On long-term relative survival rates. **J Chron Dis** 30: 431–443, 1977
25. Park WW, Lees JC: The absolute curability of cancer of the breast. **Surgery Gynec Obstet** 93: 129–152, 1951

26. Fox MS: On the diagnosis and treatment of breast cancer. **J Am Med Ass** 241: 489–494, 1979
27. Neuhauser D: Elective inguinal herniorrhaphy versus truss in the elderly. In: **Costs, Risks and Benefits of Surgery.** Bunker JP, Barnes BA, Mosteller F (Eds.). New York: Oxford University Press, 1977, pp. 223–239
28. Bartholomew DJ: **Stochastic Models for Social Processes.** New York: John Wiley, 1967, pp. 11–37.
29. Mann NR, Shaefer RE, Singpurwalla ND: **Methods for Statistical Analysis of Reliability and Life Data.** New York: John Wiley, 1974
30. Shepard DS: Prediction and incentives in health care policy. PhD Dissertation. Harvard University. Ann Arbor: Xerox University Microfilms, Dissertation No. 77-11744, 1977
31. Barnes BA: An overview of the treatment of end stage renal disease and a consideration of some of the consequences. In: **Costs, Risks and Benefits of Surgery.** Bunker J, Barnes B, Mosteller F (Eds.). New York: Oxford University Press, 1977
32. Cretin S: A model of the risk of death from myocardial infarction. Cambridge, Mass.: MIT Operations Research Center. Technical Report 0974, 1974
33. Truett D, Cornfield D, Kannel WB: A multivariate analysis of the risk of coronary heart disease in Framingham. **J Chron Dis** 20: 511–524, 1967
34. Knox EG: **Cervical Cancer, Screening in Medical Care.** New York: Oxford University Press, 1968, pp. 43–54
35. Perks W: On some experiments in the graduation of mortality statistics. **J Inst Actuaries** 83: 12–40, 1932
36. Shepard DS, Zeckhauser RJ: The assessment of programs to prolong life recognizing their interaction with risk factors. Boston: Center for the Analysis of Health Practices, Harvard School of Public Health, Discussion Paper, 1975
37. Keyfitz N: **Introduction to the Mathematics of Population.** Reading, Mass.: Addison-Wesley, 1968
38. Keyfitz N, Littman G: Mortality in a heterogeneous population. **Popul Stud** 33: 333–342, 1979
39. Haynes RB, Taylor DW: Annotated and indexed bibliography on compliance with health actions and therapeutic regimens. Prepared for Second McMaster Workshop/Symposium on Compliance with Therapeutic Regimens, 25–27 May 1977
40. Kannel WB, Gordon T (Eds.): **Some Characteristics Related to the Incidence of Cardiovascular Disease and Death: Framingham Study, 18-year Follow-up, the Framingham Study: An Epidemiological Investigation of Cardiovascular Disease.** DHEW Publication No. (NIH)74-599. Washington, D.C.: Government Printing Office, 1974
41. Pooling Project Research Group: Relationship of blood pressure serum cholesterol smoking habit, relative weight and ECG abnormalities to incidence of major coronary events: final report of the Pooling Project. **J Chron Dis** 31: 201–306, 1978
42. National Center for Health Statistics: **Life Tables for the Geographical Divisions of the United States 1959–1961.** DHEW Publication No. (PHS)1252. Washington, D.C.: Government Printing Office, 1965
43. Shepard DS, Zeckhauser RJ: Interventions in mixed populations: concepts and applications. Cambridge, MA: Kennedy School of Government, Harvard University, Discussion Paper No. 49, 1977
44. Weinstein MC, Stason WS: **Hypertension: A Policy Perspective.** Cambridge, MA: Harvard University Press, 1976
45. Gordon T, Kannel WB, McGee D: Death and coronary attacks in men after giving up cigarette smoking. A report from the Framingham Study. **Lancet** 2: 1345–1348, 1974
46. National Center for Health Statistics: **Vital Statistics of the United States 1973, Life Tables.** Vol. 2. DHEW Publication No. (HRA)75-1104. Washington, D.C.: Government Printing Office, 1975
47. Hambright TZ: **Comparability of Age on the Death Certificate and Matching Census Record.** Vital and Health Statistics Series 2, No. 29. Washington, D.C.: US National Center for Health Statistics, 1968
48. United Nations: **Demographic Yearbook 1974.** New York: Department of Economic and Social Affairs, United Nations, 1975
49. National Center for Health Statistics: **Vital Statistics of the United States 1970.** Vol. 2. **Mortality.** DHEW Publication No. (HRA)74-1101. Washington, D.C.: Government Printing Office, 1974
50. Nam CB, Weatherby NL, Ockay KA: **Causes of Death which Contribute to the Mortality Crossover Effect.** Tallahassee, FL: Institute for Social Research, Florida State University, 1978