NBER WORKING PAPER SERIES

THE CHOICE OF HEALTH POLICIES WITH HETEROGENEOUS POPULATIONS

Richard J. Zeckhauser

Donald S. Shepard

Working Paper No. 612

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

December 1980

This paper was prepared for the National Bureau of Economic Research's Conference on Economic Aspects of Health, Palo Alto, California, July 30-August 1, 1980. Research supported by grant number SOC 77-16602 from the National Science Foundation to the Kennedy School of Government and by the Health Services Research and Development Service of the Veterans Administration. Kenneth Arrow, Victor Fuchs, other participants in the NBER Conference on Economic Aspects of Health, Benjamin Barnes and Nancy Jackson gave us helpful comments. Michael Riddiough, Claudia Sanders and John Bell kindly shared their data from the study of the pneumococcal pneumonia vaccine by the Office of Technology Assessment. The research reported here is part of the NBER's research program in Health Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER CONFERENCE PAPER SERIES

Papers Available from the Conference

ECONOMIC ASPECTS OF HEALTH

Palo Alto, California

July 30-August 1, 1980

Conference Paper No.	
CP 84	"Prenatal Medical Care and Infant Mortality," by Jeffrey E. Harris
CP 85	"The Behavior of Mothers as Inputs to Child Health: The Determinatants of Birthweight, Gestation, and Rate of Fetal Growth," by Mark R. Rosenzweig and T. Paul Schultz
WP 539	"Time Preference and Health: An Exploratory Study," by Victor R. Fuchs
WP 611	"Healthiness, Education and Marital Status," by Paul Taubman and Sherwin Rosen
CP 86	"The Status of Health in Demand Estimation: Beyond Excellent, Good, Fair, and Poor," by Willard G. Manning, Jr., Joseph P. Newhouse and John E. Ware, Jr.
WP 609	"Physical Disabilities and Post-Secondary Educational Choices," by Robert A. Shakotko and Michael Grossman
CP 87	"Employment, Earnings, and Psychiatric Diagnosis," by Lee and Alexandra Benham
CP 88	"Children's Health Problems: Implications for Parental Labor Supply and Earnings," by David S. Salkever
WP 613	"The Choice of Health Policies with Heterogeneous Populations," by Donald S. Shepard and Richard J. Zeckhauser
CP 89	"Medical Care, Medical Insurance and Survival Probability -- The True Cost of Living," by Theodore C. Bergstrom

These papers have been submitted to the University of Chicago Press to be considered for publication in a volume tentatively titled, Economic Aspects of Health, edited by Victor R. Fuchs.

Copies of these papers may be obtained by sending $1.50 per copy to Conference Papers, NBER, 1050 Massachusetts Avenue, Cambridge, MA 02138. Please make checks payable to the National Bureau of Economic Research. Prepayment is required for orders under $10.00

List of Participants

ECONOMIC ASPECTS OF HEALTH

Palo Alto, California

July 30-August 1, 1980

Name	Affiliation
Kenneth Arrow	Stanford University
Lee Benham	Washington University
Theordore Bergstrom	University of Michigan
Victor Fuchs	Stanford University
Michael Grossman	City University of New York Graduate Center
Jeffrey Harris	Massachusetts Institute of Technology
Willard Manning	RAND
Joseph Newhouse	RAND
Sherwin Rosen	University of Chicago
Mark Rosenzweig	University of Minnesota
David Salkever	Johns Hopkins University
Paul Schultz	Yale University
Robert Shakotko	Columbia University
Donald Shepard	Harvard School of Public Health
Paul Taubman	University of Pennsylvania

NBER Working Paper #612
January 1981

The Choice of Health Policies with Heterogeneous Populations

ABSTRACT

Deciding whether to fund a given health program involves both statistical and ethical issues. Traditional statistical methods of measuring program effectiveness may give misleading results unless careful attention is paid to the question of population heterogeneity. Even within particular age and sex categories, members of a population typically differ in both their mortality rate and the extent to which they would benefit from a given medical intervention. It may or may not be possible to identify the risk factors (e.g., weight, smoking behavior) that explain these differences.

If an intervention confers unequal benefit on different risk groups, it will change their mixture within the population over time. If those helped most are those at greatest risk, a "traditional assessment" will overstate intervention benefits. Greater accuracy can be achieved through a "standardized assessment," which calculates intervention benefits separately for each distinctive risk group of the population. For example, a traditional assessment of pneumococcal pneumonia vaccine probably overstates program benefits and underestimates costs. Failure to recognize population heterogeneity also creates pitfalls in interpreting the results of clinical trials of new drugs, as illustrated by the example of sulfinpyrazone.

As more sophisticated statistical methods improve our understanding of differential program benefits, they will also raise ethical problems. Use of a standardized assessment, for instance, may make it clear that it is cost-effective to give an intervention to certain groups (e.g., nonsmokers, the elderly) but not others. Considering this problem from an "original position" may reveal an ethically acceptable basis for making such decisions on the basis of efficiency. We believe that if people were unaware of which risk group they themselves would fall into, they would elect to allocate resources according to the principle of cost-effectiveness.

Richard J. Zeckhauser, Ph.D.
John F. Kennedy School of Government
Harvard University
79 Boylston Street
Cambridge, Massachusetts 02138

(617) 495-1174

Donald S. Shepard, Ph.D.
John F. Kennedy School of Government
Harvard University
79 Boylston Street
Cambridge, Massachusetts 02138

(617) 732-1060

Table of Contents

THE CHOICE OF HEALTH POLICIES WITH HETEROGENEOUS POPULATIONS

I. INTRODUCTION

A. The Problem

Expending resources to secure health benefits is engaging in a game of chance. Thus, we might observe that, on average, a 65-year-old patient with an inguinal hernia and known medical characteristics who chooses an elective herniorrhaphy has about a 90 percent chance of completely successful surgery--surviving the operation without recurrence of the hernia--and a 10 percent chance of failure (0.3 percent chance of death, and 9.7 percent chance of recurrence) (Neuhauser 1977). In decision theory, we are used to considering such problems as if they were analogous to playing a roulette wheel, the outcome determined by the fall of a ball.

Many medical interventions, we shall argue here, do not fit this paradigm. It is not an entirely random process that determines which individuals have successful operations. They may tend to be different,in ways that may or may not be observable,from the 10 percent that have poor outcomes. That is, the population is heterogeneous.

Heterogeneous or mixed populations are ones in which the probability of loss from the population varies among members, in its initial value, in its evolution over time, or in its response to an intervention. In a health context, the population is people who have not yet suffered some sickness, complication, or death; the loss is the

onset of that condition. Variations along such dimensions as sex and age are well known; hence we traditionally standardize mortality statistics with respect to such variables. Medical characteristics, such as smoking or more general medical history, may also be employed as classifying variables. Here our concern is not only such directly recorded variables, but others that may go unnoticed or that are even inherently unobservable.

The benefits of medical procedures often vary according to the patient's characteristics. With herniorrhaphy, for example, an occupation requiring heavy lifting and straining, obesity, use of steroid drugs which discourage healing (perhaps to treat asthma), or chronic lung disease (which may cause coughing) would all reduce the chances of successful surgery. Where explanatory variables can be readily observed and classified, we are likely to standardize on them. Thus, for example, we might express surgical mortality as corrected by such variables as age, type of operation (elective or emergency), and the patient's general health. Frequently, however, the explanatory variables are either unmonitorable, unrecognized, or sufficiently difficult to classify that they go unmentioned. Although surgeons attempt to ascertain the risk of heart attack during or after surgery through medical history and routine tests, they typically forego the greater predictive accuracy that could be achieved through invasive cardiac tests, which have their own dangers.

Individuals differ, then, in their chances of contracting an illness, failing to respond to treatment, or dying. This paper is concerned with the policy implications of that heterogeneity in the

population, with the ultimate purpose of helping officials who must decide where to direct health interventions. Understanding the effects of population heterogeneity will permit more accurate assessment of the benefits and costs of various possible interventions, thus contributing to better policy.

The salient characteristics of heterogeneity differ depending on whether the characteristics that predict an individual's probability of loss are observed and are used as a basis for prediction. If so, we say that the population embodies "observed heterogeneity." Accurate assessments then require that we trace the way each of the several observed risk groups evolves over time. That is, we predict the expected health benefits each risk group receives from alternative interventions, and the expected resources the interventions will require each year. Policy choices involve not only these epidemiological and modeling considerations, but ethical and economic issues entailed in setting priorities for offering health programs to different categories of individuals.

By "latent heterogeneity" we denote the situation in which different subgroups of the population are at different levels of risk, but the differentiating factors are not used as a basis for prediction. They may be unobservable, unobserved, or observed and ignored because they are not known to be relevant. With latent heterogeneity, the economic and ethical issues related to setting priorities within the population are obviated; distinctions simply are not made among individuals. However, important and generally unrecognized statistical questions arise that must be attended to if accurate assessments of benefits and costs are to be made.

B. Observed Heterogeneity

In cases of observed heterogeneity, we assume that the benefits and costs returning to members of the different risk groups can be determined by established methods. The policy task is to establish priorities for the risk groups in receiving interventions. This is essentially a problem in public expenditure theory, where the objective is to maximize the difference between benefits and costs, or in the constrained resource case, to maximize benefits subject to some resource limit. Complications arise in three areas: (1) a metric must be established for health benefits, (2) in addition to their immediate costs, medical interventions have implications for future costs which may be paid from other pockets, (3) determining the order in which different groups shall be offered medical interventions on the basis of predicted benefits and costs has significant ethical implications. These issues are explored below in Section IV.

C. Latent Heterogeneity

Even though a human population appears homogeneous, there will be differentials in risk among its members; i.e., there will be latent heterogeneity. Most interventions directed towards heterogeneous populations will offer differential benefits to the members of that population at various risk levels. Hence, they will change the mix among the surviving population. Even when the nature of the different risk groups is unclear, recognizing the fact of latent heterogeneity will make it possible to predict the effects of interventions more accurately.

Evidence. The existence of heterogeneity in a population's risk levels can be diagnosed by observing losses from the population over time and comparing that pattern with a theoretical norm that assumes

homogeneous risk. As the simplest case, suppose we could confidently assume that loss rates for each individual were constant over time; heterogeneity would then reveal itself by a decline in loss rates over time as the high-risk members were eliminated from the population. This sort of population heterogeneity could explain why students' failure rates decline over successive years in colleges, or why rates of recidivism decline after longer periods of abstention from alcohol or smoking. (See Shepard 1977.)

With smoking, for example, it is frequently alleged that the first few weeks of forbearance are the most difficult. Surely there is another phenomenon at work. Some individuals, for whatever reasons, are more likely to return to smoking in any given period. Persons with a high probability of recidivism tend to be eliminated from the "not smoking" population early. That effect tends to reduce recidivism in the population over time.

In earlier analyses (Shepard and Zeckhauser 1977, 1980a) we have looked at the relapse problem in relation to hernia recurrence. The annual rate of recurrence drops from 30 per thousand person-years during the first year after the initial hernia repair, to 9.1/1,000 during years one through five, to 3.4/1,000 for years five to ten. Heterogeneity with respect to per-period probability of recurrence offers a straightforward explanation for this observed pattern: each individual remains at the same level of risk indefinitely, but a larger proportion of the recurrence-prone individuals are removed early. The alternative explanation is that the recurrence rate for each recipient of the operation falls over time. Physicians whom we consulted told us, however, that it was highly unlikely that the observed

patterns reflected a strengthening of tissues over time. Wounds gain maximum strength in a few weeks, not a few years. A further piece of evidence supports the hypothesis that there is patient-to-patient heterogeneity in recurrence rates. The probability that a patient undergoing herniorrhaphy ever experiences a first recurrence is about 10 percent; the probability that he suffers a second recurrence, given that he had a first recurrence, is 35 percent (Neuhauser 1977). The higher rate for second recurrences is consistent with the fact that persons with a first recurrence are primarily those with above-average recurrence probabilities.

Even if individuals' loss rates do not remain constant over time, we will still be able to detect heterogeneity, as long as loss rates conform to some predictable patterns. Consider, for instance, a group of people who have stopped smoking. We assume that there is substantial heterogeneity in the initial relapse rates back into smoking, and that each individual's loss rate increases exponentially over time. It may nevertheless turn out that, for the population as a whole, observed relapse rates decline over time. The selection factor that applies to the population as a whole overcomes the increasing risk factor that applies to each individual separately. Consider now an individual whose initial relapse rate is unknown. His risk of relapsing appears to decline over time, as he remains in the population, whereas actually it is increasing. The source of the paradox is that the individual is simultaneously undergoing two effects. By surviving he reveals that he is a better risk than the initial population average. Whatever his true risk level, however, that risk is increasing on a period-by-period basis.[1]

Possible Biases in Assessment and Interpretation. Latent heterogeneity must be considered when choosing interventions or when interpreting data derived from experiments, e.g., the differential survival of those given a new drug in a randomized controlled trial.

Without proper attention to latent heterogeneity, we will misestimate the long-term benefits and costs of medical interventions. We want to state at the outset that our problem is not the familiar bias due to confounding of treatment differences with patient differences. We are assuming that outcomes with and without an intervention are compared for populations that are initially identical. The bias that we discuss arises because an intervention selectively eliminates certain members of the mixed population from further followup, so that after a time, the makeup of the population differs from what it would have been without the intervention.

Latent heterogeneity usually biases inferences about the effects of an intervention. One type of inference, termed a population projection, estimates the effect of a treatment on a population from its effect on homogeneous risk groups. Projections that ignore heterogeneity will generally make the intervention appear better than it really is. A second type of inference estimates the effect of the intervention on each individual, or homogeneous risk group, from its observed effect on the population. Generally, the effect on each individual will be stronger than the observed effect on the population. Heterogeneity generally attenuates the effect on the population compared to the effect on the individual.

The next section discusses ways of improving our predictive ability by recognizing heterogeneity.

II. PREDICTING THE BENEFITS OF AN INTERVENTION ON A POPULATION WITH LATENT HETEROGENEITY

Analytic models enable us to predict the gains when interventions reduce loss rates in heterogeneous populations. The recommended procedure, which we shall refer to as a "standardized assessment," begins by classifying the population into homogeneous strata. Within each stratum, the time-specific (which implies age-specific when age is of consequence) loss rates are computed with and without the intervention. For each stratum an output measure of interest, such as mortality, life expectancy, or duration of freedom from a disease, is computed. The overall outcome measure for the population is computed by averaging the measures for each stratum weighted according to initial prevalences.

This paper compares the standardized assessment procedure with the most widely used methodology for evaluating interventions, which we shall refer to as a "traditional assessment." A traditional assessment starts with the pattern of losses under baseline conditions varying with both age and time, derived from some observed data or a statistical model. (If no adjustment is made for age and time, a procedure we term a "naive assessment," then greater biases may result [Shepard and Zeckhauser 1980a].) The active intervention is assumed to change age- and time-specific loss rates. The traditional assessment, in contrast to the recommended procedure, assumes (usually implicitly) that the presence of the intervention does not alter the population mix at any point in time remains the same. Output measures are computed using mortality rates that are adjusted to allow for the direct effect of the intervention, but take no account of the effect of the intervention in changing the mix in the population.

Recent trends in U.S. mortality (discussed in Section III) and a number of other illustrations presented below suggest that interventions change the mixture of risk groups in the population, most often increasing the proportion of high-risk persons. For example, a simulation has shown that the continued availability of a mobile coronary care unit would increase the percentage of surviving males over age 30 who had had heart attacks from 12 percent to 15 percent (Zeckhauser and Shepard 1976). This group is at substantially higher risk for future coronary events.

When carrying out the recommended standardized assessment, a critical problem is to stratify the population appropriately. Stratification variables might include lifestyle characteristics, such as smoking and drinking habits, single medical variables, or multivariate risk scores. In some instances, socioeconomic factors may be employed as proxies for underlying causal factors. Many factors that influence mortality risk may be unobserved; some may never be observed, though their presence can be inferred through experiment.

The important point is that stratification should proceed far enough that the individuals within a risk category are relatively homogeneous. The reward for painful efforts at classification is an assessment procedure that avoids the systematic bias inherent in traditional assessments.

A. Formal Concepts for Assessing Interventions

Consider a mixed population whose risk categories are indexed by j, with the initial prevalence of each being r_j. Interventions are indexed over i; i = 1 represents the baseline intervention. If we let $\mu_{ij}(x)$ denote the instantaneous hazard rate, survival at age x will be

(1) $\ell_{ij}(x) = \exp[-\int_0^x \mu_{ij}(t)dt]$,

where $\ell_{ij} = 1$ at $x = 0$. The mixed population's survival at age x, denoted by $\ell_{i.}(x)$, is a weighted average of $\ell_{ij}(x)$,

$$\ell_{i.}(x) = \sum_j r_j \ell_{ij}(x).$$

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

$$(2) \quad r_{ij}(x) = r_j \frac{\ell_{ij}(x)}{\ell_{i.}(x)} .$$

The overall hazard rate under intervention i at age x, $m_i(x)$, is a weighted average by prevalence of the hazard rates for the individual risk groups:

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

B. Biases Inherent in Traditional Assessments

In a traditional assessment, the prevalence of risk category j at age x is taken to be its baseline value, $r_{ij}(x)$. Bias arises when the population is subjected to an active treatment, i = 2 for illustration. Risk is estimated to be

$$m_2'(x) = \sum_j r_{1j}(x)\mu_{2j}(x) ,$$

where comparison with (3) reveals that the unbiased estimate replaces r_{1j} with r_{2j}.

Our interest is in the difference between the (incorrectly weighted) traditional assessment and the standardized assessment. This difference, $\Delta m(x)$ at age x, is defined by

$$\Delta m = m_2(x) - m_2'(x)$$

where a positive difference indicates that the traditional assessment overstates the benefit of the intervention in reducing mortality.

In an earlier analysis (Shepard and Zeckhauser 1980a) we identified this bias. Here we shall prove some of its properties.[2] Let us define comparative survival gain as

$$g_j(x) = \frac{\ell_{2j}}{\ell_2} - \frac{\ell_{1j}}{\ell_1}. \tag{4}$$

Thus, g_j measures the gain in survival to group j relative to overall survival gain in treatment. The group that gains the most has the highest value for g_j. We prove below that the bias of the traditional assessment depends on which risk group gains relatively most in survival.

Theorem 1

Let $m_2'(x)$ be mortality calculated by the traditional assessment and $m_2(x)$ be mortality standardized for appropriate risk indicators. Let μ_{ij} be mortality rates and g_j be the relative survival gain as defined above. Then $\Delta m(x) \equiv m_2(x) - m_2'(x) = \text{covariance}(\mu_{2j}, g_j)$.

Proof

First we establish the chain of equations in (5) by substituting from the definitions of $r_{ij}(x)$ in (2) and of g_j in (4):

$$r_{2j} - r_{1j} = r_j(0)\left[\frac{\ell_{2j}}{\ell_{2.}} - \frac{\ell_{1j}}{\ell_{1.}}\right] = r_j(0)g_j(x) . \tag{5}$$

Summing both sides over j, we have

$$\sum_j r_{2j} - \sum_j r_{1j} = \sum_j r_j(0)g_j(x) .$$

The left-hand side is 1 - 1 = 0 and the right-hand side is the mean of the g_j, weighted by $r_j(0)$. Thus the weighted mean of g_j is zero. Now multiply both sides of (5) by μ_{2j} and sum over j:

$$(6) \qquad \sum_j \mu_{2j}(r_{2j}-r_{1j}) = \sum_j r_j(0)\mu_{2j}(x)g_j(x) \ .$$

Recalling the definitions of $m_2(x)$ and $m_2'(x)$, the left side in (6) is $\Delta m(x)$. The right side is the weighted cross product of $\mu_{2j}(x)$ and $g_j(x)$.

Recalling that the weighted mean of $g_j(x)$ is zero, we can subtract the product of the means of $g_j(x)$ and $\mu_{2j}(x)$ (termed $g_.$ and $\mu_{2.}$, respectively) from the right-hand side, yielding

$$\Delta m(x) = \sum_j r_j(0)\mu_{2j}(x)g_j(x) - \mu_{2.}(x)g_.(x) = \text{covariance}(\mu_{2j}(x),g_j(x)) \ .$$

Q.E.D.

This theorem represents our central result with respect to bias. For example, if the absolute benefit of the intervention is greater for high-risk individuals, then the greater values of g_j will be associated with the larger values of $\mu(x)$. This implies that the covariance is positive. The bias will be positive as well. From this it follows:

> <u>Traditional assessment methods will overstate the benefits of interventions that offer a greater reduction in force of mortality for high-risk than for low-risk individuals.</u>

The expression for $\Delta m(x)$ reveals that the magnitude of the bias is time-dependent. At time zero, there is no bias, for selective mortality has not yet affected the composition of

the population. After a long period of time virtually all survivors with or without the intervention will be low-risk individuals; the bias is small. In the intermediate run the bias is greatest. Traditional assessments implicitly underestimate the number of high-risk individuals for intermediate times. In a numerical example involving a hypothetical intervention to reduce the risk of dropping out of a medical treatment, such as anti-hypertensive therapy, we found that the maximum bias occurred at 3.5 years after the start of the intervention. Loss rates were underestimated by 31 percent in a traditional assessment (Shepard and Zeckhauser 1980a).

A common complaint in recent years has been that our dramatically increased national expenditures on health care and other health-promoting activities have not done much to lengthen life expectancy. (See, for example, Fuchs 1974.) But perhaps they have, at least for an individual at any particular risk level. The problem may be that heterogeneity masks such effects by producing a weaker overall population of survivors.

How then should an individual feel about health care? Suppose that health improvements had reduced age-specific mortality for all initial risk levels regardless of age by X percent from 1930 to 1970. If an individual knows his prior risk level, he can calculate his current risk level as X percent less. If he does not know his prior risk level, then he must compare the age-specific mortality rates for men of his age today and 40 years ago. This comparison will generally show an expected decline in age-specific mortality that is less than X percent. The reason

is that an individual alive now at an intermediate age, say age 60, has a higher probability of being a high-risk individual than an unselected individual of the same age 40 years ago.

To illustrate numerically, assume as above, that there were two equal-sized risk groups, with forces of mortality of 5 percent (indexed by $j = 1$) and 10 percent ($j = 2$) per year. A vaccine would cut both of these rates in half (X equals 50 percent). If a person survived for 10 years without the vaccine, the chance that he was a high-risk ($j = 2$) person is $r_{12}(10)$, or 38 percent. If he survived for 10 years with the vaccine, his chance of being a high-risk person is higher, $r_{22}(10)$, or 44 percent. Though each individual has improved his chances of survival, the population as a whole contains a larger proportion of high-risk people than before the vaccine.

C. Bias for Common Epidemiological Models

Most of the examples we have encountered in our study of heterogeneity suggest that traditional assessments will overstate the benefits of interventions. However, this is an empirical rather than a logical proposition. If the covariance between benefit and risk were negative, the bias would be in the opposite direction. To discover an example of negative covariance--what we believe to be the unusual case--simply look for an intervention that offers its greatest benefits to healthier individuals. Programs that benefit employed persons might meet this criterion, since those who are employed are likely to be healthier than the population overall. Presumably workplace health and safety programs fall in this category. Even with these programs, however, if we restricted our attention to a single category of the explanatory variable (employment), the covariance would probably turn positive again.

For example, among employed individuals, those in the least healthy jobs might gain most from an occupational health program.

Two of the most widely used models of mortality both illustrate a positive bias. They are the multiplicative and logistic models. The multiplicative model assumes that the effects of an intervention are independent of other determinants of risk. The intervention merely applies a constant multiple, a, to whatever level of instantaneous risk previously held. The term for this multiple in epidemiology is the risk ratio. Thus

$$\mu_{2j}(x) = a\mu_{1j}(x) \qquad \text{where } o < a < 1.$$

Frequently the multiplicative model is written with $\mu_{ij}(x)$ as an increasing function of age (such as the Gompertz function), which enables it to reproduce realistic mortality experience from around age 30 onwards.

The logistic model applies a constant factor to the odds of death in a discrete time interval. It has been widely applied in cardiovascular epidemiology. The logistic model requires that for an interval of time Δx starting at time x,

$$q_{2j} = 1/[1 + \exp(b_j+\alpha)],$$

where b_j is defined so that

$$q_{1j} = 1/[1 + \exp(b_j)].$$

and $\alpha < 0$.

For each of these two widely applied models of risk, the covariance defined in Theorem 1 is positive. Traditional assessments of interventions will overstate their benefits. See Appendix A for theorems and proofs.

D. Conditions When Latent Heterogeneity is Important

Explicit attention to latent heterogeneity has been shown to eliminate a source of bias in predicting the effects of population interventions, but it also increases the complexity of an analysis. In circumstances where the gain in precision is small, the additional complexity may not be worthwhile. As the source of bias in traditional assessments is the changing mixture of risk groups due to an intervention, the magnitude of bias depends on the extent of selection. The selection effect will be significant if the baseline mortality rate or loss rate is high, the difference in loss rates between risk groups is large, and the intervention has a powerful effect on losses. A high loss rate means that a substantial portion (roughly at least 20 percent) experiences the event under consideration within the time period at issue. A large difference in loss rates means that the high-risk group has at least twice the risk of the low-risk group; and a powerful intervention is one that cuts the risk of losses by, say, 30 percent or more.

More formally, manipulating Shepard's (1977, p. 154) finding shows that the proportional bias in mortality, $\Delta m(x)/[m_2(x) - m_1(x)]$, is approximated to first order by the product of the following factors: (1) the number of years over which the intervention is considered; (2) the average loss rate over that period in the baseline intervention; (3) the square of the coefficient of variation in initial loss rates among risk groups in the baseline intervention.

In an example developed elsewhere (Shepard and Zeckhauser 1980a), cigarette smoking was an important source of latent heterogeneity in the analysis of the benefits of controlling blood pressure in a hypertensive

male from age 50 onward. Suppose we wished to know the intervention's effect at age 75, that is, after 25 years. In the baseline intervention, the cohort with a diastolic blood pressure of 110 mm. Hg. faces a mortality pattern that rises exponentially (at rate of 0.08 per year) from an initial level at age 50 of 20 per 1000 per year. Thus the average mortality over the period is .064 per year. Since smokers consuming a pack or more a day have about twice the risk of"others"(non-smokers and light smokers), the coefficient of variation is 0.33. The approximation gives a proportional error of 17 percent (25 x 0.064 x 0.33^2, expressed as a percent), close to the directly calculated bias of 16 percent (Shepard and Zeckhauser 1980a, p. 428).

When death is the loss being modeled, then latent heterogeneity is important only where mortality risks are substantial. This occurs when advanced age is combined with a chronic condition of moderate importance, as in the example of moderate hypertension above, or where a medical condition imposes extremely high risk, as with survivors of a recent heart attack in the discussion of sulfinpyrazone below. For other kinds of losses, the rates are much higher. Examples include dropping out of a treatment program, relapsing in a behavior such as abstinence from smoking, or the recurrence of a medical problem such as a hernia. Here the bias within a few years can be substantial. (See Shepard and Zeckhauser 1980a, p. 423.)

E. Latent Heterogeneity and the Bias in Future Health Cost Estimates

Since the difference between the traditional and standardized assessment lies only in the evolution of risk groups over time, the two procedures will give identical estimates of short-run costs. Heterogeneity, however, is likely to be of significance in assessing the

long-run costs of survivors. Let us assume, as we have previously, that the high-risk individuals receive differentially greater benefits. Usually we would also expect them to have the highest expected annual costs should they survive. However, this factor alone is not sufficient to prove the direction of bias in lifetime costs. High-risk individuals are likely to live a shorter time. Hence their costs (even discounted costs) might actually be less.

Our policy concern is not with expected lifetime costs themselves, but rather with costs per unit of benefit. We should normally expect high-risk individuals to be at a disadvantage in costs per unit of benefit. If they are, and if the intervention benefits the high-risk group more, then the use of a traditional assessment will be too favorable to an intervention. We shall return to this subject in Section IV under our discussion of cost-effectiveness.

III. DRAWING INFERENCES ABOUT AN INTERVENTION FROM POPULATION DATA

In the preceding section, we assumed that the effect of an intervention on a homogeneous risk group was known, and established procedures for predicting its effect on a mixed population. In the two examples of this section, we work in the opposite direction. We assume that we have observed the longitudinal effects of an intervention on a population. We want to infer the effects on individuals (or homogeneous risk groups) to better understand the structural effect of the intervention. We also want to be able to extrapolate from observed data to effects for different forms of the intervention, or to time periods beyond those for which we have direct data.

Our first example traces age-specific mortality reductions over the past 40 years. The intervention is a general improvement in health conditions (medicine, environment, and standard of living). We believe the observed patterns are suggestive of a heterogeneous population, with higher risk groups receiving greater benefits from mortality reduction.

Sulfinpyrazone (brand name Anturane), a newly proposed and highly controversial drug used to prevent deaths in the months after a heart attack, is examined second. The original study reported that while the drug appeared to be effective initially, it apparently offered no protection beyond six months (Anturane 1980). That inference is not valid, as we shall demonstrate, if those who would have died in the absence of the drug are at higher risk after six months than those who would have survived without it.

A method for drawing appropriate inferences in such situations is of general importance. Frequently, when mortality curves of populations

under different treatments coincide after a period of time, it is thought that the treatment no longer has any effect. Our models provide a different interpretation. Two opposite effects, one of selection, the other of differential risk, may be canceling each other.

A. Inferring the Presence of Heterogeneity: Changes Over Time in U.S. Mortality

Figure 1 shows the decline in age-specific death rates for white males in the U.S. from 1930 to 1970. (The age range of 25 through 84 was selected because simple exponential [Gompertz] mortality models can be applied over that range.) It is striking that the reduction in mortality rates around age 70 is only a fourth as large as around age 30. The higher the age group (except for age 80), the smaller is the reduction in mortality. Percentage gains in life expectancy follow a similar pattern. Life expectancy at birth has increased from 59.1 to 68.2 years over this period, a gain of 9.1 years or 15 percent. At age 40, however, the gain in remaining life expectancy is only 2.8 years or ten percent. For other sex and color groups, a similar pattern could be reported.

What factors might be responsible for this pattern of gains? Improved nutrition, sanitation, and medical care have undoubtedly figured prominently. Improvements in medical technology in this interval, particularly the introduction of sulfa and antibiotic drugs, may have been of greatest benefit to children and young adults. Yet these drugs have also been important in the treatment of diseases that affect the aged, like bacterial pneumonia. For example, the annual death rate between ages 70 and 75 from pneumonia, influenza, and respiratory disease has fallen by 70 percent from 1930 to 1960, from 4.6 to 1.4 per 10,000.

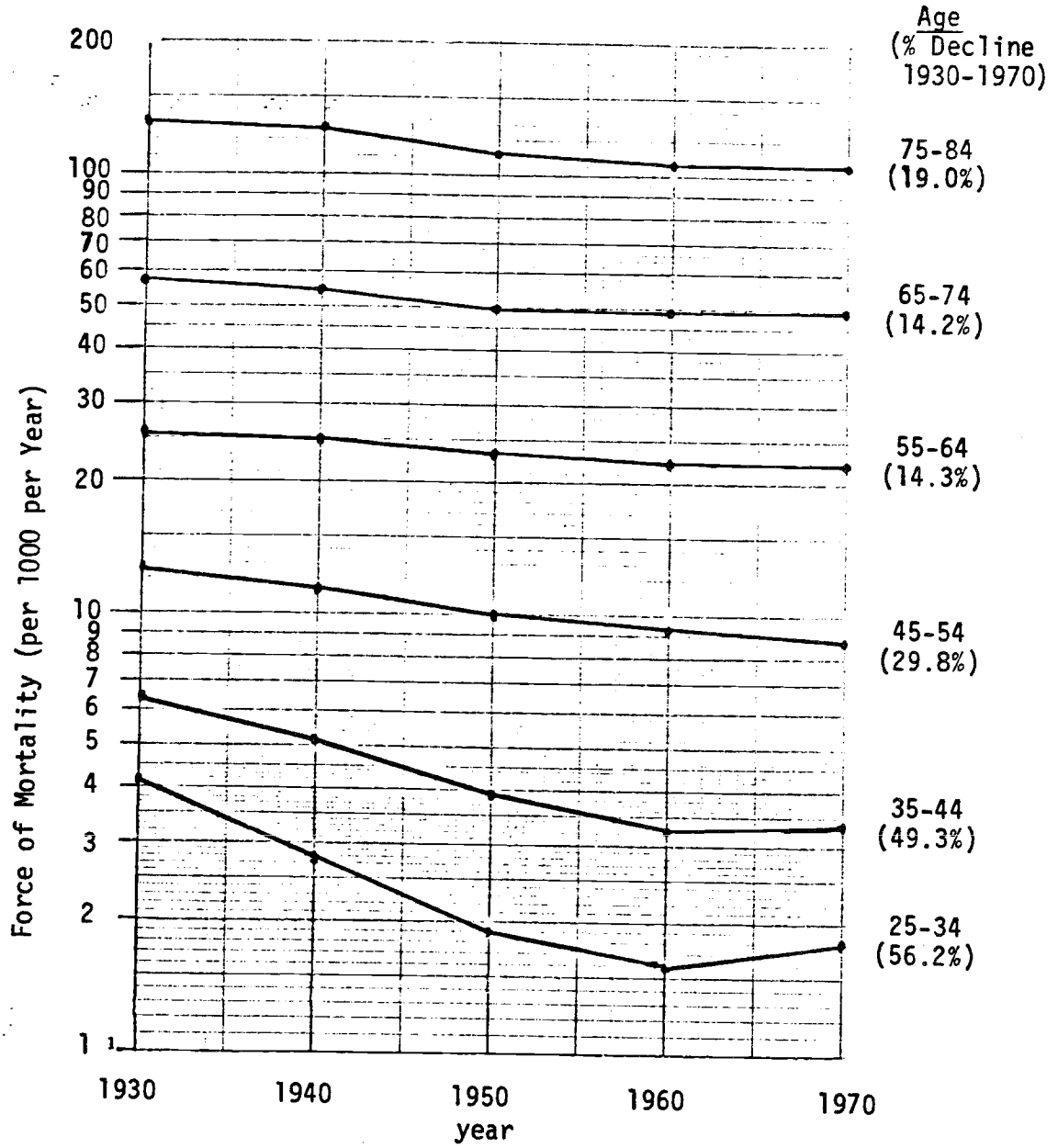

Fig. 1. Age-Specific Mortality Rates for White Males in the U.S., 1930 to 1970 (With Percentage Declines in Parentheses)

Source: See Table A-1 in Appendix B.

We offer an explanation for these data that emerges from our earlier discussion: The population consists of persons at different risk levels, and improvements in medicine, nutrition, and sanitation have interacted with the risk factors that were operating. Suppose that the population consists of persons who fall into one of two classes, "constitutionally weak," or "normal." The weak are especially prone to disease and remain at high risk until they succumb. They have only a small chance of surviving to middle age. The medical and environmental changes since 1930 benefited the normal to some extent, but were of greater benefit to the weak. At earlier ages, we see a substantial decline in death rates because weak persons are being saved. At more advanced ages, we now have a greater proportion of weak persons among the survivors. This structural change partially offsets (though it does not overcome) the fact that the prognosis for each weak and each normal person at every age level has improved; the improvement results in a modest reduction in mortality rates at advanced ages.

Since this hypothetical constitutional weakness may be an unobservable risk factor, our explanation is unlikely to be proved conclusively. Nevertheless, a simple simulation shows that our explanation yields results generally consistent with the data. Our model, described in Appendix B, fits the data markedly better than the competitive best polynomial model with the same number of parameters. Although we do not wish to claim that mortality improvements favoring high-risk persons were the sole factor responsible for the pattern of decline in age-specific mortality rates, it is encouraging that our model generates results consistent with the observed pattern.

We have reported previously (Shepard and Zeckhauser 1980a) that the hypothesis of "constitutional weakness" can also explain the "crossover" effect in remaining life expectancy. A national or racial group whose life expectancy at birth is lower (e.g., non-whites compared to whites) may experience more powerful selection; this phenomenon can explain why the remaining life expectancy in the U.S. for non-whites exceeds that for whites of the same sex beyond about age 70, while whites have a greater life expectancy at birth.

B. Deducing the Benefits of an Intervention--The Sulfinpyrazone Example

The best way to get a long-term estimate of the benefits of an intervention is to employ a randomized controlled trial. If the size of the population is large, the experiment is well controlled, the experimental group is perfectly representative of the population to which the actual program will be offered, and the intervention is to be used precisely as tested in the trial, no additional modeling is required. We need merely examine the magnitude of the benefit conferred in the randomized trial.

In many situations, however, we wish to extrapolate from results of a randomized trial to other situations, so we must resort to modeling. A recent article in the New England Journal of Medicine reported the performance of sulfinpyrazone (Anturane), a drug designed to offer protection against cardiac deaths to post-heart attack patients (Anturane 1980). A controlled trial was undertaken over a 24-month period. The data in the original article show that sulfinpyrazone offers considerable protection over months one through six. The annualized mortality rate was 5.0 percent with the drug as opposed to 10.3 percent with the placebo. During months 7 through 24, however, there was

virtually no difference between the two groups in annualized mortality (4.1 percent with the drug versus 3.7 percent with the placebo).

The article concluded that sulfinpyrazone offered protection for six months, but not beyond. The implied clinical recommendation was that physicians should prescribe the drug for a six-month period but no more.[3] Unfortunately, there has been no controlled trial that compares six months of use of the drug with continued use. All experience to date under a trial is with continued use or no use (placebo). The intervention that appears to be recommended by the data is only a minor modification (as to duration) from the protocol actually followed in the trial. Yet, as we shall see, any assessment of what would happen to a population that stopped taking the drug after six months is at best a careful speculation. The policy implications of this trial depend critically on inferring the benefits to an individual participant.

We offer a contrary interpretation of the small differential after six months, based on the heterogeneous population concept. (See Shepard and Zeckhauser 1980b for a brief informal argument.) Participants within each treatment group vary in their probability of cardiac death. Suppose, as seems plausible, given experience with other interventions, that sulfinpyrazone was most helpful to those whose mortality rate would have been highest. Then, the drug-treated group surviving at seven months would have a larger proportion of high-risk patients than would the group receiving the placebo. If that is true, then the equal experience of treatment and control groups beyond six months would illustrate a continuing positive benefit from drug treatment.

We have modeled this situation using techniques equivalent to those employed for the U.S. mortality example as described above and in

Appendix B. With two risk groups we achieved an excellent fit, fitting annualized mortality rates to less than a tenth of a percentage point for both placebo and sulfinpyrazone groups for months one through six and seven through 24. (See Table 1.) The high-risk group, which comprises 4.4 percent of the population at the outset, has a risk of 0.814 percent per day , which is reduced by 75 percent with sulfinpyrazone. The low-risk group has a mortality of 3.41 percent per year, which is reduced 32 percent by the drug.[4]

(Table 1 about here)

The important point is that the model assumes that the benefit of sulfinpyrazone treatment persists through the entire 24 months. Its continuing effectiveness is masked in the aggregate mortality rates because it is offset by the increased proportion of high-risk survivors in the treatment group. As shown in Table 2, which is derived from our model, the proportion of high-risk subjects in the sulfinpyrazone group surviving at six months is three times as great as in the placebo group.

(Table 2 about here)

If the world corresponded exactly to our model, and we have too little data to make any claims along this line, discontinuing treatment after six months would lead to an increase in the annualized cardiac mortality rate of the sulfinpyrazone group from 3.7 percent to 5.5 percent. In sum, the evidence is consistent with the hypothesis that selection within a heterogeneous population prevented sulfinpyrazone from having superior mortality experience after six months, despite the fact that it was offering continuing benefit.

Table 1. Annualized Percentage Mortality Rate

Months in Study	Placebo		Sulfinpyrazone	
	Actual	Model	Actual	Model
Up to 6	10.3	10.3	5.0	5.0
7 through 24	4.1	4.1	3.7	3.7

Table 2. Percentage of Survivors in High-Risk Group

Months After Entry	Placebo	Sulfinpyrazone
0	4.40	4.40
6	1.05	3.11
24	0.01	1.08

A controlled trial randomizing for duration of therapy beyond six months could test the hypothesis that the selection effect over months one through six had masked the beneficial impact of the intervention in months seven through 24. Leaving aside sampling questions, suppose the controlled trial found that patients on continued therapy did better than those discontinuing at six months. Such a finding would provide further information from which to infer the structure of different risk groups, but still might not provide sufficient information to identify uniquely a complex model. The finding would, however, answer the policy-relevant question concerning continued administration of the drug.

IV. POLICY CHOICE WITH OBSERVED HETEROGENEITY

With latent heterogeneity, the intellectual challenge is to infer the composition of a population and predict the way it will respond to an intervention. With observed heterogeneity, the problem is one of resource allocation: which individuals should receive which interventions? Discrimination can be made on the basis of standard demographic variables such as age and sex, medical characteristics such as blood pressure, individuals' preferences, or any other variable that is observable and that helps predict how an individual will respond to an intervention. (Since it will seldom be possible to classify the population into truly homogeneous risk groups, the methods outlined in our discussion of latent heterogeneity will have to be employed to predict how the identifiable subpopulations will respond to an intervention.)

To set priorities for interventions, it is not sufficient merely to estimate benefits. Costs must be considered as well. With costs, as with benefits, heterogeneity may play a significant role. Simple extrapolations that do not allow for heterogeneity will provide biased assessments of costs. Recent work on the high-cost users of medical care suggests that expenditures are highly skewed; a small percentage of individuals accounts for a very high percentage of costs (Zook and Moore 1980). For example, the most costly 12 percent of hospital patients in a year account for more expenditures than the remaining 88 percent. This suggests that heterogeneity is not only important conceptually when addressing the cost question, but is likely to be of policy consequence. (See Zook, Moore and Zeckhauser 1980.)

The major question we shall address first is: How should we set priorities for an intervention when we can estimate the benefits and costs it will generate for different members of the population?

A. Efficient Resource Allocation: The Cost-Effectiveness Paradigm

For simplicity, we shall consider the case in which there is only one intervention that is an alternative to the status quo. We assume that the objective is to maximize total health benefits obtained by all persons treated within a fixed total cost (in present value terms) to the medical care system. The policy must therefore establish priorities among different classes of individuals for receiving the intervention. The policy may also have to determine how much money to spend, which in effect is the question of how far to proceed down the priority list.

In the usual case, health interventions will offer positive net health effects but incur positive net costs. The policymaker then confronts a classic public expenditure problem. To maximize total benefits subject to a budget, he should follow the cost-effectiveness criterion. Cost-effectiveness (CE) is defined as the ratio (Shepard and Thompson 1979):

$$CE = \frac{\text{Net costs paid from the constrained budget}}{\text{Net health benefits}} .$$

Net health effects may be expressed in any metric common to all the alternative programs being compared. Two metrics which have proved particularly useful are years of life (improvements in life expectancy) and quality-adjusted life years, QALYs. QALYs generalize the concept of life expectancy to adjust for quality of life, applying different weights according to functional status. (QALY totals may also be adjusted for timing, discounting future years of life. It should be understood that

the discounting process is controversial, particularly with non-economists.)

Net costs of a health intervention are the sum of three costs: (1) the costs of the intervention itself, including treatment of any resulting side effects; (2) the medical costs or savings for the condition at which the intervention is aimed, called related treatment costs; and (3) general induced medical costs for other conditions, as a result of longer life.[5] In the example of the pneumococcal pneumonia vaccine, which we develop in Section V, the intervention cost is that of giving a vaccination. The related treatment savings are the avoided costs of hospitalizations for cases of pneumonia prevented by the vaccination. The general medical costs are the costs incurred during added years of life by persons who would otherwise have succumbed to pneumonia.

The inclusion of long-term treatment costs on a parallel basis might prove controversial. Do we really mean to imply that policy should be formulated with an eye to how much society will have to spend in the future, if an intervention succeeds in saving a life? The uncomfortable conclusion may be yes. If it is cost-effectiveness we seek in medical care, then not immunizing those with expensive chronic diseases might be the preferred policy, even though the immunization might yield them more expected QALYs than it would those who were healthy. The disadvantaged would be denied precisely because their disadvantage will persist in the future.

Even if the existing pattern of health expenditures has been optimized so that at the margin all individuals receive as many QALYs per unit of expenditure, the problem when considering a new potential intervention is that preserving an individual carries along his average QALYs and

dollar expenditures, not his marginal quantities. If the newly contemplated activity itself entails an expense, we will direct it first to individuals who offer us large inframarginal surpluses.

A related question is how we should account for the non-medical costs of keeping a person alive, his tax contribution, his drain on social services, etc. Should he survive, his own utility from the resources he expends on himself would be captured in his QALY stream. If he provides (drains) resources for the rest of society, the QALYs they generate to (cost) others should be added to the individual's QALY stream to get the net benefits to all from his survival.

Consider the extreme case of perfect observability in which each internally homogeneous category could be considered separately for an intervention. If we seek maximum benefits for our budget, we should simply compute the cost-effectiveness value separately for each category. For a category k, we would compute expected costs with and without the treatment; the difference would yield net costs. Then we would do the same with benefits. The cost-effectiveness of applying the treatment to individuals in this category would be

$$(7) \qquad CE_k = \frac{\text{Net costs}_k}{\text{Net health benefits}_k}.$$

Those categories with the most favorable (i.e., lowest) ratio should be accorded the highest priority to receive the intervention.

Dealing with latent heterogeneity within categories. Realistically, it will be impossible to classify the population finely enough to eliminate all within-category heterogeneity. In most of the examples mentioned above, the presence of some important variables could be inferred

only by observing responses to interventions. In other instances, physical evidence only becomes available too late to use for predictions. For example, autopsies of soldiers killed in Vietnam showed that 45 percent of these young men already had some evidence of atherosclerosis, and five percent had severe atherosclerosis (McNamara et al. 1971). The ability to classify individuals on this risk factor prospectively could greatly help us in our ability to intervene to moderate or prevent coronary heart disease. Unfortunately, there is presently no practical mass screen for monitoring of atherosclerosis. Its differential presence, among even young men, is a creator of heterogeneity.

To conduct an assessment when latent heterogeneity remains within categories, we must take the same approach to the population within each category as we did earlier to overall latent heterogeneous populations to make predictions and draw inferences. Within each category we start by identifying the prevalence for each (unobservable) group, r_j, and its associated costs and health benefits. Then we compute a standardized assessment for each of the costs and benefits for the entire category. Finally, the cost-effectiveness ratio is computed as average net costs divided by average net health benefits. Thus, we have

$$CE = \frac{\text{average net costs}}{\text{average health benefits}} = \frac{\sum_j r_j \cdot \text{net costs}_j}{\sum_j r_j \cdot \text{QALYs}_j}.$$

(Note that we do not compute a weighted average of the cost-effectiveness ratios for the different groups.)

Although there may be within-category heterogeneity, the procedure yields only a single cost-effectiveness ratio that tells us the marginal return of offering it to all or none within the category.

Nevertheless, although different groups within a category cannot be made the basis for policy, it is important to identify them in order to make an accurate assessment of both benefits and costs.

The first portion of this paper addressed the probable bias with health benefits. We shall also want an accurate assessment of costs, for our real concern is the bias in cost-effectiveness ratios if a traditional assessment if employed. If those who differentially survive because of the intervention in general cost more per QALY (what we might think of as the normal case), then there will be a downward bias--the intervention will be regarded too favorably--and vice versa. A weighted average of cost-effectiveness ratios for the different risk groups summarizes the bias. It is important to note, however, that the weights are not the relative prevalences of the different groups, but rather the relative total QALYs each group offers; that is, CE_k is weighted by $r_k \cdot QALYs_k$. This procedure is equivalent to dividing average costs (weighted by prevalences) by average health benefits (similarly weighted).

Constraint on expected costs. Let us put our cost-effectiveness approach into practice. Consider an extreme example of an intervention that reduces an individual's immediate mortality, but has no effect on his morbidity or subsequent mortality. To which group should it be offered first? To answer this question we shall have to have four pieces of information: (1) the change in probability of survival for each risk category, call it δ_j; (2) future discounted QALYs within that category for a survivor, $QALYS_j$; (3) future discounted costs for an individual in that category, $COSTS_j$; and (4) the intervention costs (including costs of side effects). If the goal is to maximize total health benefits within a budgetary constraint on total health care costs, then cost-effectiveness ratio CE_j for ordering risk groups is a special case of (7):

$$CE_j = \frac{\delta_j x\ COSTS_j + INTERVENTION\ COST}{\delta_j x\ QALYS_j}$$

The interventions should be offered to all risk groups for which CE lies below some established cutoff, the shadow price of QALYS given the budget.

Constraint on the level of intervention. In some circumstances resources other than dollars may be constrained. Particularly under present regulatory conditions, hospital days, number of operations, and physician time could each be a constrained resource. Under such conditions, it is no longer appropriate to employ traditional cost-effectiveness ratios comparing dollar cost and health effectiveness. Say the constraint is on the number of procedures undertaken, where δ_j, the decrease in current mortality, is the only health effect. In this case it is appropriate to multiply the δ_j's by the net benefits that are offered in each group. (Note they are not multiplied by cost-effectiveness ratios.) Such net benefits are appropriately computed using a shadow price on QALYS, λ. Thus,

$$(9) \qquad EXPECTED\ NET\ BENEFITS_j\ (in\ dollar\ terms) = (\lambda\ QALYS_j - COSTS_j) \times \delta_j.$$

Using the appropriate procedure could reverse the preference ordering indicated by cost-effectiveness ratios in either direction--that is, toward the high-risk or the low-risk group. Suppose an intervention reduced mortality in the high-risk category by 10 percent, and in the low-risk category by only five percent. Each QALY is assigned a shadow price of 1. As depicted in Table 3, with the number of persons who can

Table 3. Policy Choice with Constraints on the Level of Intervention

	Example A			Example B		
	Survivors		EXPECTED NET BENEFITS	Survivors		EXPECTED NET BENEFITS
	QALYS	COSTS		QALYS	COSTS	
High Risk	10	2	[.8]	10	9	.1
Low Risk	11	1	.5	11	7	[.2]

be offered the intervention constrained, it is appropriate to give the intervention first to the High Risk category in Example A and to the Low Risk category in Example B, as the boxes show.

(Table 3 about here)

An analyst who was entranced with the traditional CE ratio approach might mistakenly employ expression (8) to establish priorities. Consider a situation where the intervention cost was low. The analyst would assign the intervention first to the Low Risk category in A and to the High Risk category in B, exactly the opposite of the correct assignment defined by (9). Priorities depend upon which resource is constrained.

The same principles apply if interventions reduce morbidity, hence present treatment costs. The formulas are only a trifle more complicated.

B. Ethical Issues

Throughout history, the cornerstone of the medical ethic has been to provide for each individual the medical care that offers him the most favorable prospects. Although the rising costs of medical care have been a subject of concern since the Commission on the Cost of Medical Care of the 1930s, public policy has rarely explicitly confronted the tradeoff between health and resources. Even within current cost containment efforts, such as Certificate of Need and PSROs, the stated goal has been to foster efficiency without a sacrifice in health, rather than to save resources through the purchase of less health. Nevertheless, the notion that health might be too expensive at the margin in some circumstances appears to have been the implicit justification for many programs designed to contain medical expenditures.

The Central Ethical Problem. Once factors other than the promotion of health status enter policy decisions, as they do if resource costs

are considered at all, a central ethical problem arises. Some policies beneficial to health will be accepted, and others will not. Some may be accepted for certain classes of individuals, but not others. Society, in effect, will refuse to spend resources to provide particular health benefits for particular individuals.

On what ethical basis can such a refusal be made? Are we justified in weighing resource costs to society as a whole against health benefits to an individual, and if so, on what basis should we do the weighing? In this paper we have approached a number of problems of this sort using the tools of cost-effectiveness. But we have not justified that approach, nor have we explored its implications.

Let us take as a premise that policy priorities must be set, and that some interventions that offer some health benefits will have to be forgone. We see three bases on which a set of policies could be justified. First, if there were some widely accepted mechanism for calibrating and valuing health benefits in dollar terms, we could merely apply it and choose the set of policies that maximized expected total societal benefits. (This is the approach of benefit-cost analysis.) An appropriate mechanism for such valuations does not now exist, however, and is not likely to in the near future.

A second basis for discrimination would emerge if all members of a society agreed to a set of policies for health promotion (and for financing those policies), weighing in their own minds the associated benefits and costs. Such agreement is unlikely to be forthcoming, however, for reasons linked to the theme of the paper. Heterogeneity in the population, if observable, will hurt us if consensus is our goal. The candidate for a heart-bypass operation would certainly like to have

that procedure covered by health insurance, but most of society might think the benefits not worth the resources entailed; the 60-year-old bachelor might not support an immunization campaign against childhood diseases, whereas young mothers would be staunch partisans.

A third approach might be possible if we could get around the disagreement problem that arises because each individual judges from his own position. Suppose we could return all individuals to a hypothetical original position, where no one knew who he would end up to be, and each attached equal probability to the possibility that he would be each member of society. We could ask a representative citizen at this position what mixture of policies he would prefer. Presumably he would seek to maximize his overall expected welfare, taking both health and resource costs into account. In effect, before he is conceived, before his genetic and environmental factors become set, we would ask an individual to lay out a health-protection plan for his whole life, not knowing whether he would be born with a birth defect, contract cancer, live in a polluted city, or work as a truck driver. He would be free to choose how much money would be devoted to medical care in any circumstance, but would know that he would have to pay for that care--through taxes, insurance, or out of pocket. His program would include items that offered high benefits relative to their costs. In essence, he would be using some form of cost-effectiveness analysis to maximize his expected utility, not knowing his future identity. As a real-life approximation to an "original position" we might think of a committee of young faculty members at a university trying to draw up a health plan for the junior faculty. They would not know who would be paralyzed in an auto accident, who would have

a baby with spina bifida, who would want to spend many weekly hours with a psychiatrist.

Whatever ethical sustenance one secured from an original position argument, and however far back that original position extended, one should not expect that such policies would be widely accepted once the world played out its lotteries, i.e., once heterogeneity revealed itself, identifying who was healthy, whose health was threatened, and who was sick with which disease. A program designed on cost-effectiveness principles might lead to a highly unequal ex post result. Even though we could point out that the set of outcomes merely reflected the gambles that a disinterested individual would have been willing to take, we should still expect the actual losers to be grumblers. Consider an individual whose unfavorable health prognosis made it undesirable to spend more resources to increase his chance for short-run survival. He could justifiably complain that his welfare is being disregarded.

Our thought experiment started by placing the individual in a contingent claims market for resources, the states of the world to depend on his health. The discomforting reality is that following the efficiency dictate of securing an equal marginal return from all expenditures, he might find it optimal to transfer more resources towards high-health rather than low-health states. If so, once lotteries are played out, arguments about ex post inequity will arise. With contingent claims markets for money, in contrast to health, the individual will almost invariably wish to allocate his major transfers to situations where he would otherwise have low income. The ethical problem of transfering more to the better as opposed the worse off does not arise. So too, it would be more

convenient if the nature of the health production process were such that expected utility maximization required allocating resources in a pattern inversely proportional to health status. Any lack of negative covariance between health state and the return to expenditure generates automatic dissatisfaction.

In practice, we should expect society to respond to such dissatisfaction and the accompanying charges of inequity. Through its political processes, society would attempt to devote additional resources to poor-outcome states even though the return to those resources was below what could be secured in high-outcome states. Moreover, in contrast to the framework we laid out above, we would expect that policymakers would play down the importance of anticipated future expenditures for medical care in the decision process. That is, even though we might have accepted the lottery before the fact, we would smooth out dissatisfaction afterwards through an inefficient overallocation to poor-outcome states. Natural redistributional proclivities would reinforce any tendencies in this direction.

Social Acceptability of Categories. A number of the categories that are most useful in making medical predictions, notably sex and race, have become lightning rods for debate in a variety of policy areas. We can foresee circumstances where strong political pressures, clothed in ethical trappings, would be brought to bear requiring or prohibiting that medical priorities be set or not be set on the basis of such categorizations.

It is not beyond belief that there may be social pressures to stop classifying data or basing medical predictions on these highly charged variables. This is particularly likely if such predictions may ultimately

influence policy choices, hence the expenditure of public funds. Forces for knowledge suppression will be more difficult to counteract if, as seems likely, the classifying variables, though correlating with risk, are either not causal or cannot be shown to be causal.

Appropriate output measure. Even if we accept cost-effectiveness analysis unquestioningly, and can avoid conflicts on appropriate risk categories, we still do not escape ethical dilemmas. What output measures should be employed as we assess the benefits of interventions?

An alternative quite different in spirit from the original position approach might suggest that the QALY, or some other indicator that could be calibrated appropriately, is a basic unit of accounting. Not to maximize QALYs would be to throw away a resource that society valued. Thus, we should be converting the outputs of health policies into QALYs, just as we often convert alternative energy supplies into barrels of oil when making energy policy decisions. A major difficulty with a universal unit approach is that QALYs are not readily transferable from individual to individual. As we choose among health policies, we are basically shuttling QALYs from one individual to another. Securing maximum total QALYs and then dividing up the pie is not a possibility.

A second difficulty relates to appropriate ways for tallying QALYs within an individual. Apart from discounting considerations, should individuals be (or are they) risk averse on the total health received over a period of years? The QALY measure merely adds across years, perhaps with discounting.[6] Utility functions for longevity and health status can contain risk aversion, but they are more difficult to interpret (see Pliskin, Shepard, and Weinstein 1980). But risk aversion may raise other ethical questions. We might find ourselves denying

additional years of health to individuals who had already lived a long time, to provide lesser sums of comparable years to individuals who were younger. Would that be appropriate? We have no resolution for these problems; we merely wish to suggest that intriguing ethical issues persist.

Process. If the resources involved were small, as they may be with any single intervention, then we could afford the luxury of avoiding triage decisions. However, the costs of life preservation across all areas of our modern technological society are becoming so great that protecting ourselves against unpleasant aspects of our value system may be too costly. When the point is reached at which self-delusion is no longer worthwhile, if it has not been already, we believe that a policy of expected QALY optimization per dollar of expenditure will garner strong intellectual support.

It would have been more comfortable for us to avoid this issue, as most cost-effectiveness analyses have in the past. Given the vibrant societal debate about cost-effectiveness and risk-benefit analyses, ducking did not seem desirable.[7] If the Secretary of Health and Human Resources has to decide whether to provide federal funds for heart transplantation, as Secretary Patricia Harris has recently, we should be willing to discuss the principles that underlie such decisions. Often we can avoid the most discomforting aspects of life-versus-resource decisions. If we can remove the resources before we know who would receive them, then in effect we can capitalize on an original position type of self-interest on the part of citizens who concur. Thus, we can close down hospital beds or limit CAT scanners. (Here is not the place to judge the effectiveness of such programs for ultimate resource savings.) However, if individuals had a good idea of who the beneficiaries of those resources

might be, prior limitations might be more difficult. Thus, we would have a harder time eliminating a machine that benefited an identified group, such as asthma patients, than a CAT scanner that offered the same total benefits per year at the same cost. As Schelling (1968) has eloquently stated, an identified life is often valued more highly than a statistical life.

Voluntary versus involuntary behavior. One final element may be relevant to our ethical analysis. When we devise our optimal portfolio of policies, we consider whether an individual is denied some interventions to improve his health because of actions over which he had no control. If a choice must be made, most citizens would prefer to grant an intervention to an individual who incurred a respiratory disease because he lived in a polluted area than to one who contracted the disease by smoking. Our society seems to believe that individuals should suffer the consequences of their sins, but not of variables over which they have no control.

In practice, the determination of what behaviors are voluntary is quite complex. Consider smoking. Can we be confident that there is not a strong genetic predisposition towards smoking, at least for some individuals? And would we really say that smoking is voluntary if environmental factors beyond an individual's control, such as smoking parents, strongly predicted smoking behavior? From an economic standpoint perhaps the best way to judge the voluntariness of any behavior would be to determine how much an individual engaging in it would have to be paid to reduce or give it up. Alternative numeraires would convey different impressions. To get a numeraire other than dollars, we could look at a smoker's hourly wage, and determine how many minutes of work he would do

for a cigarette. (Because smoking is habit forming, we might also inquire about willingness-to-pay before a person has taken up the practice. Voluntariness may depend on the time that a behavior is examined.)

Denying or offering a health program based on an individual's behavior will have the side effect of changing incentives. Thus, denying an intervention to a smoker will in effect raise the price of smoking, which could have an effect on behavior if the elasticity of demand for smoking were high. In examining the effects of a pattern of interventions, we must look not only at what they accomplish vis-a-vis the status quo, but also how they change the status quo by affecting individuals' behaviors.[8]

We already tax smoking. Denying some medical interventions would represent a further tax.[9] From the standpoint of efficiency, we want to make sure that the total tax equals the social cost of smoking, which depends largely on the subsidized health care services smokers receive.

Age is obviously an uncontrollable variable, and on that basis many would find it morally objectionable as a basis for assigning priority to medical benefits. Our view, however, is that it may be one of the less objectionable variables, assuming that society is designing a fairly stable set of interventions over time. Each individual has his opportunity to be young, when he may be given priority for some interventions, and old, when he may be denied them. (The reciprocal argument applies, of course; with the pneumonia vaccine the intervention is provided first to the old.) Providing less in the way of medical resources to be old is not unlike providing them less in the way of educational resources. Those resources simply provide more value when the individual is young.

Summary. In medicine there is a well established tradition of basing predictions on factors such as age, sex and race. This tradition, however, is accompanied by the ethic that the regimen that is medically best for the patient should be prescribed. This tradition may be challenged if constraints on resources force us to forego some medically beneficial, albeit cost-ineffective treatments. Attention to medical expenditures is now forcing us to set priorities.

Any apparatus that selects one health-promoting policy over another raises profound ethical questions. Common analytic tools for making such choices, such as cost-effectiveness analysis, though often innocuous in appearance, carry strong moral overtones. Such techniques are often justified in policy debate on the grounds that they are the only logical or rational method of choice. Whatever the validity of that argument, it alone is unlikely to be sufficient to carry the day in the highly contentious policy arena. Direct ethical support is needed as well, and could perhaps be elicited by an original position argument--that is, by persuading individuals that they themselves would choose a cost-effectiveness approach were their own distinctive characteristics and interests still undetermined.

V. THE CHOICE OF HEALTH POLICIES

Given an understanding of the concept of heterogeneity, how should we design policy interventions? We illustrate the relevance of our concepts by considering a pneumonia vaccination program. We then summarize what strike us as the more important policy implications of our analysis.

A. Pneumococcal Pneumonia Vaccine: An Illustration of Heterogeneity

The pneumonia vaccine example illustrates several concepts. (1) Medical research can generally identify numerous risk factors for elevated incidence and/or case fatality from a disease (observed heterogeneity); for some factors, the effects can be quantified; other risk factors can only be assessed qualitively. (2) The cost-effectiveness framework can be applied to observed classes for which quantitative data are available to establish priorities among them for an intervention. (3) Within standard categories of observed heterogeneity (such as age), there is substantial latent heterogeneity which biases traditional assessments of benefits. (4) Assessments that control for observed heterogeneity, but not for unobserved characteristics, estimate the benefits and cost-effectiveness of an intervention too favorably.

Before proceeding with the analysis of these applications, a factual background on pneumonia and vaccination is helpful.

Pneumococcal pneumonia vaccine. "Pneumovax" is a vaccine designed to prevent the 14 most common valences of pneumococcal pneumonia; it was licensed by the Food and Drug Administration in 1977. A cost-effectiveness analysis of the vaccine by the Office of Technology Assessment (OTA 1979; Willems et al. 1980) provides much of the foundation for

this example. By conservative estimates pneumococcal pneumonia is responsible for between 5,000 and 17,000 deaths per year in the U.S. (Willems et al. 1980).

Age is an observable risk factor with marked influence on both the incidence and case fatality rate. The largest set of population-based data comes from Massachusetts for 1921 through 1930, during which time lobar pneumonia was a reportable disease (Heffron 1979, pp. 300-305). Age patterns for "lobar pneumonia" can serve as a reasonable indicator for all pneumococcal pneumonia.[10] Using Heffron's (1979, p. 299) assumption that reported cases were about half of total cases, we infer that the attack rate per year in the Massachusetts data rose steadily with age in adults, from 13 per 1000 persons aged 20-29 to 80 per 1,000 persons aged 80 and above. Data from this pre-antibiotic era reveal that the case fatality rate also rose with age. The combined effect of both factors is that the age-specific death rate due to lobar pneumonia in persons aged 80 and above was 26 times as high as that for persons aged 20 to 29. Recent data confirms that similar patterns still hold for the antibiotic era. Incidence rises with age (beyond the teens) (Oseasohn et al. 1978), as does case fatality (Sullivan et al. 1972; Austrian and Gold 1964).

A number of other observable characteristics have been found to be associated with differences in the age-specific death rate from pneumonia (and presumably pneumococcal pneumonia in particular). These include sex (males are higher), race (non-whites are higher), overcrowded housing (Heffron 1979, p. 316), probably alcoholism (Heffron 1979, p. 158), fatigue, acute infections such as influenza, and chronic diseases such as infections of the respiratory tract and cancer (Heffron 1979, p. 335).

Austrian and Gold (1964) and Mufson (1974) found that case fatality rates from bacteremic pneumococcal pneumonia were above average in persons with chronic lung disease, chronic heart disease, chronic renal failure, diabetes mellitus, and other metabolic disorders. In view of such associations, the approved indications for administering Pneumovax include "persons having severe chronic physical conditions such as chronic heart disease, chronic bronchopulmonary disease, chronic renal failure, diabetes mellitus or other chronic metabolic disorders; persons in chronic care facilities; persons convalescing from severe disease..." (OTA 1979, p. 54).

Application of observed heterogeneity: disaggregating by age. Using vital statistics data, the OTA study (1979) computed the expected health gain from vaccination against pneumococcal pneumonia as a function of age in terms of discounted quality-adjusted life years. Averting death in a younger person generally confers a larger remaining life expectancy than in an older person. Nevertheless, for pneumococcal pneumonia vaccination, the increase in discounted QALYs for a randomly chosen person rises with age. The gain for a young adult aged 25-44 is 3.7 healthy hours (.00042 adjusted years); for a person aged 65 and above, the gain is ten times as large--38.1 healthy hours (.00435 years).

Application of the cost-effectiveness paradigm. The OTA pneumonia vaccine study employed age as the primary risk category. The cost-effectiveness of pneumococcal pneumonia vaccination, according to their results, improves with age from $77,200 per QALY for children ages 2-4 to $1,000 per QALY for persons aged 65 and over. The policy implication is clear. Vaccination should probably be encouraged for persons aged 65 and above, but probably not recommended for children. Under current FDA-

approved indications, the vaccine is recommended for persons aged 50 and over (OTA 1979, p. 54).

Analyses derived from the OTA (1979) study show how policy-relevant risk factor classification can improve policy performance dramatically. If the pneumococcal pneumonia vaccine were offered to all persons aged 50 and above, and age-specific acceptance rates were similar to those observed for influenza vaccine in 1975, then a national program for the U.S. would cost $50 million and return a total of 27,200 QALYs. (Using linear interpolation, this calculation allocated three-fourths of the costs and benefits of the age category 45 through 64 to persons 50 and above.) If age could not be used for allocating vaccinations, for example because it was considered an invidious basis for discrimination, and the vaccine were offered to all persons regardless of age, the cost would be $150 million (three times as high), while the number of QALYs would be increased by only 15 percent.

CE ratios for other observable risk factors for pneumoccocal pneumonia. In the absence of quantitative data relating chronic disease and risk of death due to pneumococcal pneumonia, the OTA reported (1979, p. 77) an illustrative hypothetical calculation. Persons with selected cardiovascular, bronchopulmonary, renal, and metabolic diseases were termed "high-risk" persons. It was assumed that their instantaneous mortality rate was five times as high as that for the general population, a factor derived from Fitzpatrick, Neutra, and Gilbert's (1977) study. The proportion of deaths due to pneumococcal pneumonia in high-risk persons was assumed to be 1.8 times as high as in the general population. The OTA illustration found, consistent with the FDA-approved indications

for the vaccine, that vaccination was more cost-effective in high-risk persons than in low-risk persons. Among persons aged 25 to 44, for example, the illustrative cost-effectiveness ratio was $7,300 per QALY in high-risk persons, compared to $33,000 per QALY in low-risk persons.

Latent heterogeneity within age categories: effect on estimated benefits. Even if the vaccine were offered to all persons of a given age regardless of risk status, stratification by risk group would be important when estimating the benefits. For a 50-year-old male, for example, we calculated that failure to stratify by other risk factors overestimates the gain in undiscounted life expectancy from the vaccine by about 20 percent. We interpolated this estimate from Table 4 in Shepard and Zeckhauser (1980a).[11] If mortality were the outcome measure, rather than life expectancy, the bias would be even more substantial, particularly at advanced ages. The bias in mortality at age 80, for example, is three times as large as the bias in life expectancy (Shepard, 1977).

Latent heterogeneity and the bias in future health cost estimates. Within each age group, high-risk persons are those with other chronic conditions. A vaccination program increases the proportion of high-risk persons among survivors compared to what it would have been in the absence of intervention. Because of their chronic diseases, high-risk persons are likely to incur higher medical costs than the general population of the same age. Residents of chronic care facilities provide an extreme example. They are expected to reap large benefits from the vaccine (they were specifically mentioned in the FDA-approved indications for use), but their medical costs are probably five to ten times higher than others of the same age without the chronic condition.

In the OTA (1979) study, high-risk persons were estimated to comprise 17.3 percent of the population aged 25 to 44. (Bell 1980.) Suppose the lifetime average prevalence of high-risk persons from a standardized assessment (weighted by the number of survivors and discounted) were ten percent compared to five percent from a traditional assessment. Assume that the expected medical costs of the high-risk persons were five times that of low-risk persons. Expected medical costs for the mixed population would be 1.5 times that of low-risk persons under the standardized intervention compared to 1.25 times under the traditional assessment. Thus standardized cost-effectiveness estimates would be 20 percent (.25/1.25) more favorable than traditional estimates due to the bias in cost estimation alone. If the effectiveness bias were also overstated by 20 percent of the true value (as in the example above), the total bias would be approximately 44 percent too favorable.

B. Concluding Remarks

Heterogeneity among members of the population in their responsiveness to interventions--both beneficial and detrimental--is a central issue for policymaking. Battery plants are forced to make themselves safe for female workers in child-bearing years, and air-pollution standards are set supposedly to protect the most susceptible members of the population. Flu innoculations are dispensed according to a priority schedule based primarily on medical need. Regulatory and reimbursement policies for health care may start by examining the consequences for health and resources of offering different procedures to different categories of individuals, and then try to channel patients and providers in cost-effective directions. Society is increasingly confronting the salient issue of crafting policies that recognize heterogeneity within the population. This analysis provides some lessons and principles that might make that confrontation more productive.

We hope to have demonstrated that: (1) population heterogeneity may be an important factor even when heterogeneity is latent; (2) traditional methods for predicting the benefits of interventions in populations with latent heterogeneity are biased; (3) the bias generally causes us to overstate the benefits and cost-effectiveness of helpful interventions; (4) attention to latent heterogeneity can improve inferences and extrapolations about the benefits alternative policies will provide to populations; (5) observed heterogeneity raises interesting efficiency and equity issues in setting priorities for receipt of interventions; and (6) calculations attending to heterogeneity are feasible as a guide in making policy choices.

APPENDIX A

A COMPARISON OF TRADITIONAL AND APPROPRIATELY STANDARDIZED ASSESSMENT PROCEDURES

This appendix proves that under common assumptions, the traditional assessment underestimates mortality and overestimates life expectancy under beneficial intervention. Theorem 1 was proved in the text.

After three lemmas, Theorem 2 presents two important extensions of Theorem 1. Under two very common models of mortality, the multiplicative and logistic models, the covariance is positive so that the traditional assessment overstates the reduction in mortality. Finally, it can easily be shown that life expectancy behaves inversely to mortality. We state as a corollary to Theorem 2 that these two models overstate life expectancy. Similarly, they overstate gains in life expectancy. This statement applies also to expected utility and to quality-adjusted life expectancy.[12] Thus, if the traditional assessment gives too low an estimate of mortality, it gives too high an estimate of life expectancy.

Lemma 1[13]

Let

$$|\ell_{jk}| = \begin{vmatrix} \ell_{1j} & \ell_{1k} \\ \ell_{2j} & \ell_{2k} \end{vmatrix} .$$

Then

$$\Delta m(x) = \frac{1}{\ell_{1.}\ell_{2.}} \sum_{j} \sum_{k>j} r_j r_k (\mu_{2k} - \mu_{2j}) |\ell_{jk}| .$$

Proof

By definition,

$$\Delta m(x) = \sum_j \mu_{2j} \frac{r_j \ell_{2j}}{\ell_{2.}} - \sum_j \mu_{2j} \frac{r_j \ell_{1j}}{\ell_{1.}} \quad .$$

Now substitute the definition of $\ell_{i.}$ (from eq. (2)) and write the result over a common denominator.

$$\Delta m(x) = \frac{1}{\ell_{1.}\ell_{2.}} \left[\sum_j \sum_k \mu_{2j} r_j r_k \ell_{2j} \ell_{1k} - \sum_j \sum_k \mu_{2j} r_j r_k \ell_{1j} \ell_{2k}\right] .$$

Now the first double summation above can be rewritten with j and k reversed, since the indices of summation are symmetrical. The sum becomes

$$\sum_j \sum_k \mu_{2k} r_j r_k \ell_{1j} \ell_{2k} \quad .$$

Now the two double sums can be combined, yielding

$$\Delta m(x) = \frac{1}{\ell_{1.}\ell_{2.}} \sum_j \sum_k (\mu_{2k} - \mu_{2j}) r_j r_k \ell_{1j} \ell_{2k} \ .$$

For j = k the summand vanishes. We can group the terms with each combination of different subscripts to obtain

$$\Delta m(x) = \frac{1}{\ell_{1.}\ell_{2.}} \sum_j \sum_{k>j} (\mu_{2k} - \mu_{2j}) r_j r_k (\ell_{1j}\ell_{2k} - \ell_{2j}\ell_{1k}) \ .$$

Rewriting the last factor as a 2×2 determinant yields the required formula. Q.E.D.

Lemma 2

Let risk groups j be numbered in order of increasing risk in the absence of treatment at the initial time x_0, $\mu_{1j}(x_0)$. Assume that this numbering ranks risk groups in order of increasing risk for all ages x,

$x_0 < x < x_1$. Let treatment lower mortality according to a multiplicative model, i.e., for all x

(A-1) $$\mu_{2j}(x) = a\mu_{1j}(x) \qquad \text{where} \qquad 0 < a < 1 .$$

Then for $k > j$, $|\ell_{jk}| > 0$.

Proof

Inserting (A-1) into the definition of the survival function (1) gives

$$\ell_{2j} = \exp\left[- \int_0^x a\, \mu_{1j}(t)\, dt \right] ,$$

so

$$\ell_{2j} = \ell_{1j}^{\,a} .$$

Thus

$$|\ell_{jk}| = \begin{vmatrix} \ell_{1j} & \ell_{1k} \\ \ell_{1j}^a & \ell_{1k}^a \end{vmatrix} = \ell_{1j}\ell_{1k}^a - \ell_{1j}^a\ell_{1k} = (\ell_{1j}\ell_{1k})^a[\ell_{1j}^{(1-a)} - \ell_{1k}^{(1-a)}] .$$

The ordering of the risk classes implies that for $k > j$, $\mu_{1j}(x) < \mu_{1k}(x)$ so $\ell_{1j} > \ell_{1k}$ and $\ell_{1j}^{(1-a)} > \ell_{1k}^{(1-a)}$. Thus $|\ell_{jk}| > 0$. Q.E.D.

Lemma 3

Assume the risk groups j can be ordered as in Lemma 2. Divide the interval (x_0,x) into n equal intervals. Let $\ell_{jk}^{(m)}$ be the survival matrix on the mth interval. Assume that on each interval m, treatment lowers mortality according to a logistic model, i.e., treatment lowers the odds of mortality by a constant fraction, or for an interval of Δx starting at time x,

$$q_{2j} = 1/[1 + \exp(b_j+\alpha)\ ] \quad ,$$

where b_j is defined such that

$$q_{1j} = 1/[1 + \exp(b_j)]$$

and $\alpha > 0$. Let ℓ_{jk}^m be the survival matrix on the mth interval and let ℓ_{jk} be the composite survival matrix over the interval (x_0,x) with element ℓ_{ij} defined by

$$\ell_{ij} = \prod_{m=1}^{n} \ell_{ij}^{(m)} .$$

Then $|\ell_{jk}| > 0$.

Proof

We proceed by induction. Let $\ell_{jk}^{(M)}$ denote the composite survival matrix over the first M intervals, so $\ell_{ij}^{(M)} = \prod_{m=1}^{M} \ell_{ij}^{(m)}$.

1) First, we establish that if M = 1, $|\ell_{jk}^{(1)}| > 0$. To show this we substitute

$$|\ell_{jk}^{(1)}| = \frac{\exp(b_j)}{1+\exp(b_j)} \times \frac{\exp(b_k+\alpha)}{1+\exp(b_k+\alpha)} - \frac{\exp(b_j+\alpha)}{1+\exp(b_j+\alpha)} \times \frac{\exp(b_k)}{1+\exp(b_k)} .$$

The above expression may be rewritten with a common denominator as

$$[\exp(b_j+b_k+\alpha)]\Big\{[1 + \exp(b_j+\alpha)][1 + \exp(b_k)] - [1 + \exp(b_j)][1 + \exp(b_k+\alpha)]\Big\},$$

divided by the product of the four denominators above. Since the denominator and the first factor in the numerator are all positive, the sign of $|\ell_{jk}^{(1)}|$ is identical to the sign of the factor in braces, which simplifies to

(A-2) $$[\exp(b_k)][1 + \exp(b_j-b_k)\exp(\alpha) - \exp(b_j-b_k) - \exp(\alpha)] .$$

The steps below establish that the factor in braces is positive: Since for $k > j$, $g_{ij} < g_{lk}$, $b_j > b_k$. Thus

$$1 - \exp(b_j - b_k) < 0 .$$

Further, $\alpha > 0$ implies

$$1 - \exp(\alpha) < 0 .$$

The second factor in (A-2) is the product of the two inequalities above, now shown to be positive. This completes the proof for $M = 1$.

2) Next we establish that if $|\ell_{jk}^{(M)}| > 0$, then $|\ell_{jk}^{(M+1)}| > 0$. We define the matrices

$$A = \ell_{jk}^{(M)}$$

$$B = \ell_{jk}^{(m+1)}$$

and

$$C = \ell_{jk}^{(M+1)} .$$

Then $C_{ij} = A_{ij}B_{ij}$. Since $|A| > 0$ by hypothesis,

(A-3) $$A_{11}A_{22} - A_{12}A_{21} > 0 .$$

Since B represents the survival matrix for a single period under a logistic model, $B > 0$ by an argument identical to the one above for $|\ell_{jk}^{(1)}| > 0$. Thus

(A-4) $$B_{11}B_{22} - B_{12}B_{21} > 0 .$$

Since all the elements of logistic model survival matrices are positive, the factors $B_{11}B_{22}$ and $A_{12}A_{21}$ are both positive. The inequalities (A-3) and (A-4) may be multiplied by positive factors without changing their direction. Multiplying (A-3) by the first factor and (A-4) by the second and adding the resulting inequalities gives

$$A_{11}A_{22}B_{11}B_{22} - A_{12}A_{21}B_{12}B_{21} > 0$$

or

$$C_{11}C_{22} - C_{12}C_{21} > 0 .$$

Thus $|C| > 0$. This completes the proof of the inductive part of the lemma. Q.E.D.

Theorem 2

Assume the risk groups can be ordered as in Lemma 2, and assume that treatment lowers mortality according to a multiplicative model (as defined in Lemma 2) or a logistic model (as defined in Lemma 3). Then

$$\Delta m(x) > 0 .$$

Proof

Under either the logistic or multiplicative models, Lemmas 2 and 3 establish that for $k > j$, $|\ell_{jk}| > 0$. The ordering of risk groups assumes that for $k > j$, $\mu_k - \mu_j > 0$. With $\Delta m(x)$ written as a summation in Lemma 1, each of the summands is the product of positive factors. Thus, $\Delta m(x) > 0$. Q.E.D.

Corollary

Let $\overset{\circ}{e}_{x_0}$ and $\overset{\circ}{e}'_{x_0}$ be the life expectancy at age x_0 using mortality forces $m_2(x)$ and $m_2'(x)$, respectively. Life expectancy is defined by

$$\overset{\circ}{e}_{x_0} = \int_0^{x_0} \ell_{2.}(x)dx \ .$$

Assume treatment lowers mortality according to a multiplicative or a logistic model. Then

$$\overset{\circ}{e}_{x_0} < \overset{\circ}{e}'_{x_0} \ .$$

Proof

From the definition of the survival function in eq. (1), it is clear that lower mortality rates lead to higher survival rates. Life expectancy, as the integral of the survival function, changes in the same direction as the survival function. Q.E.D.

APPENDIX B

MODEL FOR CHANGES IN AGE-SPECIFIC MORTALITY OVER TIME

We assume that the cohort of white males attaining age 30 at the beginning of each decade consists of a specified proportion of normal and weak persons. Within both the normal and weak groups there is an average annual mortality, $\mu_{ij}(30)$, for the decade of age centered at age 30 (i.e., between 25 and 34 inclusive). The risk indicator j indicates a normal (j = 1) or weak person (j = 2). Here the "treatment" i corresponds to the year in which mortality is observed; it is an index of the state of medical technology and environmental conditions.

We further assume that the base mortality force (i.e., the rate at age 30) for each group changes in a log-linear relation with calendar year, i,

$$\ell n\, \mu_{ij}(30) = \ell n\, \mu_{1930,j}(30) + \left[\ell n \frac{\mu_{1970,j}(30)}{\mu_{1930,j}(30)}\right]\left[\frac{i-1930}{1970-1930}\right] .$$

Our model treats an individual's risk category as being fixed for life. It has mortality within a risk category in decade i increasing with age x (in years) according to the Gompertz curve, one of the simplest expressions for mortality,

$$\mu_{ij}(x) = [\mu_{ij}(30)]e^{b(x-30)} .$$

Finally, we assume that the proportion of high-risk persons in the cohort at age 30 in year i, $r_{i2}(30)$ increases in a log linear relationship

with time. (The increase is postulated because reductions in infant and child mortality have presumably benefited weak persons most.)

$$\ell n\, r_{i2}(30) = \ell n\, r_{1930,2}(30) + \left(\ell n \frac{r_{1970,2}(30)}{r_{1930,2}(30)}\right)\left(\frac{i-1930}{1970-1930}\right) .$$

Our model thus has seven free parameters: the parameter b; $\mu_{ij}(30)$ for i = 1930 and 1970, and for j = 1 (normal) and 2 (weak); and $r_{i2}(30)$ for i = 1930 and 1970. (The values for $r_{i1}(30)$ are not free since by definition they equal $1-r_{i2}(30)$.) Using this model, we computed age- and risk-specific mortality rates for cohorts attaining age 30 from 1890 to 1970, inclusive. Using those rates, the proportion of each risk cohort surviving each decade was estimated by

$$\ell_{i+10,j}(x+10) = \ell_{ij} \exp[-10\mu_{ij}(x)] ,$$

where

$$\ell_{ij}(30) = 1 \qquad \text{for all } i,j .$$

The overall survival for an age cohort is the weighted sum of the survival fractions for each risk group, weighted by the proportion of the cohort initially at each risk level,

$$\ell_{1.}(x) = (1-\alpha)\ell_{i,1}(x) + \alpha\ell_{i,2}(x)$$

where

$$\alpha = r_{i-(x-30),2}(30) .$$

From these age-specific survival rates, corresponding average annual mortality rates can be derived:

$$m_i(x) = -\frac{1}{10} \ell n \frac{\ell_{i+10,.}(x+10)}{\ell_{i.}(x)} \quad .$$

Since the risk factor j is not directly observable, only $m_i(x)$ and not $\mu_{ij}(x)$ could be observed.

The test of our model is how closely the calculated $m_i(x)$ replicate reported age-specific mortality rates. We selected as a goodness of fit criterion the weighted absolute percentage error. If $m_i^*(x)$ is the true mortality for age x in year i, the percentage error is $|m_i^*(x)-m_i(x)|/m_i^*(x)$. Reasoning that low rates were more subject to random variation, being based on fewer deaths, we wanted to give greater weight to errors on the higher mortality rates. We therefore chose as a weighting factor for each error $\sqrt{m_i^*(x)}$. Our final criterion was thus the weighted sum of percentage errors. Our sample consisted of mortality rates at the beginning of each decade from 1930 to 1970, inclusive, and for decades of age centered at 30 through 80, inclusive. They are shown in Table C-1. The number of observations was thus 5×6 = 30. Using a non-linear optimization routine[14] we found the following values for the parameters:

$r_{1930,2}(30) = .1085$, $r_{1970,2}(30) = .3849$, $\mu_{1930,1}(30) = 2.4023$,

$\mu_{1930,2}(30) = 19.2184$, $\mu_{1970,1}(30) = 1.8686$, $\mu_{1970,2}(30) = 1.6853$

$b = .0798$

Table A-1

Predicted and Actual Mortality Forces for White Males by Decade 1930 to 1970[a]

Ten Years Centering At Age		Annual Mortality Rates Per 1,000[b]					Percentage Reduction 1930-1970
		1930[c]	1940	1950	1960	1970	
30	model	4.1	3.4	2.8	2.3	1.8	56.1
	actual	4.1	2.8	1.9	1.6	1.8	56.2
40	model	7.2	6.6	5.8	4.9	4.0	43.7
	actual	6.7	5.1	3.9	3.3	3.4	49.3
50	model	12.8	12.5	11.7	10.5	9.0	29.0
	actual	12.6	11.4	10.0	9.3	8.8	29.8
60	model	26.4	25.1	24.1	22.6	20.2	23.3
	actual	25.8	25.1	23.5	22.3	22.1	14.3
70	model	58.5	55.0	51.8	48.9	45.2	22.7
	actual	57.2	54.4	49.5	48.9	49.1	14.2
80	model	130.0	120.0	114.6	107.7	101.0	22.3
	actual	130.0	127.8	111.2	106.4	105.3	19.0

a Actual rates are from National Office of Vital Statistics (1956), National Vital Statistics Division (1963), and National Center for Health Statistics (1974). Annual rates (q) per 1,000 per year were converted to instantaneous rates (μ) per 1,000 per year by the equation

$$\mu = -1000 \, \ell n(1 - q/1000) \, .$$

b Except for 1970, the rate represents the average of three years centered at year indicated. For 1970, the rate was extrapolated at the average of 1968-70 plus the increase from 1967 to 1968.

c Death registration states only (Continental U.S. excluding Texas).

where mortality rates are in deaths per thousand per year.[15] The mortality rates calculated with our model are presented above the empirically observed rates in Table A-1. The weighted average of the absolute errors was 6.23%.

Since reported mortality data are based on census rather than sample data, test of statistical significance are not appropriate to judge goodness of fit.[16] The weighted average error of 6.23% may be contrasted with the fit achieved by a naive model with the same number of parameters--a general quadratic function of i and x (with six free coefficients) multiplied by an exponential function of x, giving a seventh parameter. Using the same non-linear optimization procedure, the polynomial could achieve a fit with a weighted error of 9.27%.[17]

REFERENCES

Anturane Reinfarction Trial Research Group. 1980. Sulfinpyrazone in the prevention of sudden death after myocardial infarction. N Eng J Med 302:250-256.

Austrian, R., Gold, J. 1964. Pneumococcal bacteremia with especial reference to bacteremic pneumococcal pneumonia. Ann Intern Med 60:759-776.

Bell, J.C. 1980. Personal communication.

Fitzpatrick, G., Neutra, R., and Gilbert J.P. 1977. Cost-effectiveness of cholecystectomy for silent gallstones. In Bunker, J.P., Barnes, B.A., Mosteller, F., eds., Costs, risks, and benefits of surgery. New York: Oxford, pp. 246-261.

Fuchs, V.R., 1974. Who shall live? New York: Basic Books, Inc.

Heffron, R. 1979 (reprint of 1939 edition). Pneumonia, with special reference to pneumococcus lobar pneumonia. Cambridge, MA Commonwealth Fund, Harvard University Press.

Kolata, G.B. 1980. FDA says no to Anturane. Science 208:1130-1132.

McNamara, J.J., Molot, M.A., Stremple, J.F., et al. 1961. Coronary artery disease in combat casualties in Vietnam. JAMA 216:1185-1187.

Mufson, M.A., Kruss, D.M., Wasil, R.E., Metzger, W.I. 1974. Capsular types and outcome of bacteremic pneumococcal disease in the antibiotic era. Arch Intern Med 134:505-510.

National Center for Health Statistics. 1974. Vital statistics of the United States 1972. Vol. 2, sec. 5, life tables. U.S. Department of Health, Education and Welfare, DHEW publication no. (HRA) 75-1147. Washington, D.C.: Government Printing Office.

National Office of Vital Statistics. 1956. Death rates by age, race and sex, United States 1900-1953: selected causes. Vital statistics--special reports 43(1). U.S. Department of Health, Education and Welfare. Washington, D.C.: Government Printing Office.

National Vital Statistics Division. 1963. Vital Statistics of the United States 1960. Vol. 2, part A. U.S. Department of Health, Education, and Welfare. Washington, D.C: Government Printing Office.

Neuhauser, D. 1977. Elective inguinal herniorrhaphy versus truss in the elderly. In Bunker, J.P., Barnes, B.A., Mosteller, F., eds., Costs, risks, and benefits of surgery. New York: Oxford, pp. 223-245.

Office of Technology Assessment, U.S. Congress, 1979. A review of selected federal vaccine and immunization policies. Washington, D.C.

Oseasohn, R., et al. 1978. Pneumonia in a Navajo community. Am Rev Resp Dis 117:1003.

Pliskin, J.S., Shepard, D.S., and Weinstein, M.C. 1980. Utility functions for life years and health status. Operations Research 28:206-224.

Schelling, T.C. 1968. The life you save may be your own. In Chase, S.B. Jr., ed., Problems in public expenditure analysis. Washington, D.C.: Brookings, pp. 127ff.

Shepard, D.S. 1977. Prediction and incentives in health care policy. Ann Arbor: Xerox University Microfilms (#77-11744).

Shepard, D.S., and Thompson, M.S. 1979. First principles of cost-effectiveness analysis in health. Public Health Rep 94:535-543.

Shepard, D.S., and Zeckhauser, R.J. 1977. Heterogeneity as a factor in surgical decision making. In Bunker, J.P., Barnes, B.A., Mosteller, F., eds., Costs, risks, and benefits of surgery. New York: Oxford, pp. 56-69.

Shepard, D.S., and Zeckhauser, R.J. 1980a. Long-term effects of intervention to improve survival in mixed populations. J Chron Dis 33:413-433.

Shepard, D.S., and Zeckhauser, R.J. 1980b. The anturane reinfarction trial (letter). N Engl J Med 303:49.

Sullivan, R.J., Dowdle, W.R., Marine, W.M., et al. 1972. Adult pneumonia in a general hospital. Arch Intern Med 129:935-942.

Willems, J.S., Sanders, C.R., Riddiough, M.R., and Bell, J.C. 1980. Cost-effectiveness of vaccination against pneumococcal pneumonia. N Engl J Med 553-9.

Zeckhauser, R.J., and Shepard, D.S. Fall 1976. Where now for saving lives? Law Contemp Prob 40(4):5-45.

Zook, C.J., and Moore, F.D. 1980. The high cost users of medical care. N Eng J Med 302:996-1002.

Zook, C.J., Moore, F.D., and Zeckhauser, R.J. Winter 1980. The high cost users of medical care: patterns, predictability, and public policy. The Public Interest.

FOOTNOTES

1. For certain combinations of rates of increase for an individual's risk and heterogeneity among individuals' risk, there will be no change in observed loss rates over time. Shepard (1977) examined the special case of an exponential increase in the individual's loss rate (at rate k per year) and a gamma distribation for initial loss rates among individuals (with scaling parameter β). There is an infinite number of pairs (where k equals β) for which the two effects precisely balance.

2. To simplify notation, the age variable x in $\mu_{ij}(x)$, $r_{ij}(x)$, $g_{ij}(x)$, and $\ell_{ij}(x)$ will sometimes be omitted.

3. A factor extraneous to our analysis--the eligibility of patients and classification of deaths (Kolata 1980)--has created a further complication in assessing the validity of this particular experiment; we shall leave this problem aside and assume all classifications are correct.

4. It is not automatic that our model would be able to fit the observed data. It is true that we have six parameters (four loss rates which are constant over time, μ_{ij}, and two initial prevalences, $r_j(0)$, for i and j equal to 1 and 2), with which to fit only four pieces of data, the four rates in Table 1. However, the equality constraint

$$r_1(0) + r_2(0) = 1$$

effectively gives another piece of data, and all six parameters are also subject to non-negativity constraints. Nevertheless, since our solution is not a corner solution, an infinite number of parameter

combinations (of one dimension) within a narrow range of the foregoing values are consistent with the data and restrictions. Here, as in other examples, limited population data are not sufficient to identify a structure of risks exactly, but can suggest possibilities that merit further investigation.

5. Strictly speaking, straight addition is only appropriate in a society that organizes itself so that public expenditures are fungible among these three categories, and so that public and private expenditures are equally productive of welfare at the margin. Moreover, prices would have to reflect resource costs. The world is not perfectly efficient in this way. However, we are not prepared to discuss here what nonequal weights should be established for these different classes of resources.

6. In a subsequent analysis we hope to look at issues of discounting. For an individual who has time preference for QALYs, to discount seems unambiguous. However, this seems to us to provide no guidance on how to weight QALYs to different generations.

7. We do duck one point here, the role of income distribution. Our line of argument, for the most part, assumes that individuals have relatively equal incomes, or that for other reasons income distribution is not a primary issue in relation to these decisions. How to deal with income-distributional concerns is a matter for another day.

8. We might wish to judge the taxing of smoking from an original position. Then, if elasticities of response are low, the taxes primarily represent a random financial imposition. Given risk aversion, a population some of whose members might smoke might choose not to have smoking taxes if such taxes had little effect on behavior. (We have

asked a number of academic audiences their attitudes on smoking taxes. Over 90 percent of their members favor such taxes. This is less surprising when we observe that only a small fraction of such audiences smokes.)

9. As a means of raising the overall price of cigarettes, the strategy of denying interventions to smokers as opposed to raising per pack taxes would be reinforced if, as seems likely, the marginal damage of a cigarette is greater for smokers who manifest signs of illness or increased risk, as opposed to those who appear healthy.

In practice, some interventions designed only along grounds of cost-effectiveness, such as the large scale Multiple Risk Factor Intervention Trial, will favor smokers. If that is the predominant pattern, cigarette taxes should be correspondingly higher.

10. Various studies have shown that practically all (96 percent) of lobar pneumonia is pneumococcal (Heffron 1979, pp. 1-2). Moreover, the other major type of pneumonia, bronchopneumonia, is believed to have a similar case fatality rate (Heffron 1979, p. 304).

11. In computing the entries for the table, the odds ratio (approximately equal to the risk ratio) for the risk indicator "other chronic disease" for all causes of death was taken to be five. This is the ratio Shepard (1977) found applied to survivors of a heart attack compared to the general population of the same age and sex, and the ratio Fitzpatrick, Neutra, and Gilbert (1977) used for high-risk candidates for gall bladder surgery. The risk ratio for the "treatment," vaccination, was taken to be 1.012. This ratio is calculated as one plus the product of the share of deaths due to pneumococcal pneumonia (39,600/2,000,000) times the percentage of pneumococcal pneumonias due to valences in the vaccine

(75 percent) times the efficacy of the vaccine against included valences (80 percent) in the OTA (1979) study.

12. See for example Pliskin, Shepard, and Weinstein (1980).

13. Emmett Keeler helped us generalize this lemma to a risk factor with multiple categories.

14. The procedure was steepest descent with synthetic derivatives under the IBM scientific subroutines package.

15. Our fitting procedure did not prevent the mortality rates in the two risk groups from crossing. It so happened with the best-fitting parameters that the mortality of the "weak" risk group actually fell below that of the "normal" risk group in the last decade (between 1960 and 1970). The mortality of the "weak" group could have been constrained not to fall below that of the "normal" group with little degradation of the fit.

16. It is worthwhile to note that our model can also reproduce the backwards J-shaped decline in mortality rates as a function of age. (That is, the decline is greatest at low ages, but greater at the highest ages than for some in the middle.) To get a significant rise for the older ages, the percentages of high-risk individuals, the r_{i2}s, would have to be much greater than those estimated here.

17. The equation is:

$m(x,y) = (.1335x^2 + .1259xy + .0197y^2 - 1.4771x - .9543y + 9.4501)e^{.6848x}$,

where $m(x,y)$ is mortality force per thousand per year, $x = \frac{1}{10}$ (age-30) is age decade, and $y = \frac{1}{10}$ (year-1870) is calendar decade.

There is no unequivocal methodology by which one should compare two quite dissimilar models. That ours, with its logical underpinnings, outperformed the polynomial formulation is reassuring. The triumph is perhaps enhanced because this was the only functional form that was tried for our model.