

# Bioactuarial Models of National Mortality Time Series Data

by Kenneth G. Manton and Eric Stallard

*The incidence and prevalence of chronic degenerative disease in America's elderly population are important determinants of the need for long-term care health services. Though a wide range of data on disease incidence and prevalence is available from a variety of different health studies, a Congressional Budget Office study (1977) concluded that data limitations are a major factor in the lack of precise national long-term care cost estimates.*

*In this paper, we present a modeling strategy to make better use of existing data by using biomedically motivated actuarial models to integrate multiple data sources into a comprehensive model of population health dynamics. The development of a specific model for application to a disease of interest involves three distinct phases. First, biomedical evidence and data are used to specify a cohort model of chronic disease morbidity and mortality. Second, the model is fitted to cohort mortality data with estimates of its parameters being derived by maximum likelihood procedures. Third, the morbidity distribution in the national population is generated from the parameter estimates.*

*The model is used to examine lung cancer morbidity and mortality patterns for U. S. white and non-white males in 1977. A review of these patterns suggests that, based on current concepts of lung cancer incidence and natural history, over 2 percent of white males in the United States have lung cancer at some stage of development, though most of this prevalence is pre-clinical. The likelihood that these clinically latent morbid patterns will translate into future health care needs is a function not only of incidence and natural history of lung cancer in different birth cohorts, but also of changes in the mortality patterns of other diseases. The model demonstrates that, if for other chronic degenerative diseases a large proportion of future health care needs is determined in the present health state of the population, long-range planning models of national population health dynamics are necessary to anticipate and meet future requirements for long-term care health services.*

---

This research was supported by HCFA Contract Number P-97710/4-01 and NIA Grant Number AG-01159-04.

## Introduction

One of the most difficult problems in systematically evaluating the requirements for long-term care health services (diagnostic, therapeutic, and medical preventative) is the lack of a single comprehensive data source representing the long-term care health needs of the U. S. elderly population (CBO, 1977). Given the lack of a single broad data source, assessment of the health care needs of the elderly is typically based on multiple special purpose data sources, each with particular strengths and limitations resulting from their specific focuses. However, it is clear that the integration of multiple special purpose data sets will not completely resolve the analytic issues surrounding the determination of national estimates of need for long-term care health services. This is because the intrinsic nature of chronic degenerative diseases (for example, a lengthy "pre-clinical" period during which direct observation is not usually possible) prohibits the collection of "complete" information on the incidence and prevalence of chronic disease. Consequently, prediction strategies aimed at producing national estimates of chronic disease incidence and prevalence must be able to use information from various biomedical studies on the "natural history" of the chronic disease process in individuals to cover deficiencies in nationally representative data bases. We will employ the term "bioactuarial models" to denote models which predict population level health characteristics using both nationally representative data and insight gained from epidemiological, clinical, and biological studies.

To illustrate the development and application of a bioactuarial model, we have selected a particular example—an evaluation of changes in lung cancer mortality and morbidity for U. S. white and non-white males from 1950 to 1977. This example was chosen because of 1) the availability of the science base necessary to develop a bioactuarial model for this disease, 2) the apparent increase in the incidence and mortality of lung cancer in recent years, 3) the lack of specific attention to national cost estimates for possible long-term management of cancer viewed as a chronic disease, and 4) the example's difficult analytic challenge because of the importance of differential cohort effects, the differential health status of minority groups, and the health implications for the elderly.

In the remainder of this article we will attempt to accomplish two objectives. First, we will examine, in some detail, technical aspects in the development of a bioactuarial modeling strategy. Second, we will apply the bioactuarial model to data and derive the U. S. population morbidity prevalence patterns for lung cancer, that is, the distribution of lung cancer prevalence rates specific to age, race, sex, and severity of disease. Readers interested solely in the nature of the estimates of the disease distribution may wish to skip the section on model specification and estimation.

The organization of the remainder of this paper reflects our two goals. That is, the first section discusses the technical issues in model specification. This section has three parts. The first deals with the development of the mathematical specification of the model. This development begins with a review of the biomedical evidence on the nature of the disease process under study. We then use our conclusions to identify the functional form and parameters of a population model of a cohort illness-death process. The parameters of this model will describe both disease incidence and the forces of mortality. In the second part, we present and discuss the statistical procedures that we use to estimate parameters. In the third part, we show how life tables for the "well" and "morbid" components of the U. S. population may be calculated from the incidence and mortality of the illness-death process. The life table for the morbid population represents the prevalence of the disease.

The second section reviews the results of applying the proposed model to data from a national mortality time series and deriving distributions of lung cancer prevalence. The structure of this section parallels that of the first section, in that we discuss the results of the empirical application in terms of model specification, statistical estimation, and life table calculations of disease prevalence. The findings suggest that the total prevalence is over 2 percent for white males and near 2 percent for non-white males. Most of this prevalence is in a pre-clinical form, due to its existence in higher risk young cohorts. Actually, even these figures are somewhat misleading, since they average prevalence across age groups where there is zero prevalence. For example, 11.5 percent of white males and 12.3 percent of non-white males age 75 to 79 have lung cancer at some stage of development. Of course, depending on the age-specific mortality risks of other causes of death, not all of this disease may become manifest. For example, whereas we observed a little over 60,000 lung cancer deaths in white males in 1977, an equal number of white males with an early stage of lung cancer may have died of another disease. We review the ways a lengthy pre-clinical growth period for the disease, interacting with other population dynamics, lead to this observation. Finally, we summarize some of the general issues surrounding the application of such modeling strategies.

## Methods for Analytically Determining the National Prevalence of Chronic Diseases

In this section, we examine the methodological issues involved in developing a model to produce estimates of disease prevalence specific to age, race, sex, geographic area, disease severity, and duration. The methodological development will involve three distinct phases: model specification, statistical estimation (and inference), and procedures to compute prevalence distributions.

## Specification of a Population Model of Lung Cancer Morbidity and Mortality

The first stage in the methodological development of our strategy for generating prevalence distributions is the specification of a mathematical model of the rates at which the population moves between various health states. There are two basic types of changes in health states that this model represents: 1) acquisition of the disease of interest and 2) dying, either from the disease of interest or from other causes. (See Figure 1.)

In Figure 1, there are four boxes, each representing four different health states. Persons may possess the attributes of being "well" (disease-free), "sick" (in this case, they have a tumor growing), or they may be observed to have died from cancer or some disease other than cancer. The transfer of persons between health states is described by the functions  $\mu$ ,  $\lambda_1$ , and  $\lambda_2$ . These functions indicate how the rate of transfer of persons between compartments changes with age ( $a$ ) or time ( $t$ ) in the disease state. Note that if one had "ideal" data, these three functions could be calculated directly by dividing the age-specific numbers of persons changing from one health state to another during a time interval by the age-specific number of persons in the health state of origin at the start of the time interval. Unfortunately, the available data will not be sufficient to directly calculate the age-specific rates. In particular, we will not know the precise age at which the person entered the "sick" state, that is, the age at which the tumor first started growing. Indeed, it is likely that we observe very little of the time during which the tumor is growing, for, though the "latency" period of lung cancer is estimated to be 25 years or more (Fraumeni, 1975), the median "survival" time from diagnosis to death is only five months (Axtell *et al*, 1976). As a consequence, the transition rate into the tumor growth state ( $\lambda_1$ ) and from the tumor growth state to death either due to cancer ( $\lambda_2$ ) or some other cause ( $\mu$ ) could not be well specified. Better estimates of the health transitions involving the tumor growth state can be obtained if we employ procedures which allow us to introduce auxiliary information about these transitions from epidemiological, clinical, and biological sources. This auxiliary information is built into the model to make the transition rates  $\lambda_1$ ,  $\lambda_2$ , and  $\mu$  explicit mathematical functions of age ( $a$ ) and time ( $t$ ) in the morbid state. We derived the functional forms for these transitions by assessing the appropriate literature. To obtain an estimable form for the model outlined in Figure 1, we made three assumptions about the functional relation of the changes in health state transitions to various time measures. In the following discussion of these three assumptions we shall state the assumption, present it in formal terms, and describe the evidence that led to that assumption.

The first assumption we will make is that risk of death from non-cancer causes is not affected by the presence of a tumor. Formally, this can be expressed as:

$$\mu(a) = \mu(a_0 + t), \text{ if } a = a_0 + t \text{ and } \nu(a) = \mu(a) + \lambda(a)$$

where  $\nu(a)$  is the age-specific total force of mortality and  $\mu(a)$  and  $\lambda(a)$  are the non-cancer and cancer age-specific forces of mortality. Note that  $\lambda(a)$  represents the hazard for the two-step transition in Figure 1 (that is, from the well state through the tumor growth state to the cancer death state). This assumption is equivalent to the assumption of disease independence made in competing risk models (Chiang, 1968) and cause elimination life table calculations. Though the assumption of disease independence is frequently employed and seems useful in practice, a review of the literature about the disease of interest may suggest that disease dependence should be considered. In this case, the procedures discussed in Manton and Stallard (1980) might be applied.

For lung cancer, the independence assumption can be justified on the basis of the relatively lengthy time of tumor growth and the catastrophic nature of the disease. In general, if the tumor had grown to a size sufficient to be a factor in causing a particular death, then that death would be recorded in the vital statistics data as a "cancer death." In contrast, if the tumor had not reached a size sufficient to be a factor in causing a particular death, then that death would be recorded in the vital statistics data as a "non-cancer death." However, if the tumor is not a factor in causing death, then there is no reason to assume that  $\mu(a_0 + t)$  is different from  $\mu(a)$ . Thus, by assuming that  $\mu(a_0 + t)$  and  $\mu(a)$  are equal if  $a = a_0 + t$ , it is possible to estimate the age-specific rate at which persons with a tumor growing for  $t$  years die of non-cancer causes. This is a particularly important aspect of the bioactuarial model because it determines how much of the clinically latent prevalence of a disease eventually becomes manifest.

The second assumption is that the age increase in the rate of tumor onset may be described by the model of human carcinogenesis developed by Armitage and Doll (1954). Under this model, the relation of the cancer incidence rate to age is described by the Weibull function:

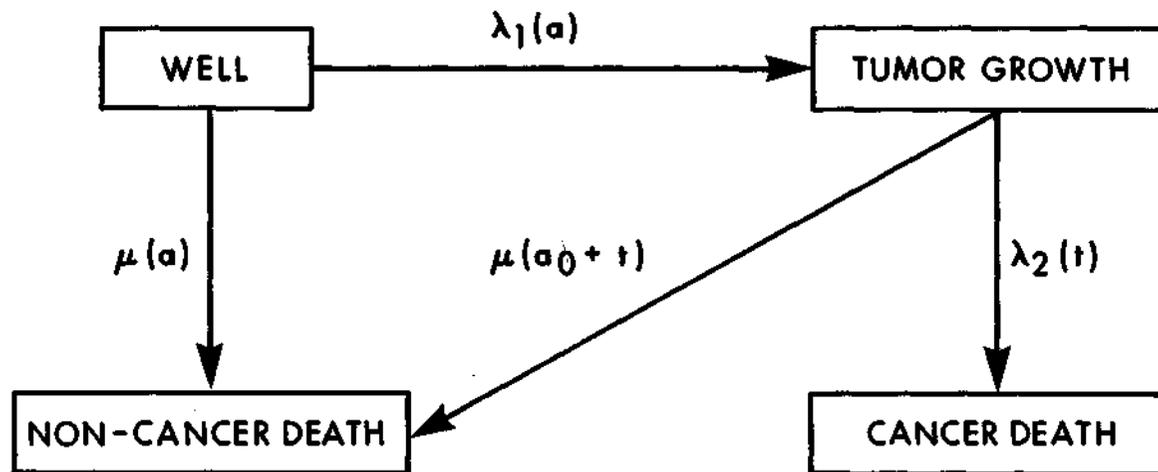
$$\lambda_1(a) = \alpha a^{m-1}$$

where  $m$  is an integer representing the number of non-lethal mutations in a cell nucleus required before that cell loses growth control, and  $\alpha$  represents the product of the probabilities associated with each of the  $m$  mutations.

This relation has three desirable properties. First, the Weibull hazard function leads to a distribution of disease events that is the extreme value distribution Type III (Mann *et al*, 1974). Thus, it permits cancer initiation in an individual to be linked to the failure process in the population of cells which led to the tumor within the individual (Watson, 1977). Second, the relation has a direct biological interpretation at the cellular level. Specifically, it suggests that a tumor starts when  $m$  non-lethal mutations occurred in the nucleus of a single cell. It is assumed that the probability of each mutation is independent of age (Peto *et al*, 1975). Finally, it was found to describe cancer incidence in a broad range of human data (for example, Cook *et al*, 1969).

FIGURE 1

A Stochastic Compartment Model Representation of an Illness-Death Process



Though the simple Weibull model conforms to a broad range of epidemiological data, it has a tendency to overpredict cancer risks at advanced age (above age 75). Since we will be interested in morbidity and mortality at advanced ages, we modified the basic Weibull function in two ways. First, the Weibull function describes the age-specific rate of occurrence of an unobservable event—the initiation of tumor growth from a single cell. This problem is solved by including a parameter,  $l$ , in the Weibull function to represent the time between tumor onset and death from the cancer. This parameter is subtracted from the age at cancer death to provide the theoretical age ( $a - l$ ) at which the cancer began. Second, the function's over-prediction of cancer incidence at advanced ages could be explained as a function of systematic mortality selection, that is, that the persons most susceptible to cancer died first, thereby lowering the average level of risk among survivors. To model the effect of selection we assumed that the standard Weibull function used to predict  $\lambda_1(a)$  applied to individuals and not populations. To model the risks manifest at the population level, it is necessary to determine which of the Weibull parameters are most plausibly allowed to vary over individuals and to specify the nature of the selection process. In examining the Weibull function, the parameters  $m$  (which determines the form of the age increase in cancer risks for individuals) and  $l$  (which represents the time between tumor onset and tumor death) can reasonably be assumed constant over individuals (the assumption of a constant  $l$  will be relaxed later). This means that individual differences will be modeled as differences in  $\alpha$ , that is, that for each individual  $i$  there is associated a value  $\alpha_i$ . The use of  $\alpha$  to model individual differences has the advantage that individual differences can be stated in terms of proportional changes in risks, or relative risks, which are independent of age increases in risk. To estimate a population level model it is necessary to assume the form of the distribution of the  $\alpha$ 's for individuals within the population. We assume that the  $\alpha_i$ 's are gamma distributed because the gamma 1) is extremely flexible and can approximate a number of common distributions, 2) remains gamma under systematic selection, and 3) is closed under sampling from a Poisson distribution. Thus, instead of estimating a single parameter  $\alpha$ , as in the simple (or individual level) Weibull function, we have to estimate two parameters. The first,  $\bar{\alpha}$ , represents the average value of the  $\alpha_i$ 's before mortality selection, that is, for the total cohort population. The second,  $s$ , is the shape parameter of the gamma distribution. This parameter governs the variance of the  $\alpha_i$ 's. Thus, the individual level Weibull may be generalized to predict the age change of the risk of death due to cancer for a cohort, as:

$$\lambda(a) = \bar{\alpha}(a - l)^{m-1} / [1 + \bar{\alpha}(a + l)^m / (ms)].$$

In this expression, we can identify our two modifications of the standard Weibull. First, instead of using the observed age at death,  $a$ , we use  $(a - l)$ , the theoretical time of tumor initiation. (Naturally  $l$  will have to be estimated.) Second, the term in the denominator  $1 + \bar{\alpha}(a - l)^m / (ms)$  represents the change in the distribution of the  $\alpha_i$ 's due to the systematic removal at early ages of persons with high  $\alpha_i$  values. This selection will cause both the mean and variance of the distribution of the  $\alpha_i$ 's to decrease with age.

The third assumption is that the risk of dying from a tumor is proportional to the size of the tumor. In general terms this means

$$\lambda_2(t) \propto g(t)$$

where  $g(t)$  is a function describing the size of the tumor after  $t$  years of growth. In specific terms, this suggests that  $l$ , the average time between tumor initiation and death, can be translated into a distribution of times between tumor initiation and death by selecting a particular function  $g(t)$ . The clinical literature on tumor growth suggests that tumors grow slightly more slowly than exponentially (Archambeau *et al.*, 1970). Hence, the risk of dying from the tumor might be modeled as proportional to any of a number of "subexponential" functions of time spent in the tumor growth state. We assumed that the Weibull function adequately described the rate of tumor growth, or

$$\lambda_2(t) = \beta t^{n-1}$$

To translate  $l$  into a distribution, it is necessary to do two things. First, we must select the parameter  $n$ . This parameter will determine how "peaked" the distribution of the time  $t$  for individuals should be. Thus, the selection of  $n$  will be based upon auxiliary data and theory about individual variability in the rate of tumor growth. With  $n$  specified and the value of  $l$  estimated, it is possible to determine the distribution of  $t$  implied by the Weibull function via the methods described in Manton and Stallard (1982).

We have shown how an estimable model of the illness-death process described in Figure 1 could be developed by making three assumptions. It should be stressed that the model in Figure 1 represents a very simple process and that it is logically possible to expand that model to include more diseases or to define multiple stages to describe a more complex disease history. With that extension, however, the analyst has to pay the price of greater computational complexity. Specifically, as the internal structure of the model is elaborated, there will be a corresponding need for additional auxiliary information to make parameter estimation practical. For example, if one wished to model specific medical interventions for clinically manifest lung cancer, it would be necessary to replace the transition rate function,  $\lambda_2(t)$ , by a more complicated model of disease progression—for example, one involving additional disease stages. Practically speaking, there appear to be few alternative strategies to translate such auxiliary information into national estimates of the specific characteristics of chronic degenerative disease processes.

## Estimation

From the prior discussion we have seen how the transition rates in Figure 1 can be described by functions involving four parameters: 1)  $\bar{\alpha}$ , which represents the average susceptibility to disease onset in a cohort before selection, 2)  $s$ , which determines the variance of individual susceptibility to disease onset, 3)  $m$ , which is the parameter describing the number of mutations required in a single cell before a tumor begins and which determines the curvature of the age trajectory of cancer incidence rates, and 4)  $l$ , which is the average time between tumor initiation and tumor death. In this section, we present statistical procedures to produce numerical estimates of  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$  from the available data and to test the fit of the model to data.

Frequently, cause-specific mortality data are analyzed using a multinomial model to describe the distribution of the cause-specific probabilities at any age. However, since we are only interested in two probabilities—the probabilities of lung cancer death and death due to all other causes—we elected to use a conditional binomial model. That is, since we were interested in adjusting for the competing risk effects of other causes of death on lung cancer, we assumed that the conditional probability of  $y_i$  lung cancer deaths in  $N_i$  total deaths in the  $i$ th age category is binomially distributed. We used this conditional model because the competing risk effects of lung cancer on all other causes of death are negligible and because the total number of deaths could be assumed fixed and estimated directly from total mortality data. The conditional binomial likelihood function is written as:

$$\mathcal{L} = \prod_i \binom{N_i}{y_i} [1 - \lambda(a_i)/\nu(a_i)]^{(N_i - y_i)} [\lambda(a_i)/\nu(a_i)]^{y_i}$$

where  $N_i$  is the observed total number of deaths, and  $y_i$  is the observed number of lung cancer deaths during the age interval  $i$ ;  $a_i$  represents the mid-point of the age interval  $i$ ;  $\lambda(a_i)$  represents the model of the average lung cancer risk among survivors to age  $a_i$ , that is,  $\lambda(a_i) = \bar{\alpha}(a_i - l)^{m-1}/[1 + \bar{\alpha}(a_i - l)^m/(ms)]$ ; and  $\nu(a_i)$ , the observed force of mortality at age  $a_i$ , is assumed known. Maximum likelihood estimates of  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$  may be derived by taking the natural logarithm of the binomial likelihood function and producing the first order partial derivatives of it with respect to  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$ . Manton and Stallard (1979) provide the derivations of the first and second order partial derivatives. With the first and second order partial derivatives, it is possible to produce maximum likelihood estimates of the parameters by determining the values of the parameters that maximize the likelihood function. These values are obtained via Newton-Raphson procedures also described in Manton and Stallard. Test statistics for evaluating pairs of hierarchical models are derived from the standard approximation that minus twice the difference in the log likelihood functions associated with each model is approximately distributed as a  $\chi^2$  variable with degrees of freedom equal to the difference in the number of parameters between the pair of models.

## Generation of Morbidity Distribution

The third phase of the analysis involves taking the maximum likelihood estimates of  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$  for selected cohorts and generating morbidity distributions for the national population. To produce these distributions, we need to define two life table functions in terms of the transition rate functions identified in Figure 1:

- $l_{W(x+1)} = l_{Wx} \cdot \exp\left\{-\int_x^{x+1} [\mu(a) + \lambda_1(a)] da\right\}$
- $l_{Tx, t+1} = l_{Tx, t} \cdot \exp\left\{-\int_x^{x+1} [\mu(a) + \lambda_2(a - x + t + 1/2)] da\right\}$

The life table function  $l_{W(x+1)}$  represents the proportion of the initial life table population alive at age  $x+1$  who do not have tumors. The conditions of being alive and not possessing a tumor represent the definition of persons in the "well" health state. Clearly, the probability of remaining in the well state to age  $x+1$  is a product of the probability of being in the well state at age  $x$  and surviving both the force of transition to the death from other cause state ( $\mu(a)$ ) and the force of transition to the tumor growth state ( $\lambda_1(a) = \bar{\alpha} a^{m-1}/[1 + \bar{\alpha} a^m/(ms)]$ ). Thus, this life table function can be calculated directly from the cohort-specific maximum likelihood estimates of the three parameters  $\bar{\alpha}$ ,  $m$ , and  $s$  and the observed  $\mu(a)$ . The second life table function represents the proportion of the initial life table population who survive to age  $x+1$  with a tumor growing for between  $t+1$  and  $t+2$  years. Again, we see that this life table function may be determined from the product of the probability of being alive at age  $x$  with a tumor growing for between  $t$  and  $t+1$  years times the probability of surviving the force of transition to either death from other causes ( $\mu(a)$ ) or from cancer ( $\lambda_2(a - x + t + 1/2)$ ).

In effect, these two life table functions represent the survival probabilities for a "two-dimensional" life table, where  $l_{Wx}$  represents survival of the primary decrements (that is, acquisition of a tumor or death due to other causes without a tumor growing), and  $l_{Tx, t}$  represents the probability of surviving the second decrement (dying from the tumor or dying from other causes but with a tumor growing).

In fact, the calculation of the  $l_{Tx, t}$  involves an important factor not present in the calculation of  $l_{Wx}$ . Specifically, while  $l_{W0}$  is a known quantity (the initial population size),  $l_{Tx, 0}$  is not known. Thus we must devise a way of estimating the  $l_{Tx, 0}$ . This quantity can be estimated from

$$l_{Tx, 0} = l_{Wx} \cdot \exp\left(-\int_x^{x+1} \mu(a) da\right) - l_{W(x+1)}$$

Simply, the proportion that develops a tumor in the age interval  $x$  to  $x + 1$  but survives both the non-cancer and cancer forces of mortality is equal to the difference of 1) the product of the probability of remaining in the well state to age  $x$  with the probability of surviving the non-cancer force of mortality over the age interval  $x$  to  $x + 1$ ,  $\exp(-\int_x^{x+1} \mu(a) da)$ , and 2) the probability of remaining in the well state to age  $x + 1$ . The derivation of this formula is provided in Manton and Stallard (1982). With these life table quantities, the morbidity distributions may be generated.

## Results

In this section we describe the application to lung cancer of the analytic strategy reviewed in the prior section. This illustration will be presented according to the three phases identified previously.

### Phase One: Model Specification

In the case of lung cancer we were able to fully specify a model based on the clinical and epidemiological literature. With other diseases, the specification of the model itself may involve one or both of two additional analyses. The first type of analysis that would be used as input involves a review and re-specification of model elements by substantive experts. The second type of analysis would involve empirical analysis of auxiliary data sources to determine either 1) parametric forms for  $\lambda_1$ ,  $\lambda_2$ , and  $\mu$  or 2) the derivation of external estimates of parameters of the functions  $\lambda_1$ ,  $\lambda_2$ , and  $\mu$ . For example,  $\lambda_2$  might be partly determined by medical follow-up studies of the mortality risks of persons who already possessed the disease.

### Phase Two: Parameter Estimation

To produce good estimates of the transition rates, it is necessary to possess appropriate data. First, to generate morbidity distributions over the policy relevant variables, all of these variables must be represented in the data. Second, the data must be representative of the national population. Third, since cross-temporal variation provides much of the information to estimate parameters, a lengthy time series must be available. This will also permit us to assess cohort-specific changes in health states. Finally, on practical grounds, the data base should contain information on a broad range of diseases so that common analytic and data management procedures can be applied to a variety of different health problems.

The national, cause-specific mortality data produced by the National Center for Health Statistics constitute one data source which fulfills these criteria. From these files we were able to obtain individual mortality records with age, race, sex, geographic region, and underlying cause of death for all persons who died in the United States of some form of cancer between 1950 and 1977. From these individual records we were able to compile race and sex-specific frequencies of lung cancer mortality by single years of age for nine cohorts, age 30, 35, 40, 45, 50, 55, 60, 65, and 70 in 1950, for each year from 1950 to 1977.

We obtained total mortality figures for these nine cohorts either from vital statistics publications (1950 to 1961) or from individual mortality records for deaths from all causes (1962 to 1977). Cohort-specific lung cancer and total mortality death rates were generated by pairing the mortality frequencies with the appropriate population figures obtained from interpolating between the 1950, 1960, and 1970 censuses and from special census estimates prepared for each year from 1970 to 1977. We adjusted the population data for enumeration error using estimates provided by Siegel (1974), Coale and Rives (1973), and Coale and Zelnick (1963).

In Table 1 we present maximum likelihood estimates of  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$  for lung cancer mortality for the nine cohorts for white and non-white males in the United States.

**TABLE 1**  
Stochastic Compartment Model Parameter for U.S. Male Cohort Lung Cancer Mortality 1950-1977

Cohort Age $x$ in 1950	$\bar{\alpha}$	White Males		
		$s$	$m$	$l$
30	$2.894 \times 10^{-11}$	$3.395 \times 10^{-2}$	6.0 <sup>1</sup>	20.3 <sup>1</sup>
35	$2.381 \times 10^{-11}$	$8.218 \times 10^{-2}$		
40	$2.121 \times 10^{-11}$	$8.621 \times 10^{-2}$		
45	$1.746 \times 10^{-11}$	$1.160 \times 10^{-1}$		
50	$1.473 \times 10^{-11}$	$1.046 \times 10^{-1}$		
55	$1.332 \times 10^{-11}$	$6.906 \times 10^{-2}$		
60	$1.054 \times 10^{-11}$	$5.262 \times 10^{-2}$		
65	$7.261 \times 10^{-12}$	$3.738 \times 10^{-2}$		
70	$5.320 \times 10^{-12}$	$2.907 \times 10^{-2}$		
		Non-White Males		
30	$4.135 \times 10^{-11}$	$8.368 \times 10^{-2}$	6.0 <sup>1</sup>	19.4 <sup>1</sup>
35	$2.858 \times 10^{-11}$	$8.982 \times 10^{-2}$		
40	$2.106 \times 10^{-11}$	$1.086 \times 10^{-1}$		
45	$1.513 \times 10^{-11}$	$1.746 \times 10^{-1}$		
50	$1.241 \times 10^{-11}$	$1.382 \times 10^{-1}$		
55	$1.004 \times 10^{-11}$	$5.769 \times 10^{-2}$		
60	$6.637 \times 10^{-12}$	$4.482 \times 10^{-2}$		
65	$4.639 \times 10^{-12}$	$2.993 \times 10^{-2}$		
70	$2.155 \times 10^{-12}$	$5.347 \times 10^{-2}$		

<sup>1</sup> parameter assumed equal for all nine cohorts

Table 1 shows that our estimate of  $m$  was assumed to apply to all cohorts and both male populations. We imposed this constraint on the estimate of  $m$  because of the argument that the parameter  $m$  is a characteristic of the tissue type in which the tumor arose. This constraint on the estimate of  $m$  also seems consistent with the findings of Cook *et al* (1969) that  $m$  is characteristic of specific tumor types even across national populations. We constrained our estimate of  $l$  to vary only by race for two reasons. First, there seems to have been little change in the median survival time for lung cancer. Second, we believed that observed differences in  $l$  might be related to racial differences in diagnosis and treatment. Two parameters,  $\bar{\alpha}$  and  $s$ , are allowed to vary over all cohorts,

indicating different levels and distributions of lung cancer risk for cohorts. The fact that  $\bar{\alpha}$  decreases systematically with cohort age (in 1950) for both races indicates the higher risk of younger cohorts. The ratio of the  $\bar{\alpha}$  for the two male groups age 30 in 1950 indicates that the risk for whites is only 70 percent of that for non-whites. For the cohort 70 years old in 1950, however, the white male risk is 2.47 that of non-white males, suggesting that there has been a substantial change in the relative risks of successive cohorts in the two racial groups. The parameter  $s$  peaks for the cohort age 45 in 1950 for both groups. The change in this parameter implies that the age at which the peak mortality risks occur changes over cohorts. As  $s$  increases, the age of peak mortality risks increases. The initial variance of the distribution of the values of  $\alpha$  for individuals (as opposed to the initial mean value of  $\alpha$  in Table 1) can be calculated from the two parameters by the relation  $\text{var}(\alpha) = \bar{\alpha}^2/s$ . As a consequence, the heterogeneity in lung cancer risk increases for younger cohorts in both groups because  $\bar{\alpha}$  is increasing and  $s$  is decreasing. For cohorts older than age 45 in 1950, the decrease in  $s$  balances the decrease in  $\bar{\alpha}$  to a degree. Manton *et al* (1982) provide a more detailed discussion of the mortality analyses.

By using these parameter estimates and fixing the parameter  $n$  in the function  $\lambda_2(t)$  to determine the translation of  $l$  into a distribution of times in the tumor growth state, we can calculate the life table parameters  $l_{wx}$  and  $l_{Tx,t}$ . These two life table parameters can then be multiplied by the appropriate population value to produce the lung cancer morbidity distribution in the U. S. population. Note that the parameter estimates determine the rate of incidence of the disease. To generate the disease prevalence distribution, the incidence parameters have to be applied over time. Consequently, the incidence parameters should be estimated from a mortality time series longer than the longest time that any individual is likely to spend in the chronic disease state. Failure to have a mortality time series of adequate length is a serious problem, as illustrated in Figure 2.

The lung cancer mortality patterns for the nine cohorts are indicated in Figure 2 by the nine continuous lines. The lung cancer mortality patterns for the periods 1970 and 1977 are indicated in Figure 2 by the two sets of dotted lines which connect the mortality rates for each of the nine cohorts during the relevant period. One can see that the age trajectory of lung cancer mortality risks represented in period data, which reflects the experience of a mixture of the nine cohorts, is different than the age trajectory of lung cancer mortality risks for any given cohort. Consequently, period estimates of the parameters  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$  will not describe the incidence rates for any cohort. In addition to producing good incidence estimates for a cohort, a long time series is necessary to represent the total history of incidence changes which produced the present prevalence.

In the case of lung cancer, with an  $l$  not greater than 20.3 years and a value of  $n$  set at 9, no persons were predicted to survive over 29 years with a tumor, and only 0.52 percent of those who died of the disease survived with the disease from 25 to 29 years. As a consequence, our 28-year mortality time series is adequate to generate the prevalence distribution, specific to the time with the disease, for the year 1977. Under other situations such as 1) a significant proportion of those with the chronic disease surviving more than 28 years or 2) one wishing to generate the prevalence distribution for an earlier date (effectively shortening the time series), the calculation will have to be based on further assumptions, that is, assumptions about changes in incidence before the start of the mortality time series.

In Figure 3, the fit of the lung cancer death rates produced from the parameter estimates in Table 1 for each of nine cohorts (pluses) to the observed death rates for those nine cohorts (circles) can be examined.

In particular, note that there are no indications of systematic deviations of the predicted death rates from the observed data. This suggests that the model faithfully reproduces the lung cancer death rates of these nine cohorts. We derived the morbidity hazards for the cohorts not explicitly in the mortality analyses from linear interpolations between the adjoining cohorts or, for cohorts younger than age 30 in 1950, by assuming that the age 30 cohort hazards applied.

### Phase Three: Generation of National Morbidity Distribution

The parameters  $\bar{\alpha}$ ,  $s$ ,  $m$ , and  $l$ , estimated from the death rates of the nine cohorts and the parameter  $n$ , can be used to derive the life table functions  $l_{wx}$  and  $l_{Tx,t}$ . These two life table functions are applied to appropriate population estimates to produce the lung cancer morbidity and mortality conditions for white males in 1977 as presented in Table 2.

The nine columns in Table 2 describe both the mortality conditions as observed (columns 4 and 9) and the morbidity and mortality conditions as inferred from the cohort time series data. For example, column 2 gives the number of persons alive and free from lung cancer for five-year age groups. For ages 75 to 79, we see that 1,380,782 of 1,560,768, or 88.5 percent of white males are free of lung cancer. Column 3 contains the number of persons in the tumor growth state for each age. For example, at ages 75 to 79 this number is 179,985 or 11.5 percent of the total. The high percentage of white males with the disease reflects the long time between tumor initiation and tumor death. As long as white male life expectancy is not increased dramatically, many of the

**FIGURE 2**  
**Period and Cohort Lung Cancer Mortality Patterns**  
**for U.S. White Males, 1950-1977**

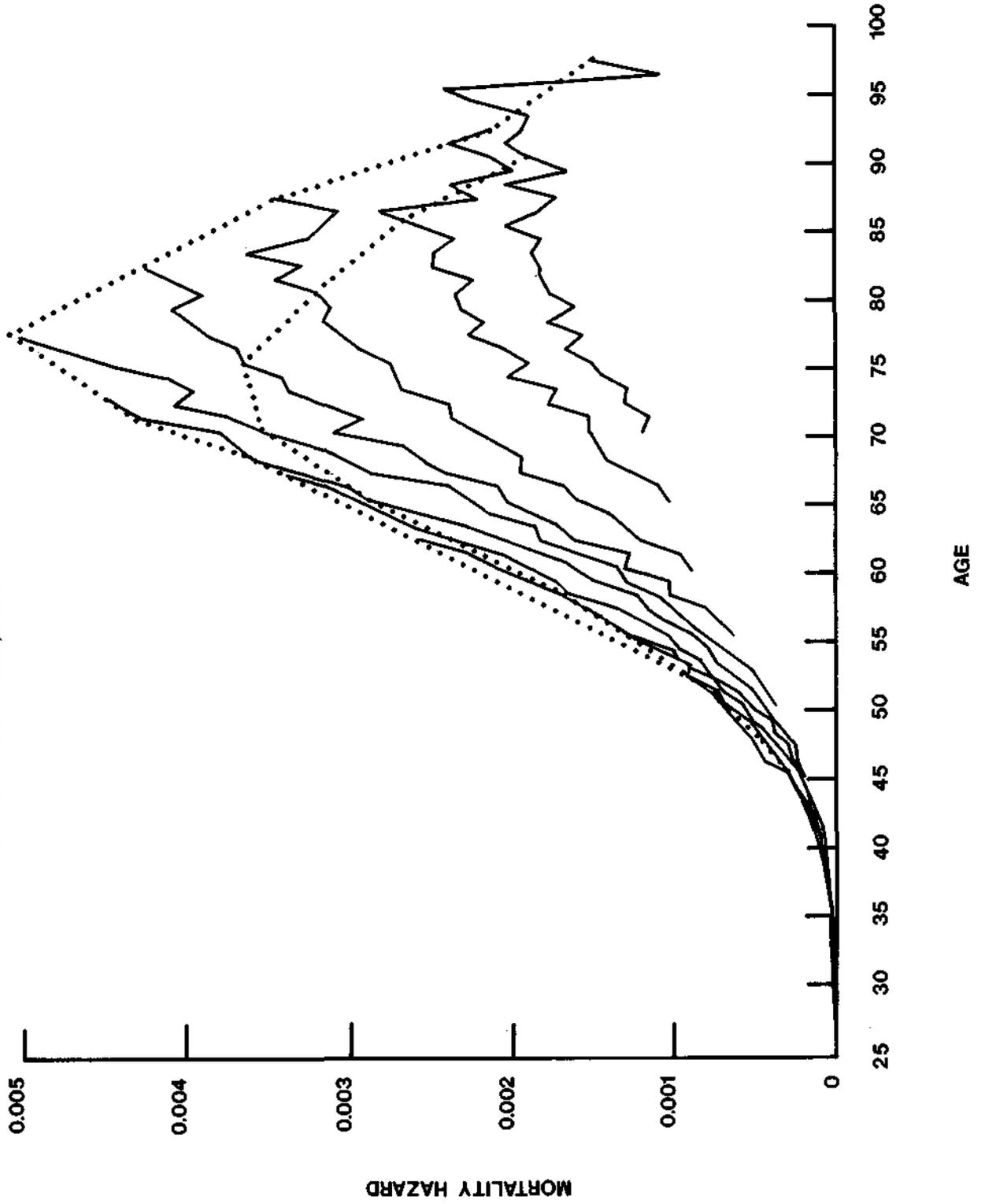
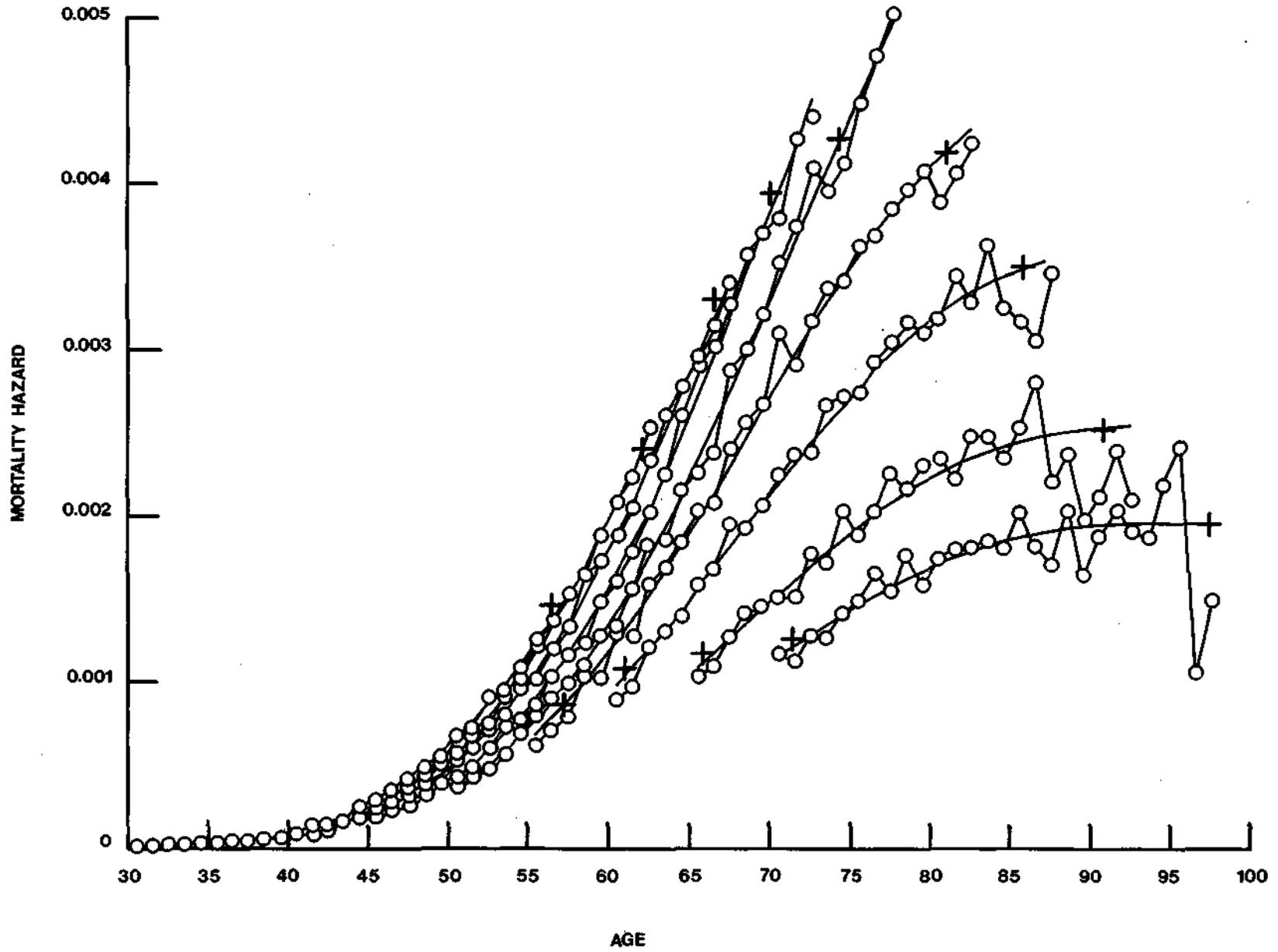


FIGURE 3

Predicted and Observed Cohort Lung Cancer Mortality for U.S. White Males, 1950-1977



**TABLE 2**  
**Observed and Predicted Lung Cancer Morbidity and Mortality Conditions for White Males in 1977**

Age	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Observed Population	Disease Free Population	Population in Tumor Growth State	Observed Dead all Causes	Dead from all Causes Predicted from Model	Dead from Other Causes and No Cancer	Dead from Other Causes with Cancer	Dead from Lung Cancer Predicted from Model	Observed Dead from Lung Cancer
0	6578500.0	6578500.0	0.0	22770.0	22770.0	22770.0	0.0	0.0	0.0
5	7472793.0	7472786.0	7.0	2798.0	2796.0	2796.0	0.0	0.0	2.0
10	8295061.0	8294910.0	150.0	3487.0	3486.0	3486.0	0.0	0.0	1.0
15	9193130.0	9191949.0	1180.0	13209.0	13203.0	13201.0	2.0	0.0	6.0
20	8787712.0	8782730.0	4982.0	16255.0	16249.0	16239.0	10.0	1.0	7.0
25	8043217.0	8028156.0	15061.0	12804.0	12786.0	12745.0	23.0	17.0	35.0
30	6994122.0	6960567.0	33554.0	11005.0	10992.0	10846.0	53.0	93.0	106.0
35	5514716.0	5457568.0	57148.0	11583.0	11558.0	11123.0	118.0	317.0	342.0
40	4969989.0	4876481.0	93508.0	16314.0	16343.0	15149.0	295.0	899.0	870.0
45	5164363.0	5013973.0	150389.0	28093.0	28034.0	24983.0	757.0	2294.0	2353.0
50	5234510.0	5027697.0	206813.0	47325.0	47214.0	40876.0	1677.0	4660.0	4771.0
55	4904591.0	4654921.0	249670.0	68530.0	68596.0	58139.0	3102.0	7354.0	7285.0
60	4118206.0	3783425.0	334781.0	93031.0	92952.0	76212.0	6725.0	10016.0	10089.0
65	3387797.0	3055655.0	332143.0	114844.0	114659.0	93326.0	10008.0	11325.0	11503.0
70	2420557.0	2138067.0	282491.0	123782.0	123724.0	100361.0	13041.0	10322.0	10366.0
75	1560768.0	1380782.0	179985.0	117355.0	117336.0	97653.0	12403.0	7280.0	7283.0
80	979411.0	889364.0	90047.0	104030.0	104148.0	91173.0	8914.0	4061.0	3923.0
85	471559.0	438084.0	33475.0	68160.0	68261.0	62107.0	4579.0	1575.0	1458.0
90	148381.0	140702.0	7679.0	29199.0	29226.0	27422.0	1443.0	361.0	328.0
95	22252.0	21361.0	890.0	5617.0	5613.0	5355.0	217.0	41.0	45.0
Total	94261634.0	92187679.0	2073955.0	910191.0	909948.0	785963.0	63369.0	60616.0	60773.0

Note: Summation reflects rounding error.

people in the morbid state will not express the disease clinically—due to the censoring effect of other causes of death. An examination of the total population, over all ages, suggests that 2.2 percent of the white male population is in the tumor growth state. One measure of how well the model predicts mortality is to compare columns 4 and 5, which are the observed and predicted total deaths, respectively. The model reproduces this figure quite well, underpredicting by less than 0.03 percent. Columns 6 and 7 are derived from the model and show, of the number predicted to die from a non-lung cancer cause of death, the numbers who die with and without a tumor growing. At ages 75 to 79, the model predicts that 11.3 percent of the white males reported dying from another cause had lung cancer at some stage of development. Actually this figure can be broken down by stage of tumor growth, as in Table 5 where it is shown how long these persons had the lung cancer before dying of another cause. The final two columns of Table 2 show how well lung cancer deaths are predicted from the model. The total over all ages shows that the model underpredicts by 0.26 percent. At ages 75 to 79 it is off by only three deaths.

Although Table 2 gives a good overall assessment of the morbidity and mortality characteristics of the white male population in 1977, it does not represent one crucial dimension—the severity of the disease process. This is

reflected under the disease model by the length of time spent in the tumor growth state. To some degree this is, practically, a definitional property of what we identify as chronic degenerative disease processes. That is, a chronic degenerative disease is a progressive disorder with little chance of recovery and with increasing disability and/or mortality risks. It is also assumed that with increasing severity there is an increasing need for diagnostic, therapeutic, and palliative health services. The mapping of the distribution of the mixture of requirements for various types of health services onto the severity distribution requires both auxiliary data on utilization and expert clinical input.

To determine the progression in the severity of the disease, it is necessary to have a model of the “natural history” of the disease process. This model can either be empirically based or derived from biomedical theory. In the case of cancer, we are fortunate to have a biological model of the progression of the disease based on a considerable range of clinical evidence. The core of this model is the concept of the “doubling time” of the growth process (that is, the amount of time it takes for the cells in a tumor to double; Archambeau *et al*, 1970) and the functional description of the doubling time process. In Table 3, we show the hazards of dying from lung cancer specific to the amount of time that the individual had the tumor.

**TABLE 3**  
**Time-Specific Hazard of Lung Cancer Death for White Males Given the Presence of a Tumor**

Time of Tumor Growth (in years)	Hazard
1	$4.03 \times 10^{-11}$
2	$3.96 \times 10^{-9}$
3	$7.87 \times 10^{-8}$
4	$7.11 \times 10^{-7}$
5	$4.04 \times 10^{-6}$
6	$1.69 \times 10^{-5}$
7	$5.71 \times 10^{-5}$
8	$1.64 \times 10^{-4}$
9	$4.18 \times 10^{-4}$
10	$9.66 \times 10^{-4}$
11	$2.06 \times 10^{-3}$
12	$4.13 \times 10^{-3}$
13	$7.81 \times 10^{-3}$
14	$1.41 \times 10^{-2}$
15	$2.45 \times 10^{-2}$
16	$4.09 \times 10^{-2}$
17	$6.64 \times 10^{-2}$
18	$1.05 \times 10^{-1}$
19	$1.61 \times 10^{-1}$
20	$2.43 \times 10^{-1}$
21	$3.59 \times 10^{-1}$
22	$5.21 \times 10^{-1}$
23	$7.43 \times 10^{-1}$
24	1.04

The mortality hazards rise very rapidly after about 20 years. We generated these hazards by assuming that the rate of progression of a tumor could be well described by a Weibull function with an exponent of 9, that is, that tumor progression was less rapid than exponential. These hazards can be used as an actuarial index of the mortality risk after having the disease for a given period of time and, consequently, as an index of the severity of the disease. Obviously, different types of health services are required at different levels of severity. For example, given that a tumor is highly lethal after 40 doublings and clinically detectable after 30 doublings, we could project, assuming a constant rate of doublings, that 40 doublings have occurred, on average, by the time  $t = 20.3$  and that 30 doublings have occurred, on average, by about the fifteenth year of tumor growth. Consequently, individuals who have had their tumors for 14 to 15 years would need diagnostic services and therapeutic services (such as surgery and acute care hospitalization). Those with tumors for about 19 years would presumably have clinically manifested the disease and would require therapeutic services and, given the generally poor prognosis of lung cancer, palliative services (for example, radiation and chemotherapy). Those with the tumor about 20 years would probably need acute hospitalization services, palliative therapy, and, perhaps, hospice services.

In Table 4 we provide the age-specific distributions of the time spent in the tumor growth state for white males alive and with a tumor growing.

As indicated earlier, we can make certain assumptions about the typical requirements for health services for a given amount of time spent in the tumor growth state. For example, 82.4 percent of the group surveyed had lung cancer for less than 15 years. Given the selected model of the rate of progression of the tumor, this group is probably not in immediate need of health services because before 15 years, the tumor is not detectable. The 14.4 percent who had the tumor 15 to 19 years were potentially detectable and would require diagnostic as well as therapeutic health services. The 3.19 percent with the tumor 20 years or more compose the group for which the tumor would almost certainly be clinically manifest and for which therapeutic health services would be mandated. Given the present prognosis for people with advanced lung cancer, this group would also be likely to need full disability, acute hospitalization services, and possibly hospice services.

In Table 5, we present the age- and time-specific distribution of white males who died from something other than lung cancer but who had a tumor growing.

One can see that a slightly larger number of persons died of other causes than died of the lung cancer (63,369 versus 60,616). As mentioned, this is due to the lengthy time of progression of the disease and its occurrence at relatively advanced ages when other causes of death have high risks. This censoring by other causes is manifest primarily for white males with up to 19 years of tumor growth. For 20 or more years of tumor growth, the risk of lung cancer death is dominant.

In Table 6 we present the age- and time-specific distribution of white males who died of lung cancer. The table shows that males who died of lung cancer had their disease for from 10 to 24 years. This distribution is clearly a function of the assumption of  $n = 9$ , indicating a rapidly increasing hazard function.

In Table 7 we present the lung cancer morbidity and mortality conditions for non-white males in 1977 generated from the parameter estimates provided in Table 1.

The model for non-white males also does a good job of reproducing total mortality and lung cancer mortality, over-predicting by 0.18 percent and 3.1 percent, respectively. The non-white population is a more difficult population to fit because of problems in enumerating population and in reporting age at mortality. An examination of the populations dying free of lung cancer, having lung cancer but dying of another cause, and dying of lung cancer itself shows that white and non-white males have broadly similar mortality experiences. For example, about as many non-whites die of another cause, but with lung cancer growing, as those who die of lung cancer directly—the same as for whites. Also, of those who died, 87.5 percent of non-white males were free of lung cancer, compared to 86.4 percent of white males. These similarities tend to hide more significant age-specific differences due to the different cohort risks.

**TABLE 4**

**Distribution of White Males Alive in 1977 with a Tumor Growing, by Age and the Length of Time in the Tumor Growth State**

Age	Total	%	Years in Tumor Growth State					
			0-4	5-9	10-14	15-19	20-24	25-29
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
5	7.0	0.00	7.0	0.0	0.0	0.0	0.0	0.0
10	150.0	0.01	142.0	8.0	0.0	0.0	0.0	0.0
15	1180.0	0.06	1020.0	152.0	9.0	0.0	0.0	0.0
20	4982.0	0.24	3875.0	958.0	141.0	7.0	0.0	0.0
25	15061.0	0.73	10490.0	3573.0	879.0	117.0	3.0	0.0
30	33554.0	1.62	20860.0	8978.0	3023.0	654.0	39.0	0.0
35	57148.0	2.76	31253.0	16551.0	7068.0	2084.0	192.0	0.0
40	93508.0	4.51	44057.0	28342.0	14900.0	5538.0	670.0	2.0
45	150389.0	7.25	60076.0	45860.0	29221.0	13265.0	1961.0	7.0
50	206813.0	9.97	69973.0	61049.0	46114.0	25215.0	4443.0	18.0
55	249670.0	12.04	75142.0	70302.0	59240.0	37336.0	7613.0	37.0
60	334781.0	16.14	101664.0	93832.0	78529.0	50109.0	10592.0	54.0
65	332143.0	16.01	94352.0	90431.0	79976.0	54812.0	12502.0	70.0
70	282491.0	13.62	77477.0	75489.0	68645.0	49050.0	11760.0	70.0
75	179985.0	8.68	46489.0	46656.0	44396.0	33736.0	8653.0	56.0
80	90047.0	4.34	21745.0	22476.0	22443.0	18274.0	5073.0	36.0
85	33475.0	1.61	7835.0	8188.0	8353.0	7041.0	2043.0	15.0
90	7679.0	.37	1778.0	1862.0	1913.0	1637.0	485.0	4.0
95	890.0	.04	205.0	215.0	222.0	191.0	57.0	0.0
Total	2073955.0		668440.0	574922.0	465071.0	299067.0	66086.0	369.0
%			32.23	27.72	22.42	14.42	3.19	.02

**TABLE 5**

**Distribution of White Males with a Lung Cancer in 1977, Who Died of a Non-Lung Cancer Cause, by Age and Length of Time in Tumor Growth State**

Age	Total	%	Years in Tumor Growth State					
			0-4	5-9	10-14	15-19	20-24	25-29
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
15	2.0	0.00	2.0	0.0	0.0	0.0	0.0	0.0
20	10.0	0.02	7.0	2.0	0.0	0.0	0.0	0.0
25	23.0	0.04	17.0	5.0	2.0	0.0	0.0	0.0
30	53.0	0.08	33.0	14.0	2.0	1.0	0.0	0.0
35	118.0	0.17	65.0	35.0	15.0	4.0	1.0	0.0
40	295.0	0.47	139.0	90.0	47.0	17.0	2.0	0.0
45	757.0	1.19	302.0	233.0	149.0	65.0	8.0	0.0
50	1677.0	2.65	571.0	500.0	378.0	199.0	29.0	0.0
55	3102.0	4.90	949.0	887.0	744.0	448.0	76.0	0.0
60	6725.0	10.61	2072.0	1914.0	1596.0	972.0	170.0	0.0
65	10008.0	15.79	2892.0	2773.0	2443.0	1598.0	300.0	1.0
70	13041.0	20.58	3644.0	3552.0	3217.0	2193.0	432.0	1.0
75	12403.0	19.57	3270.0	3284.0	3113.0	2257.0	476.0	2.0
80	8914.0	14.07	2204.0	2279.0	2267.0	1761.0	402.0	2.0
85	4579.0	7.23	1100.0	1150.0	1167.0	938.0	224.0	1.0
90	1443.0	2.28	343.0	359.0	368.0	299.0	74.0	0.0
95	217.0	0.34	51.0	54.0	55.0	45.0	11.0	0.0
Total	63369.0		17661.0	17131.0	15568.0	10800.0	2203.0	7.0
%			27.87	27.03	24.57	17.04	3.48	0.01

**TABLE 6**

**Distribution of White Males Who Died in 1977 of Lung Cancer, by Age and Length of Time in Tumor Growth State**

Age	Total	%	Years in Tumor Growth State					
			0-4	5-9	10-14	15-19	20-24	25-29
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
20	1.0	0.00	0.0	0.0	1.0	0.0	0.0	0.0
25	17.0	0.02	0.0	1.0	6.0	9.0	1.0	0.0
30	93.0	0.15	0.0	2.0	24.0	53.0	14.0	0.0
35	317.0	0.52	0.0	4.0	59.0	181.0	72.0	0.0
40	899.0	1.48	0.0	8.0	131.0	502.0	257.0	2.0
45	2294.0	3.78	0.0	13.0	268.0	1244.0	763.0	6.0
50	4660.0	7.69	0.0	19.0	440.0	2437.0	1749.0	16.0
55	7354.0	12.13	0.0	22.0	580.0	3696.0	3023.0	32.0
60	10016.0	16.52	0.0	29.0	767.0	4965.0	4208.0	47.0
65	11325.0	18.68	0.0	28.0	788.0	5475.0	4974.0	60.0
70	10322.0	17.03	0.0	24.0	676.0	4900.0	4664.0	60.0
75	7280.0	12.01	0.0	15.0	437.0	3369.0	3413.0	47.0
80	4061.0	6.68	0.0	7.0	220.0	1819.0	1985.0	30.0
85	1575.0	2.60	0.0	2.0	81.0	691.0	787.0	13.0
90	361.0	0.60	0.0	1.0	18.0	157.0	182.0	3.0
95	41.0	0.07	0.0	0.0	2.0	18.0	21.0	0.0
Total	60616.0		1.0	175.0	4496.0	29514.0	26114.0	316.0
%			0.0	0.29	7.42	48.69	43.08	.52

**TABLE 7**

**Observed and Predicted Lung Cancer Morbidity and Mortality Conditions for Non-White Males in 1977**

Age	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Observed Population	Disease Free Population	Population in Tumor Growth State	Observed Dead all Causes	Dead from all Causes Predicted from Model	Dead from Other Causes and No Cancer	Dead from Other Causes with Cancer	Dead from Lung Cancer Predicted from Model	Observed Dead from Lung Cancer
0	1502632.0	1502632.0	0.0	8825.0	8823.0	8823.0	0.0	0.0	2.0
5	1585129.0	1585127.0	2.0	758.0	756.0	756.0	0.0	0.0	2.0
10	1647978.0	1647936.0	42.0	857.0	856.0	857.0	0.0	0.0	0.0
15	1686303.0	1685998.0	305.0	2364.0	2364.0	2363.0	1.0	0.0	0.0
20	1505621.0	1504423.0	1198.0	3792.0	3787.0	3783.0	4.0	0.0	6.0
25	1289888.0	1286469.0	3419.0	4133.0	4132.0	4116.0	11.0	5.0	6.0
30	1026277.0	1019205.0	7072.0	3636.0	3634.0	3584.0	26.0	25.0	27.0
35	858360.0	845577.0	12783.0	3948.0	3963.0	3818.0	58.0	86.0	71.0
40	769523.0	748508.0	21015.0	5107.0	5104.0	4734.0	133.0	236.0	239.0
45	737011.0	705813.0	31199.0	7001.0	6995.0	6185.0	274.0	536.0	541.0
50	668880.0	630315.0	38565.0	9713.0	9694.0	8239.0	501.0	954.0	973.0
55	575201.0	534092.0	41109.0	11785.0	11840.0	9773.0	743.0	1324.0	1268.0
60	439284.0	398771.0	40513.0	13474.0	13479.0	11019.0	1108.0	1352.0	1346.0
65	377684.0	337046.0	40638.0	15029.0	15054.0	12181.0	1446.0	1427.0	1400.0
70	238552.0	207528.0	31024.0	14375.0	14412.0	11611.0	1707.0	1095.0	1054.0
75	146510.0	128463.0	18047.0	12197.0	12267.0	10157.0	1400.0	710.0	636.0
80	95445.0	87985.0	7459.0	8357.0	8415.0	7456.0	615.0	344.0	283.0
85	48392.0	45705.0	2687.0	4909.0	4916.0	4528.0	257.0	131.0	123.0
90	18165.0	17430.0	735.0	2179.0	2183.0	2062.0	84.0	36.0	32.0
95	4148.0	3952.0	197.0	562.0	564.0	529.0	26.0	9.0	7.0
Total	15220982.0	14922976.0	298006.0	133001.0	133238.0	116574.0	8394.0	8270.0	8016.0

A comparison of Tables 7 and 2 shows a number of differences between white and non-white lung cancer mortality—differences not only due to the cohort differences in mortality risks but also due to the relative youth of the non-white male population distribution compared to the white male population distribution.

For example, while 4.2 percent of non-white males ages 45 to 49 had tumors and were alive (column 3/column 1), only 2.9 percent of white males had a tumor growing at this age. This is a function of the higher lung cancer incidence for the younger non-white cohorts. This contrasts with 7.1 percent of the white male population ages 85 to 89 alive with a tumor growing versus the 5.6 percent of the non-white male population. Overall, the white male population was estimated to have a slightly higher prevalence of lung cancer (2.20 versus 1.95 percent) due to the different age structure of the two populations. With the aging of both populations and the higher risks of the young, non-white cohorts we could project that, in the future, non-white males will have a higher prevalence of lung cancer.

In Table 8 we provide the distribution of non-white males alive with a tumor growing, stratified by age and the length

of time they had the tumor. In this table, we see the effects of the younger population age structure and the high incidence rates for younger cohorts, since 86.2 percent of the non-whites with a tumor had the tumor for less than 14 years, while only 82.4 percent of the whites did. At younger ages, the differences between the groups are less (91.1 percent at ages 45 to 49 for non-whites; 89.9 percent for whites) than at older ages. Extrapolation of the higher cohort rates for the younger non-whites into future years suggests the greater rate of increase of lung cancer as a health hazard for non-whites.

In Table 9 we present the distribution of deaths from non-lung cancer causes for non-white males with a lung cancer growing. A comparison of Tables 9 and 5 shows that there are differences between the white and non-white distributions of deaths among those who had lung cancer but died from another cause. Specifically, the proportion of non-whites in this group who had a tumor growing for less than 10 years is 5.1 percent higher than for whites. This is due to the high early mortality rates from other causes for non-white males, as evidenced by the smaller proportions of non-whites with lung cancer dying of other causes after age 70.

**TABLE 8**  
**Distribution of Non-White Males Alive in 1977 with a Tumor Growing, by Age and Time in the Tumor Growth State**

Age	Total	%	Years in Tumor Growth State					
			0-4	5-9	10-14	15-19	20-24	25-29
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
5	2.0	0.00	2.0	0.0	0.0	0.0	0.0	0.0
10	42.0	0.01	40.0	2.0	0.0	0.0	0.0	0.0
15	305.0	0.10	263.0	39.0	2.0	0.0	0.0	0.0
20	1198.0	0.40	933.0	229.0	33.0	2.0	0.0	0.0
25	3419.0	1.15	2387.0	809.0	197.0	25.0	0.0	0.0
30	7072.0	2.37	4414.0	1891.0	633.0	129.0	5.0	0.0
35	12783.0	4.29	7056.0	3702.0	1565.0	432.0	27.0	0.0
40	21015.0	7.05	10051.0	6397.0	3322.0	1150.0	94.0	0.0
45	31199.0	10.87	12762.0	9625.0	6031.0	2530.0	251.0	0.0
50	38565.0	12.94	13449.0	11631.0	8637.0	4340.0	508.0	0.0
55	41109.0	13.79	12679.0	11897.0	9960.0	5791.0	782.0	1.0
60	40513.0	13.59	12451.0	11613.0	9779.0	5843.0	827.0	1.0
65	40638.0	13.64	12264.0	11509.0	9867.0	6097.0	901.0	1.0
70	31024.0	10.41	9387.0	8750.0	7496.0	4683.0	708.0	1.0
75	18047.0	6.06	5111.0	4960.0	4492.0	3000.0	484.0	0.0
80	7459.0	2.50	1930.0	1964.0	1912.0	1404.0	250.0	0.0
85	2687.0	0.90	673.0	694.0	693.0	528.0	98.0	0.0
90	735.0	0.25	183.0	189.0	189.0	146.0	28.0	0.0
95	197.0	0.07	51.0	52.0	50.0	37.0	7.0	0.0
Total	298006.0		106087.0	85952.0	64359.0	36135.0	4971.0	4.0
%			35.60	28.84	21.76	12.13	1.67	0.0

**TABLE 9**  
**Distribution of Non-White Males with a Lung Cancer in 1977, Who Died of a Non-Lung Cancer Cause, by Age and Length of Time in a Tumor Growth State**

Age	Total	%	Years in Tumor Growth State						
			0-4	5-9	10-14	15-19	20-24	25-29	
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	1.0	0.01	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	4.0	0.05	2.0	1.0	0.0	0.0	0.0	0.0	0.0
25	11.0	0.13	8.0	3.0	1.0	0.0	0.0	0.0	0.0
30	26.0	0.31	16.0	6.0	3.0	0.0	0.0	0.0	0.0
35	58.0	0.69	32.0	17.0	7.0	2.0	0.0	0.0	0.0
40	133.0	1.58	64.0	41.0	21.0	7.0	0.0	0.0	0.0
45	274.0	3.26	113.0	86.0	54.0	22.0	2.0	0.0	0.0
50	501.0	5.97	176.0	153.0	133.0	54.0	5.0	0.0	0.0
55	743.0	8.85	233.0	219.0	182.0	99.0	11.0	0.0	0.0
60	1108.0	13.20	347.0	323.0	270.0	151.0	17.0	0.0	0.0
65	1446.0	17.23	445.0	417.0	355.0	205.0	24.0	0.0	0.0
70	1707.0	20.34	526.0	491.0	417.0	244.0	28.0	0.0	0.0
75	1400.0	16.68	405.0	393.0	353.0	221.0	28.0	0.0	0.0
80	615.0	7.33	163.0	166.0	160.0	110.0	16.0	0.0	0.0
85	257.0	3.06	66.0	69.0	68.0	48.0	7.0	0.0	0.0
90	84.0	1.00	22.0	22.0	22.0	16.0	3.0	0.0	0.0
95	26.0	0.31	7.0	7.0	6.0	5.0	1.0	0.0	0.0
Total	8394.0		2625.0	2414.0	2032.0	1183.0	140.0	0.0	0.0
%			31.27	28.76	24.21	14.09	1.67	0.0	0.0

In Table 10 we present the age- and time-specific distributions of non-white males dying of lung cancer. A comparison of Table 10 with Table 6 shows that a higher proportion of lung cancer deaths occur at older ages for whites than non-whites. The proportion of white males dying with a tumor growing from 20 to 24 years is 13.7 percent higher than for non-white males.

### Discussion

We selected cancer to illustrate the modeling strategy for two reasons. First, there is an extensive theoretical and empirical base available from which the age incidence function and the rate of disease progression function could be specified. An alternate strategy is to derive those functions not from theoretical concerns but from specific studies. For example, renal dysfunction and associated morbidity and mortality have been examined in several medical follow-up studies (for example, Singer and Levinson, 1976). The findings from those studies could be used to empirically specify a function for incidence or for the progression of disease severity.

Second, cancer represents a major area of future need for long-term care services which has not been fully explored. Specifically, even the most lethal forms of adult cancers are chronic diseases which require therapeutic services. Cancer, unlike circulatory diseases, is exhibiting increasing incidence and mortality risks. The facts that cancer affects the elderly, that early treatment improves prognosis, and that much of the early disease progression remains undetected suggest the need for diagnostic services among the elderly. Finally, there has been increasing discussion of the utility of high technology medical care for such terminal diseases as cancer and alternative strategies for managing the terminal patient, such as hospice care. All these reasons suggest that, in the future, cancer will be a disease requiring considerable long-term care expenditures.

In this context, the results of the analysis, though intended to be illustrative, demonstrate one very important fact about chronic disease prevalence—that, frequently, large proportions of the population have the disease, though much of the prevalence may be clinically latent. Additionally, though our lung cancer prevalence estimates of 2.2 percent and 1.9 percent of white and non-white males may appear high, it is quite possible that these underestimate the "true" prevalence of this disease. To see this, it is first necessary to realize that the prevalence estimates will vary approximately in direct proportion to the

**TABLE 10**  
**Distribution of Non-White Males Who Died in 1977 of Lung Cancer by Age and Length of Time in Tumor Growth State**

Age	Total	%	Years in Tumor Growth State					
			0-4	5-9	10-14	15-19	20-24	25-29
0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0
25	5.0	0.06	0.0	0.0	2.0	3.0	0.0	0.0
30	25.0	0.30	0.0	1.0	7.0	15.0	3.0	0.0
35	86.0	1.04	0.0	1.0	19.0	52.0	13.0	0.0
40	236.0	2.85	0.0	3.0	43.0	145.0	46.0	0.0
45	536.0	6.48	0.0	4.0	81.0	328.0	122.0	0.0
50	954.0	1.15	0.0	5.0	121.0	579.0	249.0	0.0
55	1324.0	16.01	0.0	5.0	143.0	791.0	384.0	1.0
60	1352.0	16.34	0.0	5.0	140.0	800.0	406.0	1.0
65	1427.0	17.26	0.0	5.0	142.0	837.0	442.0	1.0
70	1095.0	13.24	0.0	4.0	107.0	639.0	345.0	1.0
75	710.0	8.59	0.0	2.0	64.0	409.0	234.0	0.0
80	344.0	4.16	0.0	1.0	28.0	194.0	121.0	0.0
85	131.0	1.58	0.0	0.0	10.0	73.0	47.0	0.0
90	36.0	0.44	0.0	0.0	3.0	20.0	13.0	0.0
95	9.0	0.11	0.0	0.0	1.0	5.0	3.0	0.0
Total	8270.0		0.0	38.0	909.0	4890.0	2429.0	4.0
%			0.0	0.46	10.99	59.13	29.37	0.05

value of the parameter  $l$  which represents the average time between tumor onset and death due to the tumor. The best available epidemiological estimates of the range of  $l$  suggest that our estimates of  $l = 20.3$  and  $19.4$  years for white and non-white males may be low (Fraumeni, 1975). Thus our prevalence estimates are likely to also be conservatively low. Second, although the Weibull function with  $n = 9$  appears to imply a satisfactory lower bound of about 10 years on the shortest latency period for both white and non-white males, the upper bound of about 25 years on the longest latency time is probably too short when compared with the Fraumeni (1975) estimates of up to 40 years or more. Our preliminary efforts in this area suggest that this truncation to about 25 years could result in our prevalence estimates being about 30 percent too low. Third, we estimated the incidence parameters which determine the prevalence levels from lung cancer mortality data. As a consequence, our prevalence estimates will not include a component representing those persons who are effectively cured of lung cancer through appropriate medical management of the disease. From Axtell *et al* (1976), we find that with about a 5 percent recovery rate our prevalence estimates are about 5 percent too low.

Although it is clear that further effort is required to improve the precision of these lung cancer prevalence estimates, it is equally clear, with a lengthy clinically latent period and a relatively short clinically manifest period (20 years versus five months), that the major portion of the lung cancer prevalence is clinically latent. Such an imbalance between the clinically latent and clinically manifest portions of chronic disease prevalence will have important implications for long-term care health service policies. This is due to the need for direct medical intervention only for the clinically manifest cases of the disease. The fact that actual prevalence may be much larger than estimates of diagnosed disease is due to 1) the lengthy development time of chronic diseases, 2) the fact that competing causes of death censor much disease prevalence before it is clinically manifest, and 3) differences in cohort incidence rates so that younger cohorts may have higher incidence rates but lower prevalence because of age differences. This suggests that long-range planning for health care needs will have to consider the dynamics of population morbidity and mortality. For example, these dynamics suggest that one consequence of a continued decrease in the mortality risks of circulatory diseases will be an increase in cancer prevalence. This would occur because the level of censoring of cancer prevalence due to non-cancer causes

of death would be reduced as the mortality risks of circulatory diseases (which represent the predominant portion of non-cancer risks) are reduced. With an increase in overall prevalence of cancer, there is the likelihood that an increasing proportion of this prevalence will represent the clinically manifest stage of the disease. Thus, the ability to accurately forecast changes in this component of the population will be of particular relevance to long-term care health service providers.

Here we can usefully distinguish two types of forecasts. First, for short-run forecasts whose term is less than the average latency time, the clinically manifest portions of the prevalence will be almost completely determined by the size of the current clinically latent prevalence. For lung cancer, with  $l$  about 20 years, the term of these types of forecasts would run to about the year 2000. Second, for long-run forecasts whose term is longer than the average latency time, the incidence model will become the primary determinant of estimates of future clinically manifest prevalence.

In general terms, our bioactuarial modeling strategy has implications for evaluations of national health policy because it produces precise, quantitative estimates of morbidity and mortality specified by a measure of disease severity (time in tumor growth state) for the national population. These results permit very detailed and specific policy assessments, for they present quantitative statements in a form appropriate to evaluate the need for health services of various types. This precision and detail should not be taken to substitute for statistical confidence. That is, a detailed quantitative assessment can be generated with underlying data and theory of varying degrees of reliability and completeness. It is clear, however, that the results produced by the bioactuarial model are based on the best available biomedical evidence and theory and fit to extensive health survey data. Thus, the results must be viewed with greater confidence than simple simulation results based on models that are not biologically motivated, subjective probability models which rely only on judgmental input, or detailed results from studies of very limited populations.

The bioactuarial modeling strategy offers a second important advantage—it can be reviewed in a variety of ways. One way the model can be verified is to see if it adequately reproduces the data to which it is fit. Second, experts can review the model components to determine if they represent the best available biomedical data and theory. Third, we can examine the projections from the model structure. For example, the morbidity distributions are not directly observable but they can be compared for reasonableness to available incidence studies and to clinical studies of disease progression. Ultimately, if a component of a model is not directly reviewable (that is, adequate auxiliary evidence is not available to determine the reasonableness of the model or of the results), then one can attempt to determine the sensitivity of the outcome of the model to variations in those parameters. If the model results are sensitive to the parameters, one has at least identified an area of need for further research which will have significant implications for assessing population requirements for health care.

## References

- Archambeau, J.O., M.B. Heller, A. Akanuma, and D. Lubell, "Biologic and Clinical Implications Obtained from the Analysis of Cancer Growth Curves," *Clinical Obstetrics and Gynecology* 13: 831-856, 1970.
- Armitage, P. and R. Doll, "The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis," *British Journal of Cancer* 8: 1-12, 1954.
- Axtell, L.M., A.J. Asire, and M.H. Myers, *Cancer Patient Survival, Report No. Five*, DHEW Pub. No. (NIH) 77-992, Bethesda, Maryland, 1976.
- Chiang, C.L., *Introduction to Stochastic Processes in Biostatistics*, New York, Wiley, 1968.
- Coale, A.J. and N.W. Rives, "A Statistical Reconstruction of the Black Population of the United States, 1880-1970," *Population Index* 39: 3-36, 1973.
- Coale, A.J. and M. Zelnik, *New Estimates of Fertility and Population in the United States*, Princeton, N.J., Princeton University Press, 1963.
- Congressional Budget Office, Congress of the United States, *Long-Term Care: Actuarial Cost Estimates: CBO Technical Analysis Paper*, Washington, D.C., U.S. Government Printing Office, 1977.
- Cook, P.J., R. Doll, and S.A. Fellingham, "A Mathematical Model for the Age Distribution of Cancer in Man," *International Journal of Cancer* 4: 93-112, 1969.
- Fraumeni, J.F. (ed.), *Persons at High Risk of Cancer: An Approach to Cancer Etiology and Control*, New York, Academic Press, Inc., 1975.
- Mann, N.R., R.E. Schafer, and N.D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data*, New York, Wiley, 1974.
- Manton, K.G. and E. Stallard, "Maximum Likelihood Estimation of a Stochastic Compartment Model of Cancer Latency: Lung Cancer Mortality Among White Females in the U.S.," *Computers and Biomedical Research* 12: 313-325, 1979.
- Manton, K.G. and E. Stallard, "A Stochastic Compartment Model Representation of Chronic Disease Dependence: Techniques for Evaluating Parameters of Partially Unobserved Age Inhomogeneous Stochastic Processes," *Theoretical Population Biology* 18: 57-75, 1980.
- Manton, K.G. and E. Stallard, "The Use of Mortality Time Series Data to Produce Hypothetical Morbidity Distributions and Project Mortality Trends," *Demography*, 1982, in press.
- Manton, K.G., E. Stallard, and W. Figgan, "Strategies for Analyzing Ecological Health Data: Models of the Biological Risk of Individuals," *Statistics in Medicine*, 1982, forthcoming.
- Peto, R., F.J.C. Roe, P.N. Lee, L. Levy, and J. Clack, "Cancer and Aging in Mice and Men," *British Journal of Cancer* 32: 411-425, 1975.
- Siegel, J.S., "Estimates of Coverage of the Population by Sex, Race, and Age in the 1970 Census," *Demography* 11: 1-23, 1974.
- Singer, R.B., and L. Levinson, *Medical Risks: Patterns of Mortality and Survival*, Lexington, Mass., Lexington Books and D.C. Heath, 1976.
- Watson, G., "Age Incidence Curves for Cancer," *Proceedings of the National Academy of Sciences* 74: 1341-1342, 1977.