TRANSACTIONS OF SOCIETY OF ACTUARIES
1973 VOL. 25 PT. 1 NO. 72 AB

CAUSE-OF-DEATH LIFE TABLES: APPLICATION OF A NEW TECHNIQUE TO WORLDWIDE DATA

SAMUEL H. PRESTON,* NATHAN KEYFITZ,†
AND ROBERT SCHOEN

ABSTRACT

In an increasing population the number of persons falls off from age to age more rapidly than does the stationary population of the life table. This is true within age groups as well as between them and can introduce a systematic bias into life table calculations. The first part of the paper describes an iterative procedure to correct for the bias, extends the method to multiple decrement and associated single decrement life tables, and demonstrates the magnitude of the bias involved.

The second part of the paper summarizes the results of applying iterative life table techniques to 165 populations which encompass over a century's experience and include countries from all parts of the world. The relationship between the probability of ultimately dying from each of twelve different causes and the expectation of life at birth is explored. Quantitative dimensions are given to the rise in the likelihood of eventually dying from one of the degenerative diseases as longevity increases from 25 to 75 years, and to the corresponding decline in the probability of succumbing to an infectious and parasitic disease.

The gain in longevity from the elimination of mortality from different causes is examined. In particular, it is found that when mortality from all infectious and parasitic sources is eliminated, all male populations considered have expectations of life at birth between 69 and 72.5 years and all female populations between 72 and 76 years. This result suggests that infectious diseases bear almost exclusive responsibility for producing life expectancies below contemporary Western levels, and testifies to the almost complete lack of progress against noninfectious causes of death.

This research was partially supported by the National Science Foundation through Grants NSF GS33864 and GZ995, by the National Institutes of Health through contract 69-2200, and by the National Institute of General Medical Sciences and the Ford Foundation through Grants 5 T01 and GM01240 to the Department of Demography, University of California.

* Dr. Preston, not a member of the Society, is associate professor and director, Center for Studies in Demography and Ecology, Department of Sociology, University of Washington, Seattle, Washington.

† Dr. Keyfitz, not a member of the Society, is professor, Department of Sociology Harvard University.

SINCE the first crude model was advanced by John Graunt in 1662 [3], the life table has been a focus of actuarial study. There is no exact answer to the question of how to construct a life table, and many methods, both simple and complex, have been used. In the present article we describe briefly a recently developed method of life table construction and show how it can be extended to multiple and associated single decrement life tables. More faithful to the data than many techniques of abridged life table construction presently used, the method partially substitutes a dependence on the speed and precision of modern electronic computers for a reliance on assumption, approximation, or graduation. Our major purpose, however, is to summarize the results of applying our technique to data from 165 national populations and to present a broad-based description of the contribution of various causes of death to variations in life table parameters. The reader may proceed directly to the section on "Applications" (p. 91) without any loss of continuity.

THE BASIC LIFE TABLE

The essential problem in the construction of a life table is how to transform observed rates of mortality into probabilities of survival. To express the problem in mathematical terms, we define the following:

$D_x dx$ = Number of deaths occurring in a specified year among those in the infinitesimally small age interval x to $x + dx$;

$P_x dx$ = Number of person-years lived in a specified year in the infinitesimally small age interval x to $x + dx$;

μ_x = Death rate in the infinitesimally small age interval from x to $x + dx$ (or the force of mortality at exact age x) in a specified year, being identically equal to the ratio $D_x dx / P_x dx$; and

l_x = Number who survive to exact age x in a life table with radix l_0, being identically equal to

$$l_0\left[\exp\left(-\int_0^x \mu_t dt\right)\right].$$

In a noninfinitesimal age interval of size n, the observed population death rate will then be

$$ {}_nM_x = \frac{{}_nD_x}{{}_nP_x} = \frac{\int_0^n P_{x+t}\mu_{x+t}dt}{\int_0^n P_{x+t}dt}. \qquad (1)$$

In the life table the death rate in the age interval x to $x + n$ will be

$$ {}_nm_x = \frac{\int_0^n l_{x+t}\mu_{x+t}dt}{\int_0^n l_{x+t}dt}, \tag{2} $$

since the age composition within the interval, represented by the l_{x+t}'s, is stationary, completely determined by the force of mortality. The death rate we observe is ${}_nM_x$, which is based on a population whose age composition is generally not stationary but has been affected by migration and natural increase. When the population is growing, the typical situation in the world today, the observed population will be younger in each interval than the life table population. Consequently, the population death rate ${}_nM_x$ will be less than the true life table death rate ${}_nm_x$ in age intervals where the force of mortality is rising with age. For rapidly growing populations, ${}_5m_x$ can exceed ${}_5M_x$ by up to 1 per cent. Nonetheless, the almost universal assumption (e.g., in Jordan [5]) has been that P_{x+t} is proportional to l_{x+t} and hence that

$$ {}_nM_x = \frac{\int_0^n Cl_{x+t}\mu_{x+t}dt}{\int_0^n Cl_{x+t}dt} = {}_nm_x . \tag{3} $$

The iterative method of life table construction does not require the assumption that the observed population is stationary. Instead, it allows the more relaxed condition where the age distribution within each age interval is stable, that is, where each annual cohort exceeds the size of the previous cohort by the constant factor e^r. As the method has been described in detail elsewhere (Keyfitz [6, 7]; Preston, Keyfitz, and Schoen [10]), the argument will be abbreviated here. We assume that P_{x+t} is proportional to $e^{-r_x t}l_{x+t}$, where r_x (written r in the future to simplify notation) is the imputed stable rate of growth within the age interval. Equation (1) is then rewritten as

$$ {}_nM_x = \frac{\int_0^n e^{-rt}l_{x+t}\mu_{x+t}dt}{\int_0^n e^{-rt}l_{x+t}dt}, \tag{4} $$

or, after integration by parts,

$$ {}_nM_x = \frac{l_x - e^{-nr}l_{x+n}}{{}_nL'_x} - r, \tag{5} $$

where we write ${}_nL'_x$ for

$$\int_0^n e^{-rt} l_{x+t} dt .$$

Thus ${}_nM_x = {}_nm_x$ when $r = 0$.

It is now necessary to find an appropriate expression for ${}_nL'_x$ in terms of l_x and r values, or to relate the life table death rate ${}_nm_x$ to the probability of survival ${}_nP_x$. In working with five-year intervals, we have chosen to approximate ${}_5L'_x$ by means of a cubic polynomial fitted through adjacent l_x values, the exact relationship being

$${}_5L'_x = \tfrac{65}{24}(l_x + e^{-5r} l_{x+5}) - \tfrac{5}{24}(e^{-10r} l_{x+10} + e^{5r} l_{x-5}) . \qquad (6)$$

While equation (5) is an exact equation for the basic life table under the stable population assumption, equation (6) is a convenient and accurate approximation for the number of person-years lived. The reader should bear in mind that two completely separate assumptions are involved and that it is the former that distinguishes the iterative method from all earlier ones.

To compute the set of l_x values, begin with an arbitrary set of l_x values and set all the r_x's equal to zero. Use equation (6) to calculate an improved set of ${}_nL'_x$ values, and use them in equation (5) to calculate new l_x's. Values for r_x can then be found from

$$r_x = \tfrac{1}{10} \ln \left(\frac{{}_5P_{x-5} / {}_5L'_{x-5}}{{}_5P_{x+5} / {}_5L'_{x+5}} \right), \qquad (7)$$

which supposes stability over a fifteen-year period. The next iterative cycle begins when the improved l_x and r values are introduced into equation (6), and iteration continues until convergence on the l_x's is achieved. A simplified but efficient FORTRAN computer program that will carry out all the necessary calculations is given in Preston, Keyfitz, and Schoen [10]. The radix is taken as 100,000, and the initial l_x values are obtained from

$$l_{x+5} = l_x \left(\frac{1 - 2.5\, {}_5M_x}{1 + 2.5\, {}_5M_x} \right). \qquad (8)$$

Alternative methods are used for the beginning and end of the table. Convergence to within 0.001 on all l_x values is achieved in roughly fifteen iterations.

THE MULTIPLE DECREMENT LIFE TABLE

The general multiple decrement case can be reduced to the case where a person can leave the life table only through one of two possible causes,

designated (i) and $(-i)$. Maintaining the stable population assumption, the observed population death rate from cause (i) is

$$ {}_nM_x^{(i)} = \frac{{}_nD_x^{(i)}}{{}_nP_x} = \frac{\int_0^n e^{-rt} l_{x+t} \mu_{x+t}^{(i)} dt}{\int_0^n e^{-rt} l_{x+t} dt} , \tag{9} $$

where the force of mortality from cause (i) is defined in the usual manner as

$$ \mu_x^{(i)} = -\frac{1}{l_x} \frac{dl_x^{(i)}}{dx} \tag{10} $$

and $l_x^{(i)}$ represents the number living at age x who will later leave the life table due to cause (i). As the total force of mortality is equal to the sum of $\mu^{(i)}$ and $\mu^{(-i)}$, l_x is equal to the sum of $l_x^{(i)}$ and $l_x^{(-i)}$.

The integral expression in equation (9) can be simplified by integration by parts and the result rearranged to

$$ l_x^{(i)} = e^{-5r} l_{x+5}^{(i)} + {}_5M_x^{(i)} \, {}_5L_x' + r \, {}_5L_x'^{(i)} , \tag{11} $$

where

$$ {}_5L_x'^{(i)} = \int_0^5 e^{-rt} l_{x+t}^{(i)} dt . $$

Equation (11), which is analogous to equation (5), leads the way to the iterative solution for the multiple decrement life table. The previously calculated r and L' values are available from the basic life table, and $L'^{(i)}$ follows from a cubic through the $l^{(i)}$'s.

Once the basic life table has been completed, only two to four iterations are required to produce an $l_x^{(i)}$ column. The procedure begins with the assumption that the proportion of life table deaths due to cause (i) in an age interval is the same as $M^{(i)}$ divided by M, and iterates to the point where life table deaths from cause (i) are in the proportion that $m^{(i)}$ is of m.

THE ASSOCIATED SINGLE DECREMENT TABLE

The associated single decrement table, like the multiple decrement table, can be treated generally through the case in which only two causes of decrement are involved. The essential difference between the two types of tables is that, while the multiple decrement table depicts survivorship when both cause (i) and cause $(-i)$ are acting on the population, the associated single decrement table (ASDT) attempts to predict what survivorship would be if only cause (i) were operating. The multiple decrement table should reflect reality as closely as possible, but the

ASDT requires not only accurate measurement but also important assumptions.

The key assumption underlying the calculation of an ASDT is that the absence of cause $(-i)$ in no way affects the vitality of the population, that is, that $\mu^{(i)}$ is unchanged. What will happen to $\mu^{(i)}$ in an actual situation cannot be predicted in advance. Two "natural experiments" suggest that $\mu^{(-i)}$ and $\mu^{(i)}$ may be positively correlated. Death rates from all causes rose in the United States during the 1918–19 influenza pandemic (Grove and Hetzel [4]), while death rates from many causes fell dramatically as a result of Ceylon's successful antimalaria campaign of 1945–47 (Newman [9]). On the other hand, Lew [8] says that, at least for individual sufferers, "medical intervention usually tends to give greater scope to other components of mortality." A decrease in one cause would then be associated with increases in others, and $\mu^{(-i)}$ and $\mu^{(i)}$ would therefore be negatively correlated. Lew suggests that a zero correlation is another possibility. For example, the decline in mortality from industrial accidents may have neither substantially increased nor decreased mortality from other causes. The assumption that the disappearance of $\mu^{(-i)}$ will not affect $\mu^{(i)}$ is strong and open to challenge, but it is probably the most reasonable until more knowledge of the interdependence of causes of death becomes available.

Let us designate the number of survivors to age x, in the ASDT in which only cause (i) is acting, as $\bar{l}_x^{(i)}$, or simply $\bar{l}_x$. If both the basic life table and the ASDT begin from a common radix, the $\bar{l}_x$ values will always exceed the l_x values, since the latter are decreased by a larger force of mortality. We assume that, within an age interval, the relationship between l_x and $\bar{l}_x$ can be expressed as

$$l_x = e^{-hx}\bar{l}_x . \tag{12}$$

The constant h depends upon the excluded cause $(-i)$. With a common unit radix

$$l_x = \bar{l}_x^{(i)}\bar{l}_x^{(-i)} = \bar{l}_x^{(i)}\left[\exp\left(\int_0^x \mu_t^{(-i)}dt\right)\right] \tag{13}$$

(see Jordan [5], chap. 14). Thus h can be written as

$$h = \frac{1}{x}\int_0^x \mu_x^{(-i)}dt , \tag{14}$$

or as the mean value of $\mu^{(-i)}$ in the interval. The constant h, or, more precisely, ${}_nh_x$, is free to vary from interval to interval. As with the analogous constant r, the subscripts of h will usually be omitted.

If we begin the calculation of an ASDT with multiple decrement table values, we know that our base population is in fact stationary. It follows that

$$ {}_nm_x^{(i)} = \frac{\int_0^n l_{x+t}\mu_{x+t}^{(i)}dt}{\int_0^n l_{x+t}dt} = \frac{\int_0^n e^{-ht}\bar{l}_{x+t}\mu_{x+t}^{(i)}dt}{\int_0^n e^{-ht}\bar{l}_{x+t}dt}. \tag{15} $$

Since $\mu^{(i)}$ is the total force of mortality in the ASDT, equation (15) is identical in form with equation (4) for the basic life table and can be solved for the $\bar{l}_x$ values in the same manner. Only the procedure for improving values of h needs to be changed, and that is easily done. Applying equation (12) to ages x and $x + n$ yields

$$ {}_nh_x = \frac{1}{n}\ln\left(\frac{\bar{l}_{x+n}/\bar{l}_x}{l_{x+n}/l_x}\right), \tag{16} $$

and the iterative circle is complete.

In constructing ASDT's, a number of writers have supposed that death rates from the several causes are constant during each age interval, or that the several causes act in a fixed ratio to one another (e.g., Chiang [1]). While those assumptions may be perfectly acceptable for some causes of death, it introduces distortions when such causes as accidents, tuberculosis, or maternal mortality are considered. The procedures described here instead assume that the death rate from the cause of interest may vary during the interval but that the death rate from the excluded cause is constant. Put another way, the iterative method treats the exclusion of cause $(-i)$ as if it produced a deformation in the age composition within an age interval, changing the previously stationary population into a stable population with growth rate h. Iteration begins by calculating h from l_x values computed by the Chiang method and converges in two or three cycles to within 0.001 for all the $\bar{l}_x$'s.

EVALUATING THE ITERATIVE METHOD

The values in an ASDT are much more sensitive to the assumptions regarding disease interdependencies than to the choice of calculation technique. The Chiang method yields values very close to those of the iterative method in almost all instances. If in fact $\mu^{(i)}$ and $\mu^{(-i)}$ are varying in an interval but remain in a fixed ratio to one another, the Chiang method is more appropriate. On the other hand, if the cause of death excluded is in fact constant during the interval, the iterative technique is preferable.

In contrast, the iterative method of multiple decrement life table construction has an important advantage over other techniques for abridged life table construction. Because the iterative method provides a way to adjust for nonstationarity in the age composition of the observed population, a condition whose effect on death rates varies considerably with cause of death, it produces a life table that more accurately describes the level of mortality in the observed population.

The quantitative difference between ${}_nM_x^{(i)}$ and ${}_nm_x^{(i)}$ can be investigated by means of equation (9). Substituting $1 - rt$ for e^{-rt} in both the numerator and denominator of equation (9), we have

$$ {}_nM_x^{(i)} = \frac{{}_nd_x^{(i)} - r\int_0^n tl_{x+t}\mu_{x+t}^{(i)}dt}{{}_nL_x - r\int_0^n tl_{x+t}dt} . \tag{17} $$

Let us now define ${}_na_x^{(i)}$, the average number of years lived past age x by those who die from cause (i) between the ages of x and $x + n$, as

$$ {}_na_x^{(i)} = \frac{\int_0^n tl_{x+t}\mu_{x+t}^{(i)}dt}{\int_0^n l_{x+t}\mu_{x+t}^{(i)}dt} , \tag{18} $$

and ${}_nA_x$, the average age past x of persons in the life table population between x and $x + n$, as

$$ {}_nA_x = \frac{\int_0^n tl_{x+t}dt}{\int_0^n l_{x+t}dt} . \tag{19} $$

Equation (17) can then be written as

$$ {}_nM_x^{(i)} \doteq \frac{{}_nd_x^{(i)} - r\,{}_nd_x^{(i)}\,{}_na_x^{(i)}}{{}_nL_x - r\,{}_nL_x\,{}_nA_x} \doteq \frac{{}_nd_x^{(i)}}{{}_nL_x}\,\frac{1 - r\,{}_na_x^{(i)}}{1 - r\,{}_nA_x} . \tag{20} $$

Rearranging terms and again making use of the partial expansion for e^{-rx}, we have

$$ \frac{{}_nm_x^{(i)}}{{}_nM_x^{(i)}} \doteq 1 + r({}_na_x^{(i)} - {}_nA_x) . \tag{21} $$

Therefore, the error in assuming that $r = 0$ is approximately equal to r times the difference between ${}_na_x^{(i)}$ and ${}_nA_x$. Since that difference is positive

if death rates from (i) are rising monotonically in the interval, ${}_nm_x^{(i)}$ will exceed ${}_nM_x^{(i)}$ in age intervals where the population is growing and mortality rising, and the error will vary directly with r.

Examples of errors in death rates introduced by assuming stationarity in the observed population can readily be shown. Table 1, based on data that we presented in *Causes of Death: Life Tables for National Populations* [10], shows the ratio of ${}_nm_x^{(i)}$ to ${}_nM_x^{(i)}$ for a scattering of ages, causes, times, and places. In all cases shown, the error in the death rate can be seen to be of the order of some 1–7 parts per thousand, with the observed

TABLE 1

EXAMPLES OF THE ERROR IN THE DEATH RATE INTRODUCED BY ASSUMING STATIONARITY IN THE OBSERVED POPULATION

Population	Age	Cause	Growth Rate (r)	Ratio of Iterated to Observed Death Rates ${}_nm_x^{(i)}/{}_nM_x^{(i)}$
United States, females, 1964	75–79	All	0.0307	1.0066
England and Wales, males, 1861	70–74	Cardiovascular	0.0092	1.0020
France, females, 1951	50–54	Neoplasms	0.0035	1.0013
Norway, males, 1920	15–19	Tuberculosis	0.0075	1.0027
New Zealand, females, 1891	35–40	Maternal mortality	0.0253	1.0050
Mexico, males, 1960	60–65	Cardiovascular	0.0353	1.0064

SOURCE.—Preston, Keyfitz, and Schoen [10].

death rate below the true life table death rate. While it is true that there may be larger errors in the data themselves, that does not mean that the error due to population growth should be ignored. If we cannot increase the accuracy of our data, we should at least retain as many of them as possible.

APPLICATIONS

Our knowledge of the factors responsible for gains in life expectancy in the past century has relied primarily on cause-of-death data from England and Wales and the United States, the representativeness of which is unknown. It is now possible to prepare a more broadly based description of mortality variation. The techniques described above were applied to data from 180 national populations, and basic parameters of cause-of-death life tables were published for each (Preston, Keyfitz, and Schoen [10]). Of the 180 populations, 165 are studied here, the remainder

having been discarded on the grounds of palpably poor coding practices or duplication of coverage. The 165 populations represent the period from 1861 to 1964, 48 nations, and life expectancies from 26 to 76 and are distributed regionally and temporally as shown in Table 2. Eleven causes of death are distinguished. The causes of death, and their corresponding numbers in the A or B list of the Seventh Revision of the International Classification of Causes of Death, are as follows: respiratory tuberculosis (B1); other infectious and parasitic diseases (B2–B17); malignant and benign neoplasms (B18–B19); cardiovascular disease (B22, B24–B29; A85, A86); influenza, pneumonia, bronchitis (B30–B32); diarrhea, gastritis, enteritis (B36); certain chronic diseases (B20, B33, B37, B38) (these numbers represent, respectively, diabetes mellitus, ulcer of stomach

TABLE 2

DISTRIBUTION OF POPULATIONS FORMING THE BASIS OF COMPUTATIONS

	Pre-1900	1900–29	1930–59	1960–64	Total
Northern and western Europe	4	8	20	27	59
Southern and eastern Europe	2	4	5	18	29
Overseas Europe*	2	11	15	8	36
Africa, Asia, and Latin America†	.....	4	11	26	41
Total	8	27	51	79	165

* Includes South Africa (white).

† Includes Israel (Jewish population) and South Africa (colored).

and duodenum, cirrhosis of liver, and nephritis and nephrosis); maternal mortality (B40); certain diseases of infancy (B42–B44); violence (BE47–BE50); and all other and unknown causes. The reader is referred to *Causes of Death: Life Tables for National Populations* [10] for a more detailed description of the cause-of-death categories.

We begin by examining scatter plots of the relationship between $l_0^{(i)}$, the chance that a randomly drawn person of a particular sex aged zero in the period life table will die from cause i, and $\mathring{e}_0$, the period life expectancy at birth for that sex. Figures 1–3 show examples of such scatter plots for the female populations. All observations are pooled; we do not consider in detail in this paper how relationships might vary with time or space. If data were available for a country at two points in time separated by less than five years, the plotting of the later observation was suppressed, although its value is incorporated in all calculations. The figures are computer-drawn from punched output produced simultaneously with the final run of tables appearing in *Causes of Death: Life Tables for*

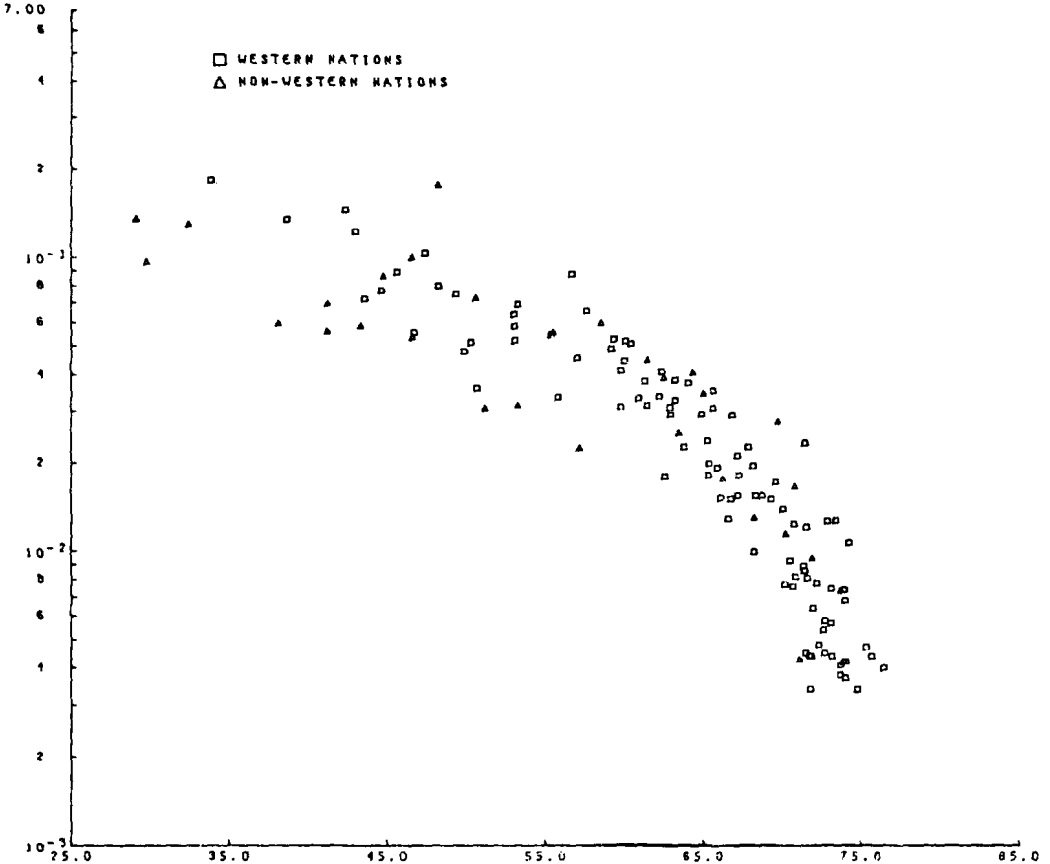

FIG. 1.—Probability of eventual death from other infectious and parasitic diseases for females aged 0. Abscissa is expectation of life at birth.

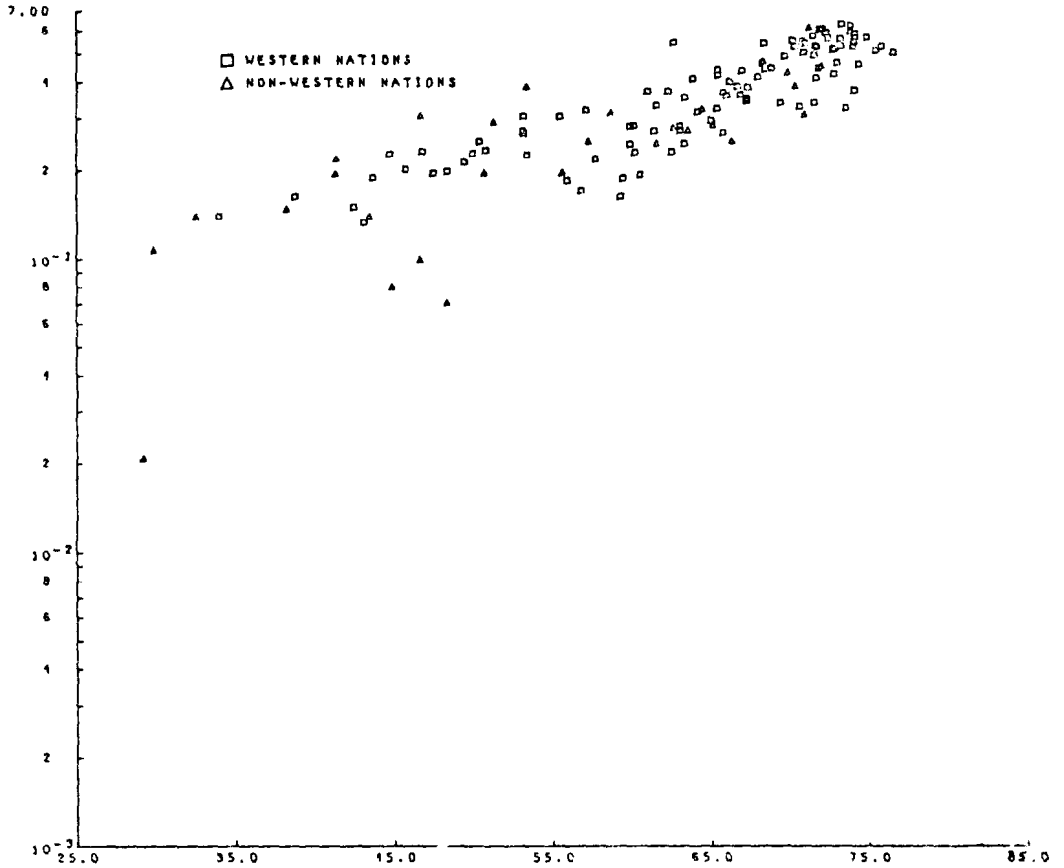

FIG. 2.—Probability of eventual death from cardiovascular disease for females aged 0. Abscissa is expectation of life at birth.

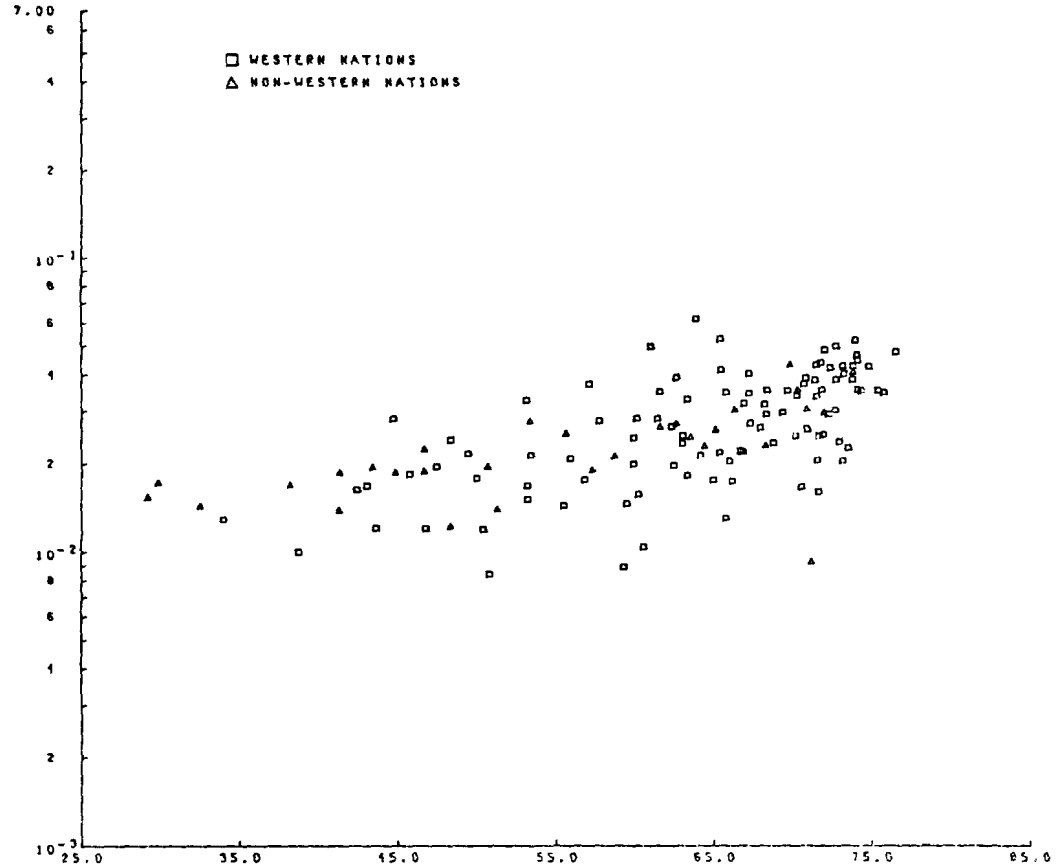

FIG. 3.—Probability of eventual death from all accidents and violence for females aged 0. Abscissa is expectation of life at birth.

National Populations; consequently, there is minimal chance of manual or transcription error.

The three figures presented exhibit the major types of relationships present in the data. First, the probability of dying from "other infectious and parasitic diseases" (excluding respiratory tuberculosis) declines quite regularly as life expectancy advances. Second, the chance of dying from cardiovascular disease rises spectacularly, from about 10 per cent to about 60 per cent over the observed range of mortality, a rise that is completely caused by the more rapid disappearance of other causes of death (see the later sections of this paper and Preston and Nelson [11]). Finally, the chances of dying from violence (suicide, homicide, and accidents) are relatively invariant with the level of life expectancy. When death rates decline from all causes combined, they also decline from violence, and by an amount sufficient to alter the chances of eventual death from this cause by only moderate amounts.

Note that observations for non-Western countries—Africa (except South Africa whites), Asia, and Latin America—are scattered more or less randomly through the more numerous observations for Western countries. This is also true of scatter plots for other diseases and for males. At a given level of life expectancy, the cause-of-death structure of non-Western countries shows little systematic difference from that of Western countries, despite frequently alleged differences in the manner by which advances in mortality were attained in the two regions.

A more precise way of expressing these relationships is to fit curves to the scatter plots and use the coefficients of those curves to predict the chances of death from various causes at different levels of life expectancy. Curve fitting seems preferable to averaging observations within a certain range because it permits estimates at a particular level to be based upon observations at other levels and does not disregard variability within an arbitrarily chosen range. The added stability becomes quite important for the lower life expectancies, where fewer observations are available. We computed second-degree polynomials for each relationship, fitted by classical least-squares regression. The coefficients of these relationships are such that the predicted values of $l_0^{(i)}$ always sum to 1.000, a necessary feature of such a system of equations (Espenshade [2]). At very high life expectancies for minor causes of death the predictions were occasionally negative, in which case values from a logarithmic regression are shown and other values adjusted proportionally to sum to 1.000. A modest extrapolation of the relationships back to a life expectancy of 25 years and forward to one of 75 for males and 80 for females permits some educated

guesses regarding the past and future probabilities of dying from various causes.

The results, shown in Table 3, suggest that, at a life expectancy of 25, approximately 60 per cent of those born will ultimately succumb to an infectious or parasitic disease, when the group is broadened to include influenza, pneumonia, bronchitis, diarrheal diseases, and maternal mortality. Only one-tenth of that number die from this group of causes at a life expectancy of 75 or 80. Over the same range, the chance of dying from the combination of neoplasms and cardiovascular disease increases fivefold, from 15 to 75 per cent. These tendencies are not unfamiliar, of course, but Table 3 assigns more exact dimensions to them than it was previously possible to do. It also permits an assessment of the degree of abnormality in the recorded cause-of-death structures of individual populations. For example, United States males in 1964 had a life expectancy of approximately 66.9 (Preston, Keyfitz, and Schoen [10]). Interpolating in Table 3, we find that this figure yields an expected probability of dying from cardiovascular disease of 0.429, in contrast to the actual probability of 0.569. Thus, United States males had a 33 per cent greater chance of dying from cardiovascular disease than would be expected at their level of life expectancy. On the other hand, their probability of dying from the aggregate of infectious diseases was only 0.051, compared to an expected 0.104. Obviously, the United States male population achieved its life expectancy through an unusual and atypical cause of death structure, one with important social and demographic implications. Results for American females are in the same direction but are not quite so striking.

Infectious diseases are often thought to afflict children much more than adults and to modify life expectancy primarily through their activity during childhood years. However, our calculations show that in high-mortality populations a person aged 15 has about the same chance of eventually dying from an infectious disease as a child aged 0. Table 4 presents information equivalent to that in Table 3, derived in an identical manner, but referring to the chances of dying for a person who has survived to age 15. At a life expectancy of 25, the chances of ultimate death from an infectious disease are virtually identical (0.56 versus 0.57) for persons aged 15 and for infants. At higher life expectancies, both series decline, the one for age 15 declining somewhat faster. The equivalence between the series at low life expectancies results from the countervailing influences of tuberculosis, primarily a disease of young adults, and "other infectious and parasitic diseases," consisting primarily of well-defined childhood diseases.

TABLE 3

PROBABILITY ($\times 100$) THAT A PERSON AGED 0 WILL EVENTUALLY DIE FROM A PARTICULAR CAUSE AT VARIOUS LEVELS OF LIFE EXPECTANCY

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)
	Females											
25.00	8.24	14.60	0.80	16.57	24.92	8.62	0.93	1.49	5.91	2.03	15.89	57.87
30.00	8.40	13.20	1.53	14.53	22.88	8.51	1.09	1.54	5.50	1.82	21.00	54.53
35.00	8.35	11.75	2.47	13.72	20.82	8.21	1.28	1.55	5.07	1.71	25.07	50.68
40.00	8.00	10.17	3.60	14.09	18.56	7.66	2.42	1.49	4.56	1.67	27.78	45.88
45.00	7.43	8.60	4.92	15.78	16.36	6.94	3.52	1.39	4.06	1.73	29.62	40.72
50.00	6.66	7.07	6.43	18.83	14.27	6.08	4.33	1.24	3.57	1.90	29.63	35.32
55.00	5.70	5.59	8.14	23.25	12.28	5.07	4.83	1.04	3.09	2.18	28.83	29.68
60.00	4.55	4.16	10.04	29.03	10.41	3.92	5.03	0.80	2.61	2.55	26.90	23.84
65.00	3.20	2.77	12.13	36.15	8.65	2.62	4.94	0.52	2.15	3.04	23.83	17.76
70.00	1.65	1.42	14.42	44.60	6.99	1.18	4.55	0.26	1.70	3.63	19.60	11.50
75.00	0.67*	0.66*	16.48	53.09	5.31	0.52*	3.75	0.18*	1.23	4.21	13.90	7.34
80.00	0.44*	0.42*	18.26	61.23	3.75	0.34*	2.66	0.12*	0.76	4.77	7.25	5.07

* Value predicted from logarithmic regression: $\log l_0^i = a_0^i + b_0^i \mathring{e}_0$. All others are predicted from polynomial regression: $l_0^i = A_0^i + B_0^i \mathring{e}_0 + C_x^i (\mathring{e}_0)^2$.

TABLE 3—*Continued*

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)
	Males											
25.00	7.99	14.06	2.04	13.47	26.11	7.93	1.36	.....	6.26	2.13	18.65	56.09
30.00	8.55	12.79	1.70	12.00	23.56	8.00	1.56	.....	5.93	3.32	22.59	52.90
35.00	8.79	11.41	1.91	11.86	20.99	7.79	1.87	.....	5.53	4.34	25.51	48.98
40.00	8.58	9.81	2.64	12.94	18.25	7.26	3.32	.....	5.02	5.15	27.03	43.90
45.00	8.08	8.24	3.92	15.42	15.69	6.51	4.38	.....	4.48	5.78	27.50	38.52
50.00	7.27	6.65	5.77	19.33	13.33	5.55	5.04	.....	3.91	6.24	26.91	32.80
55.00	6.15	5.07	8.17	24.66	11.16	4.37	5.30	.....	3.32	6.53	25.27	26.75
60.00	4.73	3.48	11.12	31.41	9.18	2.98	5.18	.....	2.70	6.63	22.59	30.37
65.00	2.99	1.90	14.64	39.57	7.39	1.38	4.65	.....	2.06	6.56	18.86	13.66
70.00	1.40*	0.83*	18.36	48.25	5.68	0.45*	3.66	.....	1.37	6.20	13.80	8.36
75.00	0.98*	0.53*	21.81	56.20	4.09	0.28*	2.25	.....	0.66	5.52	7.68	5.88

* Value predicted from logarithmic regression: $\log l_0^i = a_0^i + b_0^i \mathring{e}_0$. All others are predicted from polynomial regression: $l_0^i = A_0^i + B_0^i \mathring{e}_0 + C_x^i (\mathring{e}_0)^2$.

TABLE 4

PROBABILITY (×100) THAT A PERSON AGED 15 WILL EVENTUALLY DIE FROM A PARTICULAR CAUSE AT VARIOUS LEVELS OF LIFE EXPECTANCY

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)
	Females											
25.00	14.33	9.53	2.13	23.21	25.30	4.79	1.42	3.14	.....	1.98	14.17	57.09
30.00	13.37	8.52	3.09	20.99	22.60	4.55	1.60	2.91	.....	1.73	20.64	51.95
35.00	12.30	7.51	4.18	19.92	20.05	4.25	1.70	2.67	.....	1.55	25.87	46.78
40.00	10.97	6.41	5.37	19.79	17.36	3.84	3.13	2.37	.....	1.46	29.30	40.95
45.00	9.60	5.37	6.69	20.95	14.96	3.38	4.22	2.07	.....	1.47	31.29	35.38
50.00	8.18	4.40	8.14	23.41	12.82	2.89	4.98	1.76	.....	1.56	31.86	30.05
55.00	6.72	3.48	9.72	27.16	10.96	2.36	5.42	1.44	.....	1.77	30.97	24.96
60.00	5.22	2.62	11.44	32.22	9.36	1.78	5.53	1.12	.....	2.08	28.63	20.10
65.00	3.68	1.83	13.30	38.59	8.02	1.17	5.31	0.79	.....	2.47	24.84	15.49
70.00	2.09	1.08	15.30	46.28	6.96	0.51	4.75	0.45	.....	2.99	19.59	11.09
75.00	0.68*	0.51*	17.26	54.78	6.11	0.36*	3.81	0.18*	.....	3.56	12.75	7.84
80.00	0.42*	0.34*	18.98	63.26	5.45	0.26*	2.53	0.12*	.....	4.14	4.50	6.59

* Value predicted from logarithmic regression: $\log l^i_{15} = a^i_{15} + b^i_{15}\mathring{e}_0$. All others are predicted from polynomial regression: $l^i_{15} = A^i_{15} + B^i_{15}\mathring{e}_0 + C^i_{15}(\mathring{e}_0)^2$.

TABLE 4—*Continued*

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)
	Males											
25.00	14.40	10.02	2.69	18.74	27.47	4.25	2.19	.....	.....	4.33	15.91	56.14
30.00	13.82	8.94	2.73	17.66	24.03	3.93	2.38	.....	.....	5.22	21.29	50.72
35.00	12.97	7.80	3.20	17.67	20.69	3.56	2.79	.....	.....	5.95	25.37	45.02
40.00	11.74	6.58	4.10	18.62	17.37	3.10	4.28	.....	.....	6.45	27.76	38.79
45.00	10.34	5.43	5.47	20.82	14.46	2.62	5.31	.....	.....	6.73	28.82	32.85
50.00	8.76	4.33	7.29	24.22	11.94	2.13	5.88	.....	.....	6.98	28.47	27.16
55.00	7.03	3.30	9.59	28.86	9.84	1.61	5.99	.....	.....	7.00	26.78	21.78
60.00	5.13	2.32	12.35	34.73	8.14	1.08	5.65	.....	.....	6.88	23.72	16.67
65.00	3.08	1.40	15.56	41.82	6.84	0.52	4.86	.....	.....	6.60	19.32	11.84
70.00	1.41*	0.64*	19.06	49.64	5.89	0.26*	3.58	.....	.....	6.11	13.41	8.20
75.00	0.94*	0.44*	22.48	57.30	5.25	0.18*	1.85	.....	.....	5.37	6.19	6.81

* Value predicted from logarithmic regression: $\log l^i_{is} = a^i_{is} + b^i_{is}\mathring{e}_0$. All others are predicted from polynomial regression: $l^i_{is} = A^i_{is} + B^i_{is}\mathring{e}_0 + C^i_{is}(\mathring{e}_0)^2$.

At low life expectancies, the probability of dying from cardiovascular disease or cancer is increased considerably if one survives to age 15. But at the highest levels of life expectancy, the chance of ultimately dying from one of these diseases is largely unaffected by survival to age 15, since almost no one dies before this age.

We pass now to the question of gains in life expectancy resulting from the "elimination" of deaths from a particular cause (assuming, as before, independence of causes). It follows from equation (22) below that the gain from eliminating cause (i), $\Delta^{(i)}\dot{e}_0$, is greater the higher the mortality is from (i) and the lower the mortality is from $(-i)$.

$$\Delta^{(i)}\dot{e}_0 = \int_0^\infty \left[\exp\left(-\int_0^x \mu_t^{(-i)}dt\right)\right]dx$$

$$- \int_0^\infty \left[\exp\left(-\int_0^x \mu_t^{(-i)}dt\right)\right]\left[\exp\left(-\int_0^x \mu_t^{(i)}dt\right)\right]dx$$

$$= \int_0^\infty \exp\left(-\int_0^x \mu_t^{(-i)}dt\right)\left[1 - \exp\left(-\int_0^x \mu_t^{(i)}dt\right)\right]dx\,. \qquad (22)$$

Since death rates from a number of causes examined in this paper varied positively with death rates from their respective $(-i)$ causes, it is not evident from equation (22) whether the effects of eliminating those causes would be greater at high or at low life expectancies. Table 5 resolves this question. It indicates that the effect on $\dot{e}_0$ of eliminating most causes decreases as life expectancy increases. Like Tables 3 and 4, it is developed by fitting second-degree polynomials to the 165 observations on $\Delta^{(i)}\dot{e}_0$ and $\dot{e}_0$. We omit extrapolations in this case, since the nature of the variable of concern is already hypothetical.

Several interesting features of Table 5 can be noted. The effect of eliminating certain diseases of infancy is relatively constant over a wide range of life expectancies. As mortality declines, fewer and fewer infants die from this cause, but those hypothetically "saved" live longer because of declines in other causes. The same reasoning accounts for the slow decline in gains from eliminating tuberculosis.

These interrelations are partially suppressed by eliminating combinations of causes. It is easy to show that the effect of eliminating two causes is greater than the sum of effects of eliminating one at a time, a proposition made intuitively obvious by considering a population where only two causes operate altogether. Table 5 indicates that the joint elimination of respiratory tuberculosis, influenza, pneumonia, bronchitis, diarrheal diseases, maternal mortality, and other infectious and parasitic diseases

TABLE 5

EXPECTED GAIN IN AVERAGE LENGTH OF LIFE FROM ELIMINATING VARIOUS CAUSES OF DEATH OR COMBINATIONS OF CAUSES

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)	Life Expectancy Eliminating All Causes in Col. 12 plus Ascribed Infectious Diseases (13)
	Females												
30.00.....	2.53	5.05	0.25	3.38	7.62	3.00	0.04	0.46	1.94	0.45	7.69	23.98	73.59
35.00.....	2.58	4.57	0.41	2.63	6.64	3.04	0.23	0.50	2.00	0.46	7.84	21.52	72.99
40.00.....	2.55	4.06	0.60	2.24	5.71	2.98	0.45	0.52	2.02	0.48	7.79	19.02	72.61
45.00.....	2.43	3.54	0.80	2.21	4.83	2.82	0.62	0.52	1.99	0.51	7.55	16.48	72.47
50.00.....	2.22	2.99	1.04	2.54	4.00	2.56	0.73	0.49	1.91	0.54	7.10	13.89	72.56
55.00.....	1.94	2.42	1.31	3.23	3.23	2.20	0.79	0.44	1.78	0.57	6.46	11.26	72.87
60.00.....	1.56	1.83	1.60	4.28	2.50	1.74	0.79	0.37	1.61	0.61	5.63	8.58	73.42
65.00.....	1.10	1.22	1.92	5.70	1.82	1.17	0.74	0.28	1.39	0.65	4.59	5.86	74.20
70.00.....	0.56	0.59	2.28	7.47	1.20	0.51	0.63	0.17	1.13	0.70	3.36	3.09	75.21
72.50.....	0.26	0.27	2.46	8.49	0.91	0.14	0.56	0.10	0.98	0.73	2.67	1.69	75.81

TABLE 5—*Continued*

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Disease (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Certain Diseases of Infancy (9)	Violence (10)	All Other and Unknown Causes (11)	(1)+(2)+(5)+(6)+(8) (12)	Life Expectancy Eliminating All Causes in Col. 12 plus Ascribed Infectious Diseases (13)
	Males												
30.00	2.25	4.68	0.26	2.46	7.28	2.79	0.06	.....	2.13	0.89	7.51	21.61	70.23
35.00	2.36	4.23	0.27	2.01	6.31	2.90	0.34	.....	2.25	1.22	7.51	19.36	69.86
40.00	2.36	3.73	0.36	1.89	5.38	2.87	0.56	.....	2.29	1.48	7.31	17.00	69.65
45.00	2.25	3.20	0.53	2.09	4.49	2.70	0.72	.....	2.27	1.67	6.92	14.52	69.60
50.00	2.03	2.63	0.78	2.63	3.65	2.38	0.81	.....	2.16	1.81	6.33	11.93	69.70
55.00	1.70	2.02	1.11	3.49	2.84	1.93	0.83	.....	1.99	1.89	5.55	9.22	69.97
60.00	1.27	1.38	1.53	4.68	2.07	1.33	0.79	.....	1.74	1.89	4.57	6.39	70.40
65.00	0.73	0.70	2.02	6.20	1.34	0.59	0.69	.....	1.42	1.85	3.40	3.46	70.99
67.50	0.41	0.34	2.29	7.09	1.00	0.17	0.61	.....	1.24	1.80	2.75	1.95	72.34

would typically produce a life expectancy above 50 regardless of the initial level of mortality. Even this figure is depressed by the high proportion of deaths assigned to the "other and unknown" category and to "certain diseases of early infancy," both of which surely contain a sizable number of deaths of infectious origin, especially in high-mortality populations. To account in a rough way for the infectious deaths lost in this manner, we allocate all deaths in these two categories to the remainder of causes in proportion to the number of deaths originally ascribed to those other causes. This allocation is performed separately for each age interval and each sex in each population, with the results displayed in *Causes of Death: Life Tables for National Populations*. We then recompute the gain from eliminating the group of actual plus ascribed infectious diseases. Column 13 of Table 5 presents the surprising results. Regardless of the initial level of mortality, eliminating the aggregate of actual plus attributed infectious diseases produces a life expectancy between 72 and 76 for females and between 69 and 72.5 for males. These results are another way of saying that infectious diseases bear almost exclusive responsibility for producing life expectancies below levels characteristic of contemporary Western countries. Conversely, the results testify to the almost complete lack of medical progress against noninfectious diseases. While gains have been registered in some areas, their effects seem to have been small and perhaps counterbalanced by less desirable by-products of affluence.

At equivalent life expectancies for the two sexes, the elimination of infectious diseases would result in considerably greater gains for females. Compared with males, females tend to attain a certain level of mortality with a higher incidence of death from infectious diseases and lower incidence from noninfectious diseases. Consequently, the sizable male disadvantage in contemporary life expectancies could have been anticipated at any level of life expectancy, given foreknowledge of the cause-pattern of mortality decline. One might say by way of summarizing Table 5 that the mortality profile of a modernized country is typically present in every population, awaiting only a successful attack on infectious diseases before emerging.

Finally, we present in Table 6 the average gain in productive years of life from eliminating various causes or combinations of causes. We define the expectation of productive life (EPL) as the average number of years lived between ages 15 and 65 for a person who survives to age 15, so that, in the basic life table,

$$EPL = \left(\int_{15}^{65} l_x dx\right) \Big/ l_{15} . \tag{23}$$

TABLE 6

EXPECTED GAIN IN AVERAGE LENGTH OF PRODUCTIVE LIFE FROM ELIMINATING VARIOUS CAUSES OF DEATH OR COMBINATIONS OF CAUSES

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Diseases (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Violence (9)	All Other and Unknown Causes (10)	(1)+(2)+(5)+(6)+(8) (11)	All in Col. 11 plus Ascribed Infectious Diseases (12)
	Females											
30.00	2.76	1.38	0.23	0.99	2.91	0.36	0.12	0.74	0.22	2.02	9.17	11.56
35.00	2.57	1.19	0.32	1.04	2.24	0.30	0.20	0.69	0.22	1.88	7.76	9.76
40.00	2.34	1.01	0.39	1.07	1.67	0.25	0.26	0.63	0.22	1.73	6.45	8.08
45.00	2.09	0.84	0.46	1.06	1.18	0.21	0.30	0.56	0.22	1.56	5.24	6.54
50.00	1.80	0.68	0.51	1.03	0.78	0.16	0.31	0.49	0.23	1.37	4.13	5.13
55.00	1.49	0.53	0.55	0.97	0.47	0.12	0.31	0.40	0.23	1.17	3.12	3.86
60.00	1.14	0.38	0.59	0.88	0.25	0.08	0.29	0.32	0.23	0.95	2.21	2.71
65.00	0.77	0.25	0.61	0.76	0.11	0.04	0.24	0.22	0.23	0.72	1.41	1.70
70.00	0.36	0.13	0.62	0.61	0.06	0.01	0.18	0.12	0.23	0.47	0.70	0.83
72.50	0.15	0.07	0.63	0.52	0.07	0.00	0.14	0.07	0.23	0.34	0.38	0.44

TABLE 6—*Continued*

Life Expectancy at Birth	Respiratory Tuberculosis (1)	Other Infectious and Parasitic Diseases (2)	Neoplasms (3)	Cardiovascular Diseases (4)	Influenza, Pneumonia, Bronchitis (5)	Diarrheal Diseases (6)	Diabetes, Cirrhosis of Liver, Nephritis, Stomach Ulcer (7)	Maternal Mortality (8)	Violence (9)	All Other and Unknown Causes (10)	(1)+(2)+(5)+(6)+(8) (11)	All in Col. 11 plus Ascribed Infectious Diseases (12)
	Males											
30.00.....	2.54	1.49	0.14	1.00	2.95	0.31	0.15	.....	1.00	2.09	8.19	10.57
35.00.....	2.39	1.27	0.18	1.00	2.24	0.26	0.24	.....	1.10	1.91	6.80	8.69
40.00.....	2.20	1.05	0.23	0.99	1.64	0.21	0.31	.....	1.17	1.72	5.51	6.97
45.00.....	1.95	0.85	0.28	0.97	1.14	0.17	0.35	.....	1.21	1.52	4.34	5.42
50.00.....	1.65	0.66	0.33	0.95	0.74	0.13	0.36	.....	1.21	1.30	3.28	4.04
55.00.....	1.31	0.49	0.39	0.93	0.44	0.09	0.35	.....	1.18	1.07	2.33	2.82
60.00.....	0.91	0.32	0.45	0.90	0.25	0.06	0.31	.....	1.12	0.83	1.48	1.77
65.00.....	0.46	0.17	0.52	0.86	0.15	0.02	0.24	.....	1.02	0.57	0.75	0.88
67.50.....	0.19	0.09	0.55	0.84	0.12	0.01	0.20	.....	0.96	0.43	0.42	0.50

The gain in *EPL* from the elimination of cause (i) is therefore

$$\Delta^{(i)} EPL = \left(\int_{15}^{65} \bar{l}_x^{(-i)} dx\right) \Big/ \bar{l}_{15}^{(-i)} - \left(\int_{15}^{65} l_x dx\right) \Big/ l_{15} . \qquad (24)$$

Once again we are reminded that infectious diseases take a heavy toll of lives during the productive years, as well as in childhood and old age. At a life expectancy of 30 years, an average of 9–10 years of potential life during the productive span are lost to mortality from the aggregated infectious diseases alone.

Respiratory diseases are responsible for the largest foreshortening of productive life at the lowest levels of life expectancy. For a long span thereafter, tuberculosis is pre-eminent, largely by virtue of its unique age incidence. While tuberculosis never ranks above third in years lost at birth, it is first over a majority of mortality levels for both sexes in terms of loss of productive life. Tuberculosis ultimately gives way to cancer for women and violence for men. Despite the major and well-publicized role of cardiovascular disease in the modern mortality structure, its effects are heavily concentrated above age 65. In the United States in 1964 cardiovascular disease held a slight margin over cancer and violence as the leading source of loss of productive life (0.643 versus 0.637 years for females and 1.385 versus 1.270 years for males). Even in this atypical population, the margin seems unexpectedly small. Combined with the probable sensitivity of the several death rates to public expenditures, the results suggest that public health moneys are probably most appropriately directed toward the elimination or reduction of cancer and violent deaths. The ranking of diseases as a threat to public health depends upon the index chosen, however. The more equally ages are treated, the more serious a public health problem cardiovascular disease becomes.

REFERENCES

1. CHIANG, C. L. *Introduction to Stochastic Processes in Biostatistics.* New York: Wiley, 1968.
2. ESPENSHADE, THOMAS. "A Generalized Matrix Proof of an Adding-up Theorem." Unpublished manuscript, International Population and Urban Research, University of California, Berkeley, 1972.
3. GRAUNT, JOHN. *Natural and Political Observations Made upon the Bills of Mortality.* Baltimore: Johns Hopkins Press, 1939 (original ed., 1662).
4. GROVE, R. D., and HETZEL, A. M. *Vital Statistics Rates in the United States, 1940–1960.* National Center for Health Statistics, Public Health Service Pub. No. 1677. Washington: U.S. Government Printing Office, 1968.
5. JORDAN, C. W., JR. *Life Contingencies.* Chicago: Society of Actuaries, 1967.

6. KEYFITZ, NATHAN. "A Life Table That Agrees with the Data. II." *Journal of the American Statistical Association*, 1968, pp. 1253–68.
7. ———. "Finding Probabilities from Observed Rates, or How to Make a Life Table." *American Statistician*, XXIV (1970), 28–33.
8. LEW, E. A. Discussion of paper by Krall and Hickman, *TSA*, XXII (1971), 163–79.
9. NEWMAN, P. *Malaria Eradication and Population Growth*. Bureau of Public Health Economics Research Series No. 10. Ann Arbor: University of Michigan, School of Public Health, 1965.
10. PRESTON, S. H.; KEYFITZ, NATHAN; and SCHOEN, ROBERT. *Causes of Death: Life Tables for National Populations*. New York: Seminar Press, 1972.
11. PRESTON, S. H., and NELSON, V. E. "Structure and Change in Mortality by Cause: An International Summary." Unpublished manuscript, Department of Sociology, University of Washington, 1972.

DISCUSSION OF PRECEDING PAPER

T. N. E. GREVILLE:

This paper consists of two parts, a part dealing with methodology and a part under the heading "applications," concerned with international comparisons of longevity, with emphasis on the effects of the manner in which deaths are distributed by major groups of causes. This latter part of the paper is extremely valuable and illuminating. The technique of analyzing life tables for many countries by cause-of-death groups and then plotting expectation of life against the probability of eventually dying from a given "cause" (i.e., group of related causes) is especially helpful. Of particular interest is the finding that the United States is in certain important respects not typical of "Western" countries.

I am inclined to be more critical of the methodological portion of the paper. I emphasize, however, that my criticisms of methodology in no way affect the conclusions drawn in the second part of the paper, as the use of a different methodology would not make a penny's worth of difference in the numerical results obtained. In brief, my criticism is that, in my opinion, the authors make exaggerated claims for their suggested methodology.

The methodological part of the paper deals with two distinct problems: (1) bridging the gap between age-specific death rates and probabilities of dying in a general life table in which there is no differentiation by causes of death and (2) construction of a multiple decrement table, and the associated single decrement tables, when certain cause-of-death groups are treated as separate decrements. As I see it, these two problems involve quite different considerations, and it is convenient to treat them separately.

Calculation of Probabilities from Rates

The calculation of the probability of dying associated with an age-specific death rate when the latter spans an age interval of more than one year is a problem that has been much studied, and many methods have been proposed. There probably will never be general agreement on any single method that is unequivocally "best."

The method used in this paper is based on the equation

$$ {}_nM_x = \frac{{}_nD_x}{{}_nP_x} = \frac{\int_0^n P_{x+t}\mu_{x+t}dt}{\int_0^n P_{x+t}dt}, \qquad (1) $$

where $P_x dx$ is the number of person-years lived in a given year in the infinitesimal age interval x to $x + dx$. If it is assumed that P_{x+t} is proportional to l_{x+t}, the population death rate ${}_nM_x$ reduces to ${}_nm_x$, the central death rate of the life table. However, the authors very properly desire to take into account the fact that the age distribution within the n-year age interval in the actual population is typically different from that in the stationary population of the life table. Accordingly, they take P_{x+t} proportional to $l_{x+t} \exp(-tr_x)$, where $-r_x$ is an "imputed stable rate of growth within the age interval." It is here that questions may legitimately be raised.

This assumption is interpreted to mean that cohort size increases or decreases in geometric progression within the n-year age interval. It is not brought out, however, that a corollary to this assumption is that there are abrupt shifts in cohort size at the ages that bound the n-year intervals.

In my opinion, the authors have concealed from us the real reason for the assumption that P_{x+t} is proportional to $l_{x+t} \exp(-tr_x)$, which is one of mathematical convenience. Let

$$P_{x+t} = l_{x+t} g(x + t) ,$$

where the function $g(x + t)$ is defined by this relation. Then, substitution in expression (1) and integration by parts in the numerator gives

$${}_nM_x = \frac{l_x g(x) - l_{x+n} g(x + n) + \int_0^n l_{x+t} g'(x + t) dt}{\int_0^n l_{x+t} g(x + t) dt} .$$

It is now clear that if $g(y)$ satisfies

$$g'(y) = c g(y) \tag{2}$$

(for all y and some constant c), a substantial simplification is effected, since there is only one integral to be evaluated rather than two. The solution of equation (2) is, of course,

$$g(y) = k e^{cy}$$

for some constants k and c. Taking $c = -r_x$ for y in the interval x to $x + n$ gives the substitution in the paper. As I have stated, I think that this is the real reason for making the assumption described and that the interpretation in terms of stability within the age interval is a rationalization after the fact.

I certainly have no serious objection to this "geometric" assumption, if it is frankly acknowledged that it is made primarily to simplify the mathematics. In this respect, it is analogous to the commonly used assumption of "uniform distribution of deaths."

However, in my opinion, a convincing case has not been made that the proposed iterative method is in any way superior to the method of interpolating populations and deaths separately in order to achieve age intervals of one year, within which the difference between M_x and m_x can be neglected.[1] The iterative method is "more faithful to the data" only if faithfulness to the data is defined in terms of the "geometric" assumption.

Cohort size varies not according to a mathematical rule but in response to wars, epidemics, severe economic crises, and other environmental factors. While the general trend is upward in time (i.e., downward with increasing attained age), there are numerous ups and downs. In my opinion, a reasonable interpolation procedure is likely to give results closer to reality than a rigid pattern of geometric increase or decrease over five-year intervals (in the typical case) with abrupt shifts every five years. (There is some inconsistency in the assumptions used in the paper, since the formula for l_x assumes stability over five-year intervals, while that for r_x supposes stability over a fifteen-year interval.)

If the desired end product is a life table containing values for all single years of age (as is not the case in the applications in this paper), the "separate interpolation" approach is certainly simpler and does not involve iteration. While it is undeniably true that the computer facilitates extensive calculations that would be impractical without it, surely its availability imposes no obligation to use cumbersome methods when simpler ones will do as well.

In this paper ${}_nP_x$ is defined as the number of years lived during the given year by persons in the age interval x to $x + n$. It is tacitly assumed, but never explicitly stated, that this will be approximated by the population in the age interval on some fixed date. This is a usual assumption in demographic work, but it should be stated.

Cause-of-Death Analysis

This part of the methodology involves construction of a multiple decrement table in which the decrements are groups of related causes of death, and construction of the associated single decrement tables

[1] See Nathan Keyfitz, "A Life Table That Agrees with the Data. II," *Journal of the American Statistical Association*, 1968, pp. 1253–68.

(ASDT's). I quote two sentences in full, as I wish to comment on them at some length.

> In constructing ASDT's, a number of writers (e.g., Chiang [1]) have supposed that all death rates are constant during each age interval, or that the several causes act in a fixed ratio to one another. While that assumption may be perfectly acceptable for some causes of death, it introduces distortions when such causes as accidents, tuberculosis, or maternal mortality are considered.

The first quoted sentence mentions two quite different assumptions, of which the second is much more plausible and defensible than the first. Yet the second sentence continues with "that assumption" as if only one had been mentioned. It is also not clear which assumption is being attributed to Chiang, and I would like to reassure the reader that he recommends only the fixed-ratio assumption.

The iterative procedure recommended for computing the quantities $l_x^{(i)}$ in the multiple decrement table has the virtue that if $l_x^{(i)}$ and $l_x^{(-i)}$ are independently calculated by this method, they should, at least in theory, satisfy the consistency requirement

$$l_x = l_x^{(i)} + l_x^{(-i)} .$$

However, careful attention to rounding error might be needed to achieve this result numerically.

The same requirement is satisfied more easily by taking

$${}_nd_x^{(i)} = ({}_nM_x^{(i)}/{}_nM_x){}_nd_x ,$$

derived from the fixed-ratio assumption, and one wonders whether there is a significant numerical difference in the results obtained by the two methods.

In the light of the above quotation we are startled to find that the recommended procedure for constructing ASDT's assumes that "during the interval ... the death rate from the excluded cause is constant"! There is no suggestion that this might introduce any distortion.

Note that if this recomendation is followed, the "causes" designated by (i) and $(-i)$ are not treated in a symmetrical, interchangeable fashion. One has a constant force within the interval, while the other does not. The reader should be aware that the fixed-ratio assumption leads to the simple formulas

$${}_np_x^{(i)} = ({}_np_x)^{{}_nM_x^{(i)}/{}_nM_x} , \qquad {}_np_x^{(-i)} = ({}_np_x)^{{}_nM_x^{(-i)}/{}_nM_x}$$

which clearly satisfy the consistency requirement

$${}_np_x = {}_np_x^{(i)}\,{}_np_x^{(-i)}$$

I am impressed by the insight this paper provides into the relationships between longevity and causes of death in countries at different stages of social and economic development, but I am far from convinced that the suggested methodological techniques have the superiority claimed for them.

CECIL J. NESBITT AND HANS U. GERBER:*

This is a fascinating and formidable paper. In one introductory sentence—"More faithful to the data than techniques presently used, the method partially substitutes a dependence on the speed and precision of modern electronic computers for a reliance on assumption, approximation, or graduation"—it sweeps away our traditional approaches to the construction of life tables from population data. Mathematical formulas are superseded by numerical processes, and one finds oneself in new and strange territory in attempting to assess the results. In the applications the authors have performed a huge volume of calculations but have nevertheless refined out some observations which challenge the imagination and at the same time put one on the alert against the possibility of overstretching of interpretation.

In regard to the iterative method, we have the following comments:

1. We are inclined to question how one general iterative process can be applied to the available basic data for a great many different countries. Without being knowledgeable concerning such data, we guess that special adjustments would be required in many cases. For instance, in the United States population life tables have been based on data on births, on deaths over a three-year period, and on a mid-period census. How does this iterative process take all these data into account? Also, on viewing the graph of live births in the United States from 1910 to 1970, we are dubious about how an exponential function e^{-rt} (even with variable r) can take account of the changes in population growth in this country. However, we believe that the authors should be commended for attempting to take account of the growth bias instead of ignoring it.

2. The authors indicate that the iterative process converges to values of l_x and r which satisfy equation (5). In the process, ${}_5L'_x$ is approximated by means of a cubic polynomial fitted through four quinquennial values of $e^{-rt}l_{x+t}$ with, it appears, r constant over a fifteen-year span. On the other hand, it appears that r (and ${}_nh_x$) can vary with x, which is inconsistent with the fifteen-year constancy of r. This mathematical inconsistency may not be of numerical importance.

* Dr. Gerber, not a member of the Society, is assistant professor, Department of Mathematics, University of Michigan.

3. If we denote by ${}_nm'_x$ the central death rate in the stable population, we may view the authors' process as one of identifying ${}_nm'_x$ with the observed ${}_nM_x$. There remains the problem of obtaining the life table function ${}_nL_x$ in order to calculate life expectancies. We are curious as to how ${}_nL_x$ is obtained. One possibility would be by the use of ${}_na_x$, the average number of years lived past age x, by those who die in the interval x to $x + n$:

$$ {}_nL_x = {}_na_x \; {}_nd_x + nl_{x+n} . $$

There is the related remark that the authors' equation (21) may also be applied to the single decrement table in the form

$$ {}_nm_x \doteq 1 + r({}_na_x - {}_nA_x) . $$

4. We do not understand the second part of the statement concerning the process for the multiple decrement table, namely: "The procedure begins with the assumption that the proportion of life table deaths due to cause (i) in an age interval is the same as $M^{(i)}$ divided by M, and iterates to the point where life table deaths from cause (i) are in the proportion that $m^{(i)}$ is of m." We would substitute for the second part the following: "iterates to the point where the deaths in the stable population from cause (i) are in the proportion that $M^{(i)}$ is of M."

5. T. N. E. Greville has laid emphasis on a consistency property whereby, if a life table process is applied to two subsets of data (say, male and female) and then applied to the complete set of data, the total mortality rates are intermediate to the rates for the subsets. We can conceive of cases where the authors' iterative process might not have this property.

The "Applications" section of the paper is highly interesting and provocative in its interpretations. Perhaps the warning should be made that the expectancies and probabilities are based on that somewhat artificial construct the current-year life table. In other words, if a population actually experienced over a period of a year the rates shown in the current-year life table, then the expectancies and probabilities would follow. For populations with rapidly changing size and mortality, the current-year life table approach may not be realistic. However, we do not have any good countersuggestions to offer.

In Table 4 it might have been more relevant to use regression relations on e_{15} rather than on $\mathring{e}_0$.

We consider the paper, and the concomitant book *Causes of Death: Life Tables for National Populations* (authors' ref. [10]), to be a major contribution to the analysis of mortality by cause of death, and welcome its presentation to the Society.

(AUTHORS' REVIEW OF DISCUSSION)

SAMUEL H. PRESTON, NATHAN KEYFITZ, AND ROBERT SCHOEN:

We are grateful to Mr. Greville and to Messrs. Nesbitt and Gerber for their useful responses to our paper. Their discussions will be instrumental in the further development of methodology for cause-of-death life tables. We are also gratified that the discussions raise no serious objection to the major purpose of the present paper, that of summarizing basic parameters of cause-of-death life tables for 165 populations. The methodological sections were appended in order that readers might know and evaluate the procedures used to derive the parameters being summarized; they were not intended to provide an exhaustive description of the method or comparison with other approaches. Greater detail on these matters is available in the volume presenting the basic parameters and elsewhere. Since we tend to agree with Mr. Greville that the choice of methodology has little effect on the outcome, we shall keep our remarks brief.

Mr. Greville raises the question of whether we chose an exponential graduation in the interests of mathematical convenience. If so, he says, we should admit our "guilt." This question reveals the considerable gap between our approaches to numerical estimation. The stable population is the most useful and powerful model in demography, since it provides unique mathematical relationships among demographic variables for a numerically large set of populations that have a simple and interpretable behavioral basis. The exponential graduation is simply an extension of this basic model to the problem of life table construction. It is developed deductively from this prominent model and is hardly a "rationalization after the fact." Despite having only one parameter, this graduation is clearly the most appropriate form when the condition of stability applies, which is the case in a large number of our populations. In most other populations, *sectional* stability is a reasonable assumption (of which total stability is a special case), and it is only this weaker assumption that we impose on the data. If the assumptions produced no formal solutions in terms of observable data (i.e., were not mathematically convenient), we obviously could not make use of them. But this is true of any numerical procedure. It is ironic that the question of convenience arises, since exponential graduation requires iteration, and the iterative computer program required the better part of a well-trained man-year to produce and increased the computational time and costs at least fivefold.

The apparent inconsistency in the paper between assuming r constant in a five- as opposed to a fifteen-year interval results from the difference between theory and application. The formulas assume r to be constant in

a five-year interval, but growth rates cannot be estimated from one observation. Thus we make use of numbers of people in the five-year age groups adjacent to that under consideration in order to estimate r. Mr. Greville claims that we offer no suggestion that our procedure for the cause-deleted tables might introduce distortion; apparently he overlooked the section on "evaluating the iterative method," where we make clear the separate conditions under which our procedure works best and those under which the fixed-ratio assumption works best. We would also like to make clear that we examined our results to ensure that iteration did not lead to any significant departures from the additive consistency requirement for the $l_x^{(i)}$ of the multiple decrement tables or from the multiplicative consistency requirement for the $\bar{l}_x^{(i)}$ of the associated single decrement tables.

Messrs. Nesbitt and Gerber inquire about our formula for ${}_nL_x$. We use the customary procedure of cubic graduation on the l_x values. We derive ${}_na_x$ after the fact through use of the formula which they cite; we do not adopt ${}_na_x$ values from a "standard" population. They also seek clarification of the procedure for deriving the multiple decrement table. The point here is that we do not want life table deaths by cause to be in the ratio of ${}_nM_x^{(i)}/{}_nM_x$, since this ratio is affected by the age distribution in the interval x to $x + n$. Instead, we want to take account of the shape of the $\mu_x^{(i)}$ functions and of the age distribution in the interval to estimate what the ratio ${}_nM_x^{(i)}/{}_nM_x$ would be in a stationary population. This is done by iterating to a set of mutually consistent values of $l_x^{(i)}$, using information on ${}_nr_x$ and on the shape of the cause-specific force of mortality function that is derived from comparisons of death rates in successive age intervals.

Finally, they raise the question of whether it would not have been better to compare the probabilities of dying from a particular cause at age 15 with life expectancy at 15, rather than at birth. We considered this possibility but rejected it because of the advantages of relating all parameters to one index of mortality. Had we followed their suggestion, for example, we could not have made the statement that, in a very-high-mortality population, a person faces approximately the same chance of dying from an infectious disease at age 15 that he faced at birth. There was no reason except brevity for our not showing both types of comparisons.