DEMOGRAPHY© Volume 21, Number 2 May 1984

MORTALITY ESTIMATION FROM REGISTERED DEATHS IN LESS DEVELOPED COUNTRIES

Neil G. Bennett
Department of Sociology, Yale University, New Haven, Connecticut 06520
Shiro Horiuchi
Population Division, The United Nations, New York, New York 10017

Abstract—Age-specific population growth rates were introduced to demographic analysis in earlier work by Bennett and Horiuchi (1981) and Preston and Coale (1982). In this paper, we derive a method which uses these growth rates to transform what may be a set of incompletely recorded deaths by age into a life table that accurately reflects the true mortality experience of the population under study. The method does not rely on the assumption of stability and, for example, in contrast to intercensal cohort survival techniques, is simple to implement when presented with nontraditional intercensal interval lengths. Thus we can obtain mortality estimates for less developed countries with defective data, despite departures from stability. Further, we assess the sensitivity of the method to violations in various assumptions underlying the procedure: error in estimated growth rates, existence of non-zero net intercensal migration, age dependence in the completeness of death registration, and misreporting of age at death and age in the population. We demonstrate the use of the method in an application to data referring to Argentine females during the period 1960 to 1970.

Throughout much of the time period during which indirect estimation has evolved, there have been many countries whose populations have approximated stability. Recently, however, more and more countries have been experiencing rapidly declining mortality and/or declining or fluctuating fertility, and thus have undergone a radical departure from stability. Consequently, previously successful indirect methods, grounded in stable population theory, are with greater frequency ill-suited to the task for which they were devised.

Bennett and Horiuchi (1981) have introduced the use of age-specific population growth rates into the indirect estimation of mortality in order to circumvent the increasingly limited applicability of stable population techniques. This new methodology has recently been extended by Preston and Coale (1982) to cover a wide range of demographic estimation.

There are at least three major approaches to mortality estimation using age-specific growth rates. First, the completeness of death registration may be estimated from inconsistencies among the age structure of population, the age distribution of deaths, and the age-specific growth rates, as shown by Bennett and Horiuchi (1981). Second, the age distribution of deaths found in the observed population can be converted into the distribution of deaths in the life table. Regardless of the extent of registration completeness, one can obtain accurate estimates of life expectancy at various ages using this method as long as certain assumptions approximately hold true. We present the second approach in this paper.[1] Last, the age structure of a closed population can be converted into that of the corresponding stationary population subject to the same mortality conditions. This approach has been elaborated by Preston and Bennett (1983).

In addition to these three approaches which do not necessarily depend on the use of model life tables, Preston (1983) has developed a method for estimating the birth rate and mortality level simultaneously for an intercensal period, by combining the use of age-specific growth rates with the logit model life table system and a Brass-type estimate of child mortality. This method can be conceived of as a special version of the third approach, although distinguishing it from the third is the fact that model life tables were used.

Using the method described below, one can construct life tables which are corrected for any underenumeration of the population or underregistration of deaths. Prior knowledge of the level of completeness of the recording of deaths and population is not necessary. The method is applicable to any closed population and essentially derives from extensions to stable population theory.

DERIVATION

Just over three decades ago, Paul Vincent (1951) published his seminal article describing what is known as the method of extinct generations. By way of this method, we can estimate the number of persons age a, $N(a)$, at time t-x, by cumulating all deaths experienced by that cohort of persons subsequent to that time. Clearly, the cohort must be "extinct" at the time the study is initiated, otherwise $N(a)$ will be underestimated. Unless one is focusing on historical populations, the method is not very useful in practice. However, even for the analysis of historical populations, the method is often impractical since it requires a long time series of death registration data.

If, on the other hand, we want to estimate $N(a)$ in the current population and we know the population is stationary and subject to unchanging mortality, then we can apply the method of extinct generations cross-sectionally to the current age distribution of deaths. The method can be applied in this case because stationarity guarantees that the age distribution of any cohort is identical to the current cross-sectional population. Moreover, this cross–sectional analogue can be extended to stable populations by noting that in a stable population, the following relationship holds:

$$N(a) = \int_a^\infty D^*(x)\exp[r(x - a)]dx, \qquad (1)$$

where $D^*(x)$ is the true number of deaths experienced by those aged x in the current population (Preston et al., 1980). This formula exploits the fact that, in a stable population, the number of deaths to persons age a in a given year differs from that number in the previous year by a factor of $\exp[r]$. Note that when r equals zero, equation (1) reduces to simply the summation of deaths occurring to those above age a which yields the number of persons currently aged a.

Equation (1) can be generalized such that it can be applied to any closed population, without requiring the restrictive assumption of stability (Bennett and Horiuchi, 1981). Thus we have

$$N(a) = \int_a^\infty D^*(x)\exp\left[\int_a^x r(u)du\right]dx,$$

where $r(u)$ is the rate of growth of the population aged u. Suppose though that registered deaths, $D(x)$, underestimate the true number of deaths, $D^*(x)$, by a proportion constant across age such that $D(x)$ equals $kD^*(x)$, for all $x \geq a$. $\hat{N}(a)$ may be defined as

$$\hat{N}(a) = \int_a^\infty D(e)(x)\exp\left[\int_a^x r(u)du\right]dx, (2)$$

where $\hat{N}(a)$ will equal $kN(a)$.

Bennett and Horiuchi (1981) have also shown that

$$_xp_a = \frac{N(a + x)}{N(a)} \exp\left[\int_a^x r(u)du\right],$$

where $_xp_a$ is the probability of survival from age a to age $a + x$. Given the assumption that completeness of death registration does not vary with age, we then have

$$_xp_a = \frac{\hat{N}(a + x)}{\hat{N}(a)} \exp\left[\int_a^x r(u)du\right]. \quad (3)$$

The other life table functions can be straight forwardly derived from the series of $_xp_a$'s. Thus, given the number of registered deaths by age and a set of age-specific growth rates, a life table can be constructed for the population under study.[2]

IMPLEMENTATION

Demographic data are most widely available by five-year age groups, up to the open-ended, highest age interval. Thus the equations in the previous section need to be modified in order to apply this method to the data, and form of which, that are usually available.

From equation (2) it is straightforward to derive

$$\hat{N}(a) = \hat{N}(a + 5)\exp[5_5r_a] + {_5D_a}\exp[2.5_5r_a], \quad (4)$$

where $_5r_a$ represents the rate of growth of the population aged a to $a + 5$, and $_5D_a$, the number of deaths to persons aged a to $a + 5$. At one point in the derivation of equation (4) we arrive at the following equality:

$$\int_0^5 D(a + z)\exp[z_5r_a]dz = {_5D_a}\exp[\tilde{z}_5r_a].$$

We assume $\tilde{z}$ to equal 2.5, though in an age group where the age distribution of deaths is declining rapidly, this is a poor assumption and will contribute to a biased estimate of $N(a)$ in the oldest ages. It becomes necessary, then, to develop a correction factor which will compensate for the error due to this assumption. Equation (4) is thus adjusted to be:

$$\hat{N}(a) = \hat{N}(a + 5)\exp[5_5r_a] + {_5\gamma_a} \cdot {_5D_a}\exp[2.5_5r_a]. \quad (5)$$

By simulating populations with a wide range of $_5r_x$ and $_5M_x$ (the observed death rate in the population aged x to $x + 5$), we have generated a regression equation which provides an estimate of $_5\gamma_x$. Below age 60, any adjustment is likely to be inconsequential due to the relative subtlety of the curvature in these ages and the small influence of $_5D_a\exp[2.5_5r_a]$ as compared with $\hat{N}(a + 5)\exp[5_5r_a]$ in the determination of $\hat{N}(a)$. For $x \geq 60$, the following estimation formula is recommended:

$$_5\gamma_x = 1.00 - 2.26_5r_x \cdot {_5M_x} + 0.218_5r_x - 0.826_5r_x{}^2.$$

In order to begin the estimation process implied by equation (5), we must first find $\hat{N}(A)$, where A is the lower bound of the open interval in the age distribution. The following equation has been found[3] to closely approximate the relationship among $D^*(A+)$, the true number of deaths to persons age A and above, $e(A)$, the expectation of life at age A, $r(A+)$, the rate of growth of the number of persons age A and above, and $N(A)$:

$$N(A) \simeq D^*(A+)(\exp[r(A+)e(A)] - \{[r(A+)e(A)]^2/6\}). \quad (6)$$

After obtaining $\hat{N}(A)$ (by substituting $D(A+)$, the registered number of deaths to those aged A and above, for $D^*(A+)$ in equation (6)), we can determine all other $\hat{N}(x)$'s ($x = 0,5,\ldots,A-5$) using equation (5).

We can see in equation (6) that we need an estimate of $e(A)$ as input. This is a difficult problem, at best, because the estimate cannot be derived as a function of the inverse of the death rate above age A since that death rate may be biased due to the incomplete recording of deaths and the likely nonstationarity of the population above age A (see, e.g., Horiuchi and Coale, 1982). Furthermore, the power of the technique would be enhanced if we could obtain the estimate of $e(A)$ directly from the age distribution of deaths, without having to rely on the census age distribution as well. Along these lines, then, we suggest the following procedure, which takes advantage of a relationship observed in model life table systems between the age distribution of deaths and the expectation of life at a given age.

Within each family (West, North, East, or South) of the Coale-Demeny model life tables, for example, there exists a one-to-one relationship between the ratio of adolescent and younger adult deaths (ages 10 to 40), ${}_{30}d_{10}$, to older adult deaths (ages 40 to 60), ${}_{20}d_{40}$, and the life expectancy at any age x, for $x = 60, \ldots, 95$ (Coale and Demeny, 1983). Because of this correspondence, it is useful to convert the age distribution of registered deaths into the life table death distribution. This can be accomplished by noting that

$$d(a) = \frac{1}{k \cdot B} D(a)\exp\left[\int_0^a r(x)dx\right], \qquad (7)$$

where $d(a)$ is the number of life table deaths at age a, k is the completeness of death registration, and B is the annual number of births in the population. The discrete analogue to equation (7) is

$${}_5d_a \simeq \frac{1}{k \cdot B}\, {}_5D_a\exp\left[5\sum_{x=0}^{a-5} {}_5r_x + 2.5\,{}_5r_a\right].$$

When we sum the values of ${}_5d_a$ to form the ratio ${}_{30}d_{10}/{}_{20}d_{40}$, it is not necessary to know k and B since they appear in both the numerator and denominator and thus cancel one another.

Once we compute that ratio, we refer to the appropriate family of model life tables for the corresponding $e(x)$ value, which may be obtained by interpolation. Within the Coale-Demeny system, we suggest that model West be chosen in the absence of any information which would point toward the use of another family among the West, East, and South life tables. The relationships found in these three families of life tables are nearly identical to one another, but differ from the relationship incorporated within the North model life tables. It should be emphasized, however, that if one is limiting the estimation of $e(x)$ to ages above 75 or so, then the impact of an incorrect choice of $e(x)$ (for example, due to an inappropriate choice of family) will be minimal in the estimation of life expectancy at the very young ages. For users of the present method, we provide in Table 1 the ratios of ${}_{30}d_{10}/{}_{20}d_{40}$ and the corresponding values of $e(75)$ through $e(95)$ which are associated with the Coale-Demeny West model life tables (Second edition) for males and females at many different levels of mortality. Level 3 corresponds to $e(0)$'s of 22.852 and 25.000 years for males and females, respectively, and level 25, to $e(0)$'s of 76.647 and 80.000.[4]

Using equations (5) and (6) we can generate all values of $N(x)$, for $x = 0, 5, \ldots, A-5$, and A. After computing these values, it is a simple matter to derive five-year survival probabilities, by using

$${}_5p_a \simeq \frac{\hat{N}(a+5)}{\hat{N}(a)} \exp[5\,{}_5r_a],$$

which is the five-year discrete version of equation (3). The other life table functions are derived from the sequence of ${}_5p_a$'s under certain assumptions about the distribution of l_a's within five-year age groups.[5]

SENSITIVITY ANALYSIS

In developing countries and, indeed, in many developed countries, death registration and census data are subject to several types of errors. These errors violate various assumptions underlying

Table 1.—The Ratio, ${}_{30}d_{10}/{}_{20}d_{40}$, and Corresponding $e(x)$ Values (x = 75,. . .,95) Associated with Many Levels of Mortality in the Coale-Demeny West Model Life Tables, Males and Females.

	Males						Females					
Level	Ratio	e(75)	e(80)	e(85)	e(90)	e(95)	Ratio	e(75)	e(80)	e(85)	e(90)	e(95)
3	1.161	4.55	3.35	2.41	1.71	1.19	1.376	4.88	3.57	2.54	1.78	1.23
4	1.094	4.75	3.49	2.51	1.77	1.24	1.300	5.11	3.73	2.65	1.86	1.28
5	1.034	4.95	3.63	2.60	1.83	1.28	1.233	5.33	3.89	2.76	1.93	1.32
6	.980	5.14	3.77	2.70	1.90	1.31	1.171	5.54	4.04	2.87	1.99	1.36
7	.930	5.32	3.90	2.79	1.95	1.35	1.115	5.75	4.19	2.97	2.06	1.40
8	.885	5.49	4.03	2.87	2.01	1.39	1.062	5.95	4.33	3.07	2.12	1.44
9	.842	5.67	4.15	2.96	2.07	1.42	1.012	6.14	4.47	3.16	2.19	1.48
10	.802	5.83	4.27	3.04	2.12	1.46	.964	6.33	4.61	3.26	2.25	1.52
11	.763	5.99	4.38	3.12	2.18	1.49	.918	6.52	4.74	3.35	2.31	1.56
12	.725	6.15	4.50	3.20	2.23	1.52	.872	6.70	4.88	3.44	2.37	1.60
13	.689	6.30	4.61	3.28	2.28	1.55	.827	6.88	5.00	3.53	2.43	1.63
14	.648	6.43	4.70	3.34	2.32	1.58	.787	7.02	5.11	3.60	2.47	1.66
15	.609	6.55	4.79	3.40	2.36	1.61	.729	7.16	5.21	3.67	2.52	1.69
16	.570	6.68	4.88	3.47	2.40	1.63	.673	7.32	5.33	3.75	2.57	1.72
17	.530	6.81	4.98	3.53	2.45	1.66	.617	7.48	5.44	3.83	2.63	1.75
18	.490	6.95	5.09	3.61	2.50	1.69	.560	7.65	5.57	3.92	2.68	1.79
19	.447	7.11	5.20	3.68	2.55	1.72	.501	7.83	5.70	4.01	2.74	1.82
20	.401	7.26	5.31	3.77	2.60	1.76	.438	8.01	5.84	4.11	2.80	1.86
21	.352	7.43	5.44	3.85	2.66	1.79	.365	8.22	5.99	4.21	2.87	1.90
22	.305	7.70	5.63	3.99	2.75	1.85	.298	8.54	6.22	4.38	2.98	1.97
23	.255	8.03	5.88	4.16	2.86	1.92	.235	8.94	6.52	4.59	3.12	2.05
24	.202	8.48	6.21	4.40	3.02	2.01	.175	9.46	6.91	4.86	3.30	2.16
25	.147	9.08	6.66	4.71	3.23	2.14	.117	10.17	7.45	5.24	3.54	2.31

our procedure. We will formally consider four common errors found in the data and discuss how these errors influence the estimated values of $e(x)$.

1. *Error in Estimated Growth Rates*

If age-specific growth rates computed from two successive censuses are used, then the resulting mortality estimates might be biased due to differential completeness of coverage between the two censuses. If the proportional differences in coverage completeness are invariant to age, then all age-specific growth rates are biased by the same amount,

$$\Delta r = \frac{\ln(c_2/c_1)}{t},$$

where c_1 and c_2 are the completeness of the first and second censuses, respectively, and t is the length of the intercensal interval. It should be noted that the amount of bias introduced to the age-specific rates under these circumstances is the same for all ages, even if the completeness of census enumeration varies with age. Bias in growth rates is age-dependent only when the coverage of persons in one age group has improved or worsened relative to that in another.

Differentiating equation (1) with respect to r, we have

$$\frac{de(x)}{dr} = e(x)[2A_x^s - e(x)],$$

where A_x^s is the mean age of persons above age x, measured with x as the origin, in the stationary population that corresponds to the life table produced by the true mortality conditions. By rearranging terms, we see that for Δr close to zero,

$$\frac{\Delta e(x)}{e(x)} \simeq \Delta r[2A_x^s - e(x)]. \qquad (8)$$

The proportionate error in expectation of life at age x, then, is proportional to the error in the rates of growth. The factor by which the error in $e(x)$ is proportional to Δr is presented in Table 2 for three different mortality regimes.

In Table 2, we also present the proportionate error in $e(x)$ due to errors in growth rates derived by Preston and Bennett (1983) for their census-based method of mortality estimation. Their method is analogous to the present one in that they use age-specific growth rates

Table 2.—$\frac{d \log \hat{e}_x}{dr}$ Using Coale-Demeny West Model Life Tables for Females.

Age x	Present Method	Preston-Bennett
	Level 9 (e_0=40 years)	
0	23.28	31.64
5	9.90	29.85
10	9.21	27.84
20	8.66	23.89
30	7.50	20.02
40	6.30	16.19
50	5.55	12.48
	Level 15 (e_0=55 years)	
0	14.98	34.99
5	7.00	32.88
10	6.69	30.68
20	6.51	26.34
30	5.89	22.07
40	5.34	17.89
50	5.02	13.86
	Level 21 (e_0=70 years)	
0	5.96	37.98
5	3.48	35.64
10	3.54	33.28
20	3.75	28.59
30	3.84	23.98
40	4.02	19.48
50	4.20	15.16

to convert the observed age structure of the population to that of the stationary population corresponding to the life table representing the mortality conditions of the population under study. In the present method, we convert the observed age structure of deaths to that of the stationary population. Note that the Preston-Bennett method is based on the following relationship:

$$e(x) = \frac{\int_x^\infty N(a)\exp\left[\left[\int_x^a r(u)du\right]da\right]}{N(x)}.$$

The proportionate error in $e(x)$ due to an error in the growth rate is given by

$$\frac{\Delta e_x}{e_x} = \Delta r \cdot A_x^s$$

for small Δr. The multiplier A_x^s is presented in Table 2.

The multiplier values in Table 2 reveal that the present method is significantly less sensitive to error in the growth rate than is the method by Preston and Bennett. At the extreme, the estimate of $e(5)$ under low mortality conditions is more than ten times as sensitive to an error in the growth rate using the census-based, rather than the death registration-based, method.

It is also clear from Table 2 that proportionate errors tend to be larger under higher mortality conditions. For female life expectancies of 40 and 55 years, the proportionate error in $e(x)$ is inversely related with age between the ages of five and 50. However, the reverse is true in low mortality populations ($e(0) = 70$).

One last item worth noting is that the proportionate error in $e(0)$ is substantially larger than that in $e(x)$ for ages five through 50. This is one reason among several that it is better to restrict use of the present method to ages five and above, leaving mortality under that age to be estimated using a different procedure.

2. *Net Intercensal International Migration*

Preston and Coale (1982) have shown how equation (1) can be generalized to accommodate the concept of an open population. However, the data necessary to implement this generalization, age-specific net migration rates, are rarely available. Thus, in the absence of information regarding these rates, we are forced to assume that the population is closed, or more accurately, that at each age in- and out-migration exactly offset each other. Violation of this assumption introduces biases in the estimated level of mortality. Since various age patterns of migration can be considered, we focus on two extreme types of migration.

(a) *Age-specific rates of net migration are equal.* Suppose each age group receives a net inflow of migrants during the intercensal period in proportion to its average size over the period. We denote this age-independent net migration rate, M. The life table death function for the observed population is then given by

$$d(a) = \frac{D(a)}{B \cdot k}\exp\left[-\int_0^a \rho(x)dx\right],$$

where $\rho(x) = r(x) - M$. However, if net migration is neglected and one simply uses equation (7), actually appropriate only for closed populations, then the situation is formally analogous to the preceding one in which all growth rates are distorted by the same amount of error, due to improved census coverage. We see, then, that the proportionate error in life expectancy at age x will be

$$\frac{\Delta e(x)}{e(x)} \simeq M[2A_x{}^s - e(x)].$$

With an annual rate of net inmigration of 1 per 1,000, life expectancy estimates will be too high by less than 1 percent at ages between five and 50, as implied by Table 2.

(b) *Net inmigration occurs at only one age, z.* If we estimate life expectancy for an age x younger than the age of migration z, then for small M_z, where M_z is the annual rate of net inmigration at age z, the effect of migration on estimated life expectancy is

$$\Delta e(x) \simeq M_z \cdot {}_{z-x}p_x \cdot {}_{z-x}\phi_x,$$

where ${}_{z-x}\phi_x = [e(z) + z] - [e(x) + x]$. As x recedes from z, ${}_{z-x}\phi_x$ increases more rapidly than ${}_{z-x}p_x$ declines, so that the error increases. Table 3 presents the values of ${}_{z-x}p_x$ and ${}_{z-x}\phi_x$ at various mortality levels for $z = 25$, an age which is close to the age of heaviest migration in a number of populations (see, e.g., Rogers and Castro, 1981). The table shows that the amount of error increases substantially when we proceed from age five to age zero, suggesting again that the use of the present method should be limited to ages five and above. If x is above z, then estimates of life expectancy at age x are unaffected by migration at age z.

Table 3.—Values of ${}_{25-x}p_x \times {}_{25-x}\phi_x$ in Coale-Demeny West Model Life Tables for Females.

Age x	${}_{25-x}p_x$ (1)	${}_{25-x}\phi_x$ (2)	(1)×(2) (3)
		Level 9 (e_0=40 years)	
0	.629	20.81	13.08
5	.867	6.01	5.21
10	.898	4.34	3.89
15	.922	3.14	2.90
20	.956	1.69	1.61
		Level 15 (e_0=55 years)	
0	.803	12.19	9.79
5	.933	3.43	3.20
10	.948	2.52	2.39
15	.960	1.87	1.79
20	.977	1.02	1.00
		Level 21 (e_0=70 years)	
0	.945	3.76	3.55
5	.983	0.96	0.95
10	.987	0.74	0.73
15	.989	0.57	0.56
20	.994	0.32	0.32

3. *Age-Dependent Completeness of Death Registration*

Completeness of death registration may vary with age. Deaths in early childhood tend to be underregistered to a greater extent than those at older ages. The likelihood that a death will go unregistered is especially great in infancy when the death occurs prior to the registration of the birth. For this reason, estimates of $e(x)$ for ages five and above will be more reliable than that of $e(0)$. Among adults, deaths of older persons, especially widows and widowers living alone, are more likely to be missed.

Suppose a decline in the completeness of death registration begins at age z, and the age pattern of completeness follows an exponential trajectory, such that

$$k(x) = k(z) \exp[-g(x-z)] ,$$

where $k(x)$ is the completeness of death registration at age x, and g is the parameter of the exponential function. This decline of completeness with age has the same impact on one's estimates as does a reduction of the age-specific growth rates above age z by the constant amount g. Thus for age $x \geq z$, the proportionate error in life expectancy at age x will be:

$$\frac{\Delta e(x)}{e(x)} \simeq -g[2A_x^s - e(x)].$$

Suppose, for example, completeness declines from age 10 such that completeness at age 60 is 80 percent that at age 10. In this instance, g is equal to .0045 and from Table 2 we see that the proportionate error in $e(x)$ (for $x \geq 10$) is about 2 to 3 percent given a true $e(0)$ of 55 years.

4. *Misreporting of Age at Death and Age in the Population*

Age-misreporting affects the mortality estimates produced by the present method in two ways. Misstatement of age at death distorts the age distribution of deaths (the $D(a)$'s), and inaccurate reporting of ages of the enumerated population may introduce errors in the series of age–specific growth rates (the $r(a)$'s).

In general, although random errors in reported ages tend to cancel one another, systematic over– or understatement of ages will result in biased estimates of mortality. For example, if the ages of all living persons and decedents are overstated by y years, then the entire life table for the population is shifted y years. Life expectancy is too high by:

$$\Delta e(x) = e(x-y) - e(x) .$$

It should be noted that distortions in the $D(a)$'s and $r(a)$'s due to age misreporting differ in their impact on mortality estimates. We first consider errors in age–specific growth rates stemming from age misstatement. If the rates are obtained from two successive censuses, the misreporting of ages of enumerated persons will generally introduce errors in these rates. Using the inaccurate age–specific growth rates, the age structure of deaths in the study population is incorrectly converted into that of the life table that is considered to represent the mortality conditions prevailing in the population. If growth rates are constant across age, as in a stable population, age misreporting will not result in defective rates of growth. This proposition holds as long as (a) age misreporting is independent of the three components of growth, i.e., mortality, fertility, and migration, and (b) patterns of age misreporting remain unchanged between the two censuses.[6]

If growth rates vary with age, then age misreporting leads to inaccurate estimates of these rates. The reason for the resulting bias in growth rates rests on the fact that if the group of persons reportedly age a consists of persons of different ages (including age a itself), then the observed growth rate at age a is approximately equal to the weighted average of true growth rates at the constituent ages. The weight given to the growth rate at each age x is equivalent to the proportion of those reportedly age a who are actually age x.

Suppose, for example, that growth rates for age groups 65–69 and 70–74 are .025 and .030, respectively, and, further, that among those reportedly age 70–74, only 80 percent are truly that age and 20 percent are in fact age 65–69. The observed growth rate, then, for age group 70–74 is the weighted average of .030 and .025, which in this case is about .029. If all age-specific growth rates are underestimated by .001, as in this example, then life expectancy at age x (for $x \geq 5$) is underestimated by no more than one percent.

Empirically, we find that age misreporting tends to result in the underestimation of age-specific growth rates, as was the case in the above example. In most populations there is more overstatement of age than understatement, particularly among old age groups. Consequently, the growth rate at a given age tends to be confounded with age-specific growth rates at younger ages. It can be shown that reductions in fertility and mortality tend to result in growth rates at the younger ages that are low in comparison with those at the higher ages. In the presence of net age overstatement, then, estimated age-specific rates of growth are lower than the true values.

In general, the size of the errors in estimated life expectancy due to age misstatement will most likely not be too large. The transfer of persons by age misstatement predominantly occurs between neighboring age groups, and not between the very young and very old groups. Furthermore, insofar as changes in past vital rates have been gradual, differences in growth rates among age groups tends to be smaller if the groups lie relatively close together. Thus, even though age misstatement may exist, the implicit weighting process ensures that the estimated growth rates will not differ too much from the true, underlying values (albeit under these "well-behaved" conditions).

Now we turn to the effect of misreporting of age at death on estimated mortality. It can be shown that if ($g \times 100$) percent of deaths at age z are reported to occur at age y, where $y > z$, then the expectation of life at age x ($x < z < y$) is too high by

$$\Delta e_x = \frac{\lambda(z, y)}{1 + \lambda(z, y)} \cdot \left\{ (y - x) + \frac{(y - z)}{\psi(z, y)} - e_x \right\},$$

where $\lambda(z,y) = g \cdot {}_{z-x}p_x \cdot \mu(z) \cdot \psi(z,y)$, ${}_{z-x}p_x$ is the proportion of those living at age x who survive to age z, $\mu(z)$ is the age-specific death rate at exact age z, and

$$\psi(z,y) = \exp\left[\int_z^y r(u)du\right] - 1.$$

Suppose, for example, that 10 percent of the population aged 60–64 classify themselves as 70–74, that age-specific growth rates are constant at .025 for ages between 60 and 75, and that the population is subject to the Coale–Demeny model West life table for females at level 13 (corresponding to $e(0) = 50$ years); the error in estimated life expectancy at age 10 in such a case is a mere 0.114 year. If 10 percent of *all* ages at death above age 60 are reported ten years too old, then the error increases to 1.04 years. Although the effect is still rather small, it is significantly larger than the corresponding error in life expectancy estimated from the population age structure using the census-based method by Preston and Bennett. If 10 percent of the population above age 60 reported themselves to be 10 years older than they actually were, then the estimate of $e(10)$ would be 0.266 year too high.

Because deaths are more heavily concentrated at the older ages than is the living population, age misreporting at these ages will have more serious effects on estimated mortality when the method of estimation is based on the age distribution of deaths. Therefore, although the present method is more robust than the census-based method of Preston and Bennett in the presence of migration and differential census coverage, the census-based method is less sensitive to age misreporting.[7]

APPLICATION

In order to illustrate the use of this method, we apply it to registration and census data for Argentine females covering the period 1960 to 1970. Argentina is believed to have better quality data than most other Latin American countries, although these data are still subject to a considerable amount of error. Inconsistencies in the data become evident when we attempt to evaluate the completeness of death registration in the population.

Registration and census data are available for five–year age groups up to age 85, plus the open-ended interval, 85 and above. To be able to determine the completeness, we first need an estimate of $e(85)$. Rather than use the Coale–Demeny model life tables to obtain this estimate, we instead use the Latin American model developed by the United Nations (1982b). Later, we will see whether this model is in fact an appropriate choice. Enumeration of 55–59, 60–64, and 65–69 year olds appears to be inaccurate in many Latin American countries due to age misstatement. Thus, instead of using ${}_{30}d_{10}/{}_{20}d_{40}$ to infer $e(85)$, we use ${}_{30}d_{10}/{}_{15}d_{40}$, on the assumption that the number of deaths in the age group 55–59 might be suspect. The ratio equals .636, giving us an $e(85)$ of 5.49 years.

Table 4 presents the raw data—the two census age distributions (September 30, 1960 and 1970) and the intervening ten-year sum of deaths[8]—and estimates of the completeness of death registration in the form of the two series of ratios, ${}_5\hat{N}_x/{}_5N_x$ and ${}_{85-x}\hat{N}_x/{}_{85-x}N_x$. ${}_5\hat{N}_x$ and ${}_{85-x}\hat{N}_x$ refer to the estimated number of persons aged x through $x + 4$ and x through 84, respectively. ${}_5N_x$ and ${}_{85-x}N_x$

Table 4.—Age Distribution of Argentine Female Population on September 30, 1960 and 1970, Age Distribution of Intercensal Female Deaths, and $\hat{N}/N$ Ratios Using 85 and Above for Open Interval.

Age Group (x, x+4)	Population 9-30-60	Population 9-30-70	Intercensal Deaths	${}_5\hat{N}_x / {}_5N_x$	${}_{85-x}\hat{N}_x / {}_{85-x}N_x$
0	1,054,603	1,158,350	159,124	---	---
5	1,029,209	1,133,950	7,259	1.071	1.037
10	965,393	1,086,850	5,558	1.067	1.033
15	854,136	1,039,850	9,183	1.068	1.028
20	778,130	980,550	10,729	1.030	1.023
25	775,842	860,150	11,864	1.007	1.022
30	789,746	795,650	14,542	1.001	1.024
35	724,175	767,400	18,379	1.036	1.029
40	611,018	769,600	21,664	1.034	1.027
45	590,405	698,950	27,173	.981	1.025
50	499,239	584,800	35,955	1.046	1.039
55	414,264	549,250	46,529	1.021	1.036
60	326,719	454,750	58,863	1.019	1.043
65	236,487	350,450	69,388	1.036	1.056
70	172,717	244,200	79,734	1.042	1.070
75	99,937	156,550	82,508	1.106	1.100
80	50,570	89,400	71,083	1.089	---
85	32,052	52,350	73,952	---	---

refer to the corresponding observed values given by the censuses. In Figure 1 we plot the two series so that we may more easily discern errors in the data. If the completeness of death registration and census enumeration were invariant to age, the age-specific growth rates were correct, and no age misstatement occurred, then the $\hat{N}/N$ ratios would form a flat line, the level of which would equal the proportion of deaths recorded relative to the completeness of census enumeration. The overall pattern in the cumulated $\hat{N}/N$ ratios (${}_{85-x}\hat{N}_x/{}_{85-x}N_x$) reveals the existence of overstatement of age at death. The sharp upswing in these ratios toward the end of the age distribution indicates that too many deaths are being placed in the oldest age groups, thus spuriously giving the impression that completeness of death registration improves rapidly with age above age 60 or so. Theoretically, of course, this improvement could be a real phenomenon, but such would be extremely implausible.

To circumvent the problem of overstatement of age at death in the very old ages, we collapse several age groups at the end of the age distribution and form a new open interval, one within which presumably most of the age overstatement occurs. In the present case, we create the open interval, 75 and above, and thereby assume that little systematic overstatement of age occurs below age 75. The inferred value for $e(75)$ is 9.68 years.

In Table 5 we present the two sets of ratios, ${}_5\hat{N}_x/N_x$ and ${}_{75-x}\hat{N}_x/{}_{75-x}N_x$ and plot the series in Figure 2. The series of cumulated $\hat{N}/N$ ratios has flattened out considerably. The median of this series for $x = 5, 10, \ldots, 65$ is 1.032, indicating that deaths are actually overregistered by 3 to 4 percent relative to the completeness of census enumeration.

From the $\hat{N}(x)$'s, we estimate the ex-

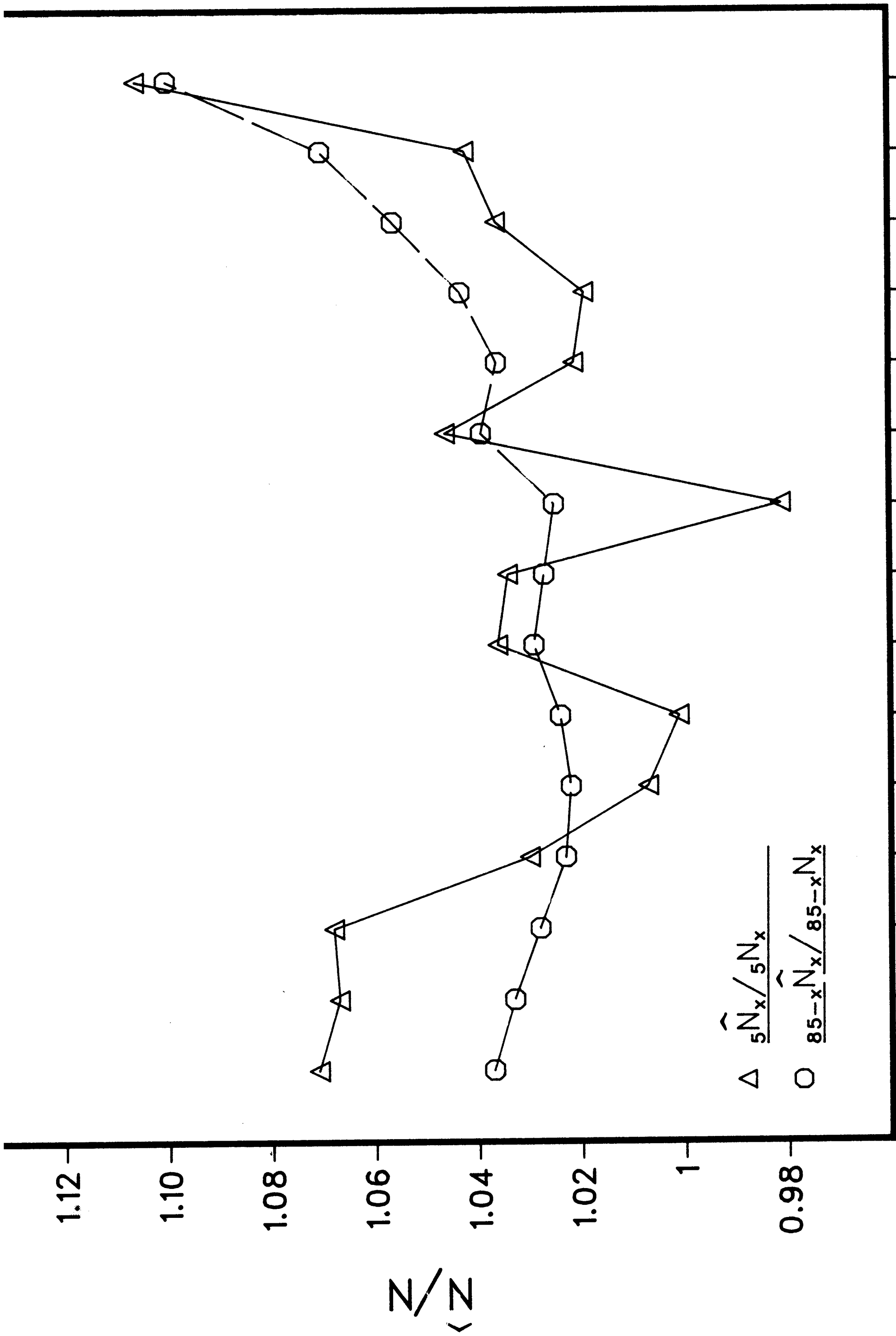

Figure 1.—$\hat{N}/N$ Ratios Derived Using 85 and Above for Open Interval.

Table 5.—$\hat{N}/N$ Ratios for Argentine Females Using Age 75 and Above for Open Interval.

Age Group (x, x+4)	${}_5\hat{N}_x / {}_5N_x$	${}_{75-x}\hat{N}_x / {}_{75-x}N_x$
5	1.077	1.042
10	1.073	1.037
15	1.074	1.032
20	1.036	1.026
25	1.012	1.025
30	1.006	1.027
35	1.042	1.031
40	1.040	1.029
45	.986	1.026
50	1.052	1.039
55	1.027	1.034
60	1.026	1.037
65	1.043	1.046
70	1.050	---

pectation of life at ages 5 through 75, shown in Table 6. Because it is likely that various assumptions implicit in the procedure may not be strictly valid, for example that registration completeness is constant across age, we choose to fit a model life table to our estimated $e(x)$ values. Using the Latin American tables for females, we infer the model $e(0)$ that is associated with each estimated $e(x)$. The array of model $e(0)$'s is also presented in Table 6. The striking result is that our derived life table for Argentine females nearly perfectly follows the Latin American model pattern of mortality.[9] The $e(0)$'s implied by the estimated $e(x)$'s for ages 5 through 65 vary only insignificantly, within a range of .3 year. We should note that the dramatically close fit justifies our earlier decision to use the Latin American model to obtain the estimate of $e(A)$, the expectation of life at the lower bound of the open interval.

Taking the median of the model $e(0)$'s derived from the $e(x)$'s for ages 10 through 55 (the ages probably least subject to error), we interpolate from the United Nations model life tables the entire life table associated with $e(0)$ equal to 69.36 years. We present life expectancy at age x for $x = 0, 1, 5, 10, \ldots, 85$ in Table 7. Adopting these values of $e(0)$ and $e(1)$ assumes, of course, that the Latin American pattern derived from observed $e(x)$'s for ages 5 and above holds for ages below 5 as well. Ideally, we would want further information regarding infant and child mortality before we make such a claim.

Our estimate of $e(5)$, 69.4 years, is slightly higher than that given by the United Nations (1982a), approximately 69.3 years. This slight difference is easily explained by our finding that the straightforward use of registered deaths and census data to form age-specific death rates will lead to a small upward bias in these rates. Our results indicate that we must deflate the observed number of deaths to bring that number in line with the actual extent of completeness of census enumeration. The lower death rates, then, result in slightly higher life expectancies.

DISCUSSION

We have derived a method from a generalization of stable population theory that enables one to estimate an accurate life table using incomplete death registration data. It is not necessary, however, to know a priori the level of completeness.

Forward and backward projection techniques have been most widely used for the construction of a life table for an intercensal period when deaths in the period are likely to be significantly underregistered or not reported at all (see, e.g., Coale and Demeny, 1967, and Palloni and Kominski, 1981). The major advantages of the present method over these standard intercensal survival techniques are the following:

First, the present method is much simpler to use, especially when one is ana-

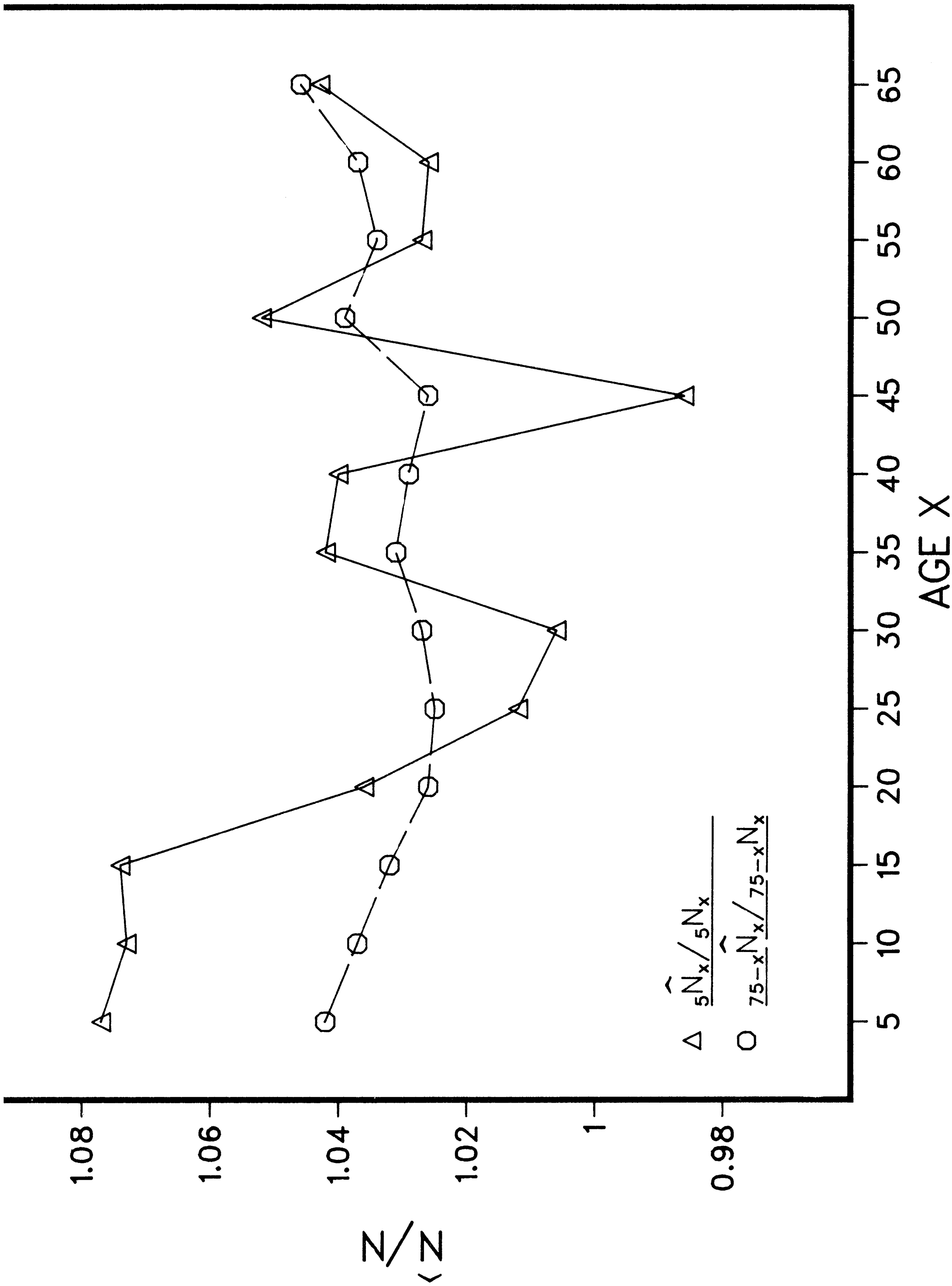

Figure 2.—$\hat{N}/N$ Ratios Derived Using 75 and Above for Open Interval.

Table 6.—Expectations of Life Estimated Using the Present Method and the Corresponding Implied e(0)s from the Latin American Model Life Tables.

Age x	Estimated e(x)	Implied Model e(0)
5	69.52	69.53
10	64.73	69.28
15	59.89	69.24
20	55.15	69.31
25	50.46	69.36
30	45.81	69.39
35	41.21	69.42
40	36.67	69.41
45	32.19	69.35
50	27.83	69.32
55	23.64	69.36
60	19.66	69.38
65	15.98	69.43
70	12.63	68.99
75	9.68	68.01

ness of enumeration between the two censuses from age patterns of $\hat{N}(x)/N(x)$ (Preston et al., 1980, Bennett and Horiuchi, 1981, and Hill et al., 1983).

There are, however, two drawbacks to the use of this method. First, we cannot safely obtain an estimate of the expectation of life at birth using this method since it is likely that deaths under age five are recorded to a lesser extent than those above age five. If we were to estimate $e(0)$, then, the assumption of constancy of completeness in death registration across age might be an incorrect one. Further, estimates of $e(0)$ are considerably more sensitive to violations in the other assumptions underlying this method than are estimates of $e(x)$ at higher ages. It is very common, however, that a country will have reasonably good estimates of infant and child mortality by virtue of Brass questions on

lyzing data referring to an intercensal interval of nonstandard length (i.e., not an integer multiple of five years). For example, if one were studying mortality during a four-year period between censuses (e.g., as with the Korean censuses of 1966 and 1970), one could by using the present method avoid reclassification of the data into unorthodox four-year age intervals. Such reclassification would require either single-year age data or putting forth some assumption about the single-year age distribution within each five-year age group.

Second, one need not rely on model life tables in order to derive estimates of $e(x)$, except perhaps in the case of $e(A)$, where A is the lower bound of the open-ended interval.

Third, the $\hat{N}(x)$ series, a by-product of this method, is very useful for analyzing the data for violations of the assumptions underlying the method. Previous work has shown how to diagnose and deal effectively with possible overstatement of age at death and differential complete-

Table 7.—Final Estimates of Expectations of Life for Argentine Females during the Period 1960 to 1970 Interpolated from Latin American Model Life Tables ($e(0)$ = 69.36 years).

Age x	Model e(x)
0	69.36
1	71.87
5	69.41
10	64.78
15	59.96
20	55.18
25	50.46
30	45.79
35	41.18
40	36.64
45	32.19
50	27.85
55	23.64
60	19.65
65	15.96
70	12.71
75	9.90
80	7.45
85	5.59

numbers of children ever born and surviving. It would thus be possible to splice the life table under five from this source with that above five obtained using the present method.

Second, since deaths are concentrated in the upper ages, it is important how we obtain $e(A)$. When A is above 75 or so, the estimate of $e(A)$ usually will not affect significantly the estimate of $e(5)$. However, should we have age groups up to only age 60 and over, and therefore have to begin the procedure with an estimate of $e(60)$, then the choice of $e(60)$ must be made carefully. In fact, the safest approach would be to estimate a range of $e(5)$ within which the true value would most likely fall, by inputting the limiting values of $e(60)$ (within which the true value of $e(60)$ would most likely fall).

Finally, a word may be needed about the comparative usefulness of this method and the other previously mentioned techniques of estimating mortality using age-specific growth rates. The method by Preston and Bennett (1983) for constructing a life table from intercensal person-years lived, and another technique by Preston (1983) based on the logit life table system, seem adequate when data from two successive censuses are available but registration of deaths during the intercensal period is virtually nonexistent. When deaths are registered but their completeness is in doubt, we recommend that the present method and the procedure by Bennett and Horiuchi (1981) for estimating the extent of underregistration be integrated as two components of one data-processing package. Both procedures require the calculation of $\hat{N}(x)$'s, and once the sequence of $\hat{N}(x)$'s is obtained, then both the estimation of the completeness of death registration and the construction of a life table from the sequence are fairly simple matters, logically consistent with each other.

NOTES

[1] This method was originally developed and applied in Bennett (1981).

[2] An alternative way to construct life tables is implied by equation (7), which converts the number of registered deaths by age into the age distribution of deaths in the stationary population using age-specific growth rates. A proportionality factor, $k \cdot B$, in equation (7) is obtained such that

$$\int_0^\infty d(a)da = 1.$$

The remainder of the life table can be derived from the sequence of $d(a)$'s. Although the starting age of the life table is set to zero in equation (7), this age can be shifted easily to any higher one.

Given the same set of data, equations (7) and (3) lead to exactly the same life table. They are essentially alternative derivations of the same method. Although equation (7) may represent the basic idea of the method more clearly, the use of $\hat{N}(a)$'s is in practice preferable, since the $\hat{N}(a)$'s are useful byproducts for diagnosing violations of basic assumptions of the method.

[3] For the derivation of equation (6), see note (3) of Bennett and Horiuchi (1981).

[4] An alternative method of obtaining $e(A)$ is suggested by Preston and Bennett (1983). They note that if the age-specific growth rates are constant for all ages above A, then

$$e(A) = \frac{\int_0^\infty N(A + y)\exp[yr(A+)]dy}{N(A)},$$

which reduces to

$$e(A) = \exp[\bar{y}r(A+)] \cdot \frac{N(A+)}{N(A)},$$

where $\bar{y}$ is defined such that the equality holds. The following equation is suggested to solve for $\bar{y}$ and is based on regression analysis of simulated data:

$$\bar{y} = e(A)[0.802 - 0.0106e(A) - 1.34r(A+)].$$

The solution for $e(A)$ is then obtained implicitly by numerical approximation. Thus $e(A)$ may be derived in this manner, given that the numbers of persons above age A, $N(A+)$, and at age A, $N(A)$, (the latter estimated by means shown in Preston and Bennett [1983]) are available.

The major problems with this approach are: (a) it requires an additional source of data, and (b) the estimates of $e(x)$ would be subject to errors in the estimation of $N(A+)$ and $N(A)$. The former problem is virtually trivial. Although we may never use the population age distribution explicitly in the procedure described in the text, we will almost always require two census age distributions in order to obtain the age-specific rates of growth. Furthermore, it should be noted that if one is forced to estimate $e(A)$ at an age as low as 60, for example, and if in addition one is uncertain about the appropriateness of one family versus another in the model life table system, then we might in fact wish to use this alternative approach rather than a method which would rely on the use of model life tables.

[5] To obtain $e(x)$ values, we must first derive the series of ${}_5L_a$'s, the person-years lived in each age category. It is straightforward to compute the l_a's from the five-year survival probabilities (beginning with a radix of one). Through a equal to age 45, we employ the formula ${}_5L_a = 2.5(l_a + l_{a+5})$. For ages 50 and above we assume that age-specific mortality rates are growing exponentially, and thus follow a Gompertz mortality schedule. We then have

$$ {}_{x\text{-}a}\hat{P}_a = \exp\{[\hat{\mu}(a)\text{-}\hat{\mu}(x)]/\xi\}, \qquad (9)$$

where ξ, which we set equal to .10, is the natural log of one plus the proportion increase in mortality with each year of age, and $\hat{\mu}(a)$ and $\hat{\mu}(x)$, the estimated death rates at exact ages a and x, respectively, are given by

$$\hat{\mu}(a) = \frac{\ln[\hat{N}(a+5)/\hat{N}(a)] + 5\,{}_5r_a}{\{1 - \exp[5\xi]\}/\xi}$$

and

$$\hat{\mu}(x) = \hat{\mu}(a)\exp[(x-a)\xi].$$

Equations (7) and (9) allow us to compute the single-year l_a's which we can then sum to obtain the ${}_5L_a$'s.

[6] Specifically, we assume that the proportion of those age a who report to be age x, for all ages, remains constant over the intercensal period.

[7] We might also note that given an acceptable (perhaps, model) life table we can correct an observed distribution of deaths for age misreporting. Inverting the approach suggested by equation (7), we can convert the *life table* death distribution using age-specific growth rates and thus generate an *expected* age distribution of deaths for the actual population.

[8] The age distribution of deaths for 1967 has not been published. We assume that distribution to be equal to the average of the 1966 and 1968 distributions.

[9] We should note that no life tables from Argentina were used as input in the creation of the U.N. model life tables for developing countries.

ACKNOWLEDGMENTS

The authors would like to thank Yat-Ming Siu for his most helpful assistance in computing. Much of this paper was written while Bennett was affiliated with The University of Michigan as a Research Associate at the Population Studies Center and Mellon Assistant Professor in the Department of Sociology.

REFERENCES

Bennett, Neil G. 1981. Estimation Techniques Derived from Structural Relations in Destabilized Populations. Unpublished Ph.D. dissertation, Princeton University. Ann Arbor, Michigan: University Microfilms.

——— and Shiro Horiuchi. 1981. Estimating the Completeness of Death Registration in a Closed Population. Population Index 47:207–221.

Coale, Ansley J. and Paul Demeny. 1967. Manual IV. Methods of Estimating Basic Demographic Measures from Incomplete Data. New York: United Nations.

——— and Paul Demeny. 1983. Regional Model Life Tables and Stable Populations. Second Edition. New York: Academic Press.

Hill, Kenneth, Hania Zlotnik, and James Trussell. 1983. Manual X. Indirect Techniques for Demographic Estimation. New York: United Nations.

Horiuchi, Shiro and Ansley J. Coale. 1982. A Simple Equation for Estimating the Expectation of Life at Old Ages. Population Studies 36:317–326.

Palloni, Alberto and Robert Kominski. 1981. Estimation of Adult Mortality Using Forward and Backward Projections. Center for Demography and Ecology, University of Wisconsin. Unpublished manuscript.

Preston, Samuel H. 1983. An Integrated System for Demographic Estimation from Two Censuses. Demography 20:213–226.

——— and Neil G. Bennett. 1983. A Census-based Method for Estimating Adult Mortality. Population Studies 37(1):91–104.

——— and Ansley J. Coale. 1982. Age Structure, Growth, Attrition, and Accession: A New Synthesis. Population Index 48:217–259.

———, Ansley J. Coale, James Trussell, and Maxine Weinstein. 1980. Estimating the Completeness of Reporting of Adult Deaths in Populations that are Approximately Stable. Population Index 46:179–202.

Rogers, Andrei and Luis J. Castro. 1981. Model Migration Schedules. Laxenburg, Austria: International Institute for Applied Systems Analysis.

United Nations. 1982a. Levels and Trends of Mortality since 1950. New York: United Nations.

———. 1982b. Model Life Tables for Developing Countries. New York: United Nations.

Vincent, Paul. 1951. La Mortalité des Vieillards. Population 6:182–204.