CURRENT ITEMS

ESTIMATING THE COMPLETENESS OF DEATH REGISTRATION IN A CLOSED POPULATION

Neil G. Bennett and Shiro Horiuchi*+

The assessment of the mortality level of a population is often based on information including the number of registered deaths. In many developing nations, however, deaths are underregistered by a significant margin, which in turn may lead to a biased estimate of the level of mortality.

Various methods have been developed in the past several years in order to correct death registration data for underreporting. The following five methods have been most useful. (1) Brass's sectional growth balance equation (Brass, 1975); (2) a method devised by Preston and his colleagues (Preston and Hill, 1980 and Preston, Coale, Trussell, and Weinstein, 1980); (3) the technique of forward projection; (4) an intercensal cohort survival method developed by Preston and Hill (1980) and elaborated by Brass (1979) and Trussell and Menken (1979); and (5) a modified growth balance method by Martin (1980).

Brass's sectional growth balance equation, which holds for stable populations, may be expressed as

$$N(a)/N(a+) \quad = \quad r + k[D(a+)/N(a+)] \;, \qquad (1)$$

where N(a) is the population age a, N(a+) is the population age a and above, D(a+) is the number of registered deaths to persons age a and above, r is the growth rate of the stable population, and k is the inverse of the completeness of death registration. In populations that are approximately stable, N(a)/N(a+) and D(a+)/N(a+) for different ages will lie on a straight line with a slope equal to k and an intercept equal to r.

Preston et al. have developed another stable method in which the age distribution of a population is estimated from the age distribution of deaths. The

---

*The work for this project was carried out while the authors were associated with the Office of Population Research at Princeton University. Neil G. Bennett is currently affiliated with the Population Studies Center, The University of Michigan, 1225 South University Avenue, Ann Arbor, MI 48109. Shiro Horiuchi is a member of the Population Division, United Nations, New York, NY 10017.

+An earlier version of this paper was presented at the Annual Meeting of the Population Association of America, Washington, D.C., March 26, 1981. The authors are indebted to Ansley Coale for his invaluable guidance throughout the development of this project. We also are grateful for the advice and encouragement offered by Samuel Preston, James Trussell, and Jane Menken. Kenneth Hill and Hania Zlotnick provided criticism which led to an improved treatment of the open interval. This research was supported in part under Grant No HD-11720 from the National Institutes of Health.

completeness of death registration is then given by the ratio of the estimated population to the observed population. We will later discuss this technique in greater detail, since the method proposed in the present paper is a logical extension of it

Forward projection uses data from two censuses and does not rely on the assumption of stability. In this method, a life table is constructed from registered deaths and person-years lived in the intercensal period On the basis of this life table, the population of the first census is projected to the time of the second census. The completeness of death registration for the adult population is estimated by $[P_1(i) - \hat{P}_2(i)]/[P_1(i) - P_2(i)]$, where $P_1(i)$ and $P_2(i)$ are the size of the open cohort i at the first and second censuses, and $\hat{P}_2(i)$ is the projected size of the cohort at the time of the second census. [1]

Forward projection requires prior correction for relative underenumeration between the two censuses. However, the technique developed by Preston and Hill (1980) is designed to estimate both the completeness of death registration and the relative underenumeration between two censuses.

The method is based on the following relationship:

$$P_1(i)/P_2(i) \quad = \quad [c_1/c_2] + c_1 k[D(i)/P_2(i)] \ , \qquad (2)$$

where c_1 and c_2 indicate the completeness of enumeration at the first and second censuses, D(i) is the number of registered deaths experienced by the cohort i during the intercensal period, and $P_1(i)$, $P_2(i)$, and k have been previously defined A regression analysis using equation (2) with data drawn from successive cohorts will result in estimates of c_1/c_2 and c_1k as the intercept and slope, respectively.

Martin (1980) has modified the Brass method to relax the assumption of stability. The modified growth balance equation, which holds for any closed population, is expressed as

$$k[D(a+)/N(a+)] \quad = \quad [N(a)/N(a+)] - r(a+)$$

where r(a+) is the growth rate of the population aged a and above. If a set of r(a+) is available or estimated in some way, the plot of D(a+)/N(a+) against [N(a)/N(a+)]-r(a+) will provide an estimate of death registration.

Two major characteristics are common to all of the above-mentioned methods. First, they provide estimates of the completeness of death registration relative to the completeness of census enumeration. This is not a serious limitation insofar as knowledge of the relative completeness of registration is sufficient to correct observed death rates. Second, all of these methods are predicated on the following assumptions: (1) the population under study is closed; (2) the completeness of death registration is constant over age; and (3) the ages of the living and the dead are accurately reported.

These techniques may be divided into two groups: those that require the assumption of stability (the first two) and those that do not (the last three). It should be noted that more information is necessary in order to implement those methods which do not assume stability. Such methods usually require the use of two censuses, whereas stable methods are applicable even when population data are available from only one census.

In this paper we focus on the stable method discussed in Preston et al. (1980) and modify it such that the assumption of stability is no longer necessary

Preston and his colleagues employ the following relationship obtaining in stable populations:

$$N(a) = \int_a^\infty D^*(x)\exp[r(x-a)]dx , \qquad (3)$$

where $D^*(x)$ is the true number of deaths experienced by persons aged x in the current population. In words, the number of persons at age a in a population is the sum over deaths at each age above a weighted by an exponential of the product of r and the difference between the age at death and age a. Note that $D^*(x)\exp[r(x-a)]$ is an estimate of the number of people currently aged a who will die at age x. This follows from the fact that, in a stable population, the number of deaths to people aged a in a given year is related to that number in the previous year by a factor of $\exp[r]$.

Therefore the method suggested in equation (3) is the period analogue of the method of extinct generations set forth by Vincent (1951), by which the number of persons aged a at a certain time in the past can be estimated by cumulating all deaths to persons aged a and above which have been experienced by that cohort. The obvious fundamental requirement is that the last of the cohort members has died. On the other hand, equation (3) provides us with a mechanism by which we can adjust the death distribution in the current population to approximate the "future" death distribution of the cohort currently aged a.

If the completeness of death registration is constant at age a and above, then

$$D^*(x) = kD(x), \qquad \text{for all } x \geq a , \qquad (4)$$

where $D(x)$ is the number of registered deaths to persons aged x, and $D^*(x)$ and k have been previously defined.

By substituting equation (4) into equation (3), we obtain

$$N(a) = k \int_a^\infty D(x)\exp[r(x-a)]dx$$

If we define

$$\hat{N}(a) = \int_a^\infty D(x)\exp[r(x-a)]dx ,$$

then the completeness of death registration can be estimated as $\hat{N}(a)/N(a)$, when the number of registered deaths by age, the number of living persons by age, and the growth rate of the population are provided. More robust measures of completeness have been suggested, such as that derived from cumulating $\hat{N}(a)$ and $N(a)$. Cumulation would tend to absorb some of the distortion resulting from age misreporting and differential registration and enumeration by age.

The formula which is the basis for computing the estimated age distribution is:

$$\hat{N}(a-5) = \hat{N}(a)\exp[5r] + {}_5D_{a-5}\exp[2.5r] \qquad (5)$$

where $\hat{N}(a)$ is the number of persons aged a estimated iteratively by this equation, and ${}_5D_{a-5}$ is the number of deaths occurring within the age group a-5 to a. This method (discussed extensively in Preston et al., 1980) is clearly appropriate for analysis of stable populations and is justified by equation (3). However, it is not robust in the context of destabilization. When a population deviates from stability, r is no longer a constant but rather varies with age. In such a case, the total population growth rate, r, is often a poor approximation of r(a), the growth rate of the population aged a.

In order to accommodate the concept of differential growth rates within a population we propose the following extension of equation (5):

$$\hat{N}(a-5) = \hat{N}(a)\exp[5\,{}_5r_{a-5}] + {}_5D_{a-5}\exp[2.5\,{}_5r_{a-5}] , \qquad (6)$$

where ${}_5r_{a-5}$ is the growth rate experienced by those in the age group a-5 to a Note that for a cohort, which otherwise may be interpreted as a stationary population, the estimated number of people at age a-5, $\hat{N}(a-5)$, is simply equal to the estimated number of people at age a, $\hat{N}(a)$, plus the number of deaths occurring in the cohort (or the stationary population age group) in the intervening period, ${}_5D_{a-5}$. That exponentials are attached to each of the last two terms in equation (6) is to account for growth in the population and the number of deaths over time. Equation (6) is justified by the following relationship which holds true for any closed population:

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

Proof of this relationship may be found in the Appendix. Thus, we have removed the assumption of stability over the entire population. Instead, we rely on the significantly less restrictive assumption that the observed number of persons in each five-year age interval is approximately equal to the corresponding number in a stable population inferred from the numbers of persons at ages a and a+5, and the observed age-specific growth rates.

After all values of $\hat{N}(a)$ are calculated, we can compute the values of ${}_5\hat{N}_a$, the estimated number of persons in the age group a to a+5, by using the following approximation formula:

$${}_5\hat{N}_a = 2.5[\hat{N}(a) + \hat{N}(a+5)] .$$

In the older portion of the age distribution, however, there is often a significant amount of curvature within each five-year age group. For computation of ${}_5\hat{N}_a$ above the age of 60, we suggest imposing a stable population curve over the five-year span and then determining the area under the curve accordingly. [2]

Estimates of completeness may be derived, for example, from the median of a series of ${}_{10}\hat{N}_{a-5}$ (the estimated number of people between ages a-5 and a+5) divided by the corresponding figures in the observed population.

The Open Interval

We have deferred discussion of the estimation procedure involving data from the highest age group, which has no upper bound, due to the relatively complex nature of this method. Although a treatment of the open interval is prescribed in Preston et al. (1980), it is not necessarily adequate in those cases in which the lower bound lies at a relatively low age. N(a) tends to be underestimated when a is low, since their formula for obtaining $\hat{N}(a)$ for the open interval omits terms, such as the variance of the age of death above age a, which are negligible when a is high but are potentially of significant magnitude when a is low. Therefore, it is necessary to develop a more accurate means by which we can estimate N(a).

From a relationship found to hold in populations that are stable above age a, we have

$$N(a) \doteq D(a+)\{\exp[r(a+)e(a)] - ([r(a+)e(a)]^2/6)\} \, . \qquad (7)$$ [3]

Given values of r(a+), the rate of growth in the open interval, and e(a), the expectation of life at the beginning of the open interval, we can then compute our estimate of N(a). Once we have $\hat{N}(a)$ we can proceed to find all other $\hat{N}(x)$'s, by iterating downwards using equation (6) The value of r(a+) comes from the data themselves, while e(a) must be obtained independently of the data. If only knowledge of the overall level of mortality is available, we suggest that one estimate e(a) from a model life table which is characterized by the appropriate level of e(0) or e(10). Although in some instances a value of e(a) may be somewhat arbitrary, the resulting estimates of completeness will not be significantly biased.

Given an approximate level of mortality, indicated by e(0) or e(10), the variation in possible e(a)'s is small.[4] Hence, the estimate of e(a) is not likely to differ much from the unknown, true value of e(a). We can see in equation (7) that a proportionate error in e(a) will result in a much smaller proportionate error in $\hat{N}(a)$. Moreover, the impact of error in e(a) is mitigated as we proceed downward in the age distribution. In fact, in a stable population, the proportionate error in $\hat{N}(x)$ is simply the product of the error in $\hat{N}(a)$ and the probability of survival from age x to age a.

Due to the fact that $\hat{N}(a+)$ is more sensitive to error in e(a) than is $\hat{N}(a)$, we choose to ignore estimation of the population in the open interval. Consequently, measurements of completeness cannot be based on values of $\hat{N}(x+)/N(x+)$, as suggested in Preston et al. (1980).

This treatment of the open interval assumes that there is neither age misstatement of deaths nor age misstatement of population across the lower bound. However, there is much evidence that, in many populations, ages of the living and the dead are less accurate among older persons, whose ages tend to be overstated Therefore, should the age referring to the lower bound be sufficiently high, the assumption of no age misstatement might very well be violated. We may overcome this difficulty by broadening the open interval to a point above which almost all age misstatement is believed to occur. For example, even if the original data available to us allow for age groups up to 85+, we may aggregate several groups and expand our open interval to 60+. To this new open interval, we may now apply equation (7).

Applications

Sweden, 1965-1970

In assessing the quality of this technique we apply it to data from a destabilized population in which deaths and population are known to be virtually completely recorded. Swedish males for the period 1965 to 1970 fulfill these criteria[5] Due to the accuracy of, and consistency between, Swedish population and death data, we would expect to be able to reproduce the age distribution of the population from the distribution of deaths. This would imply, too, that we would obtain a flat sequence of values of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$, which would equal approximately one.

After having determined the number of deaths and person-years lived of native Swedish males over the five-year period, we computed the age-specific rates of growth. The results of the application of this method are presented in Table 1.[6] All values of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ from the different age groups fall within one percent of 1.000, indicating, as expected, that registration is essentially complete.[7]

Had we assumed stability in the case of Sweden, our estimate of the completeness of death registration would have been drastically incorrect. Surely, this would be an unrealistic application of the stable method given in Preston et al. (1980). However, such an application gives a clear indication of possible biases introduced by assuming stability. Figure 1 presents the sequences of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ derived from the method using a constant r and age-specific r's. We see that the series of values using a constant r is much too low and extremely erratic, whereas that using age-specific r's shows little variation and lies close to one. A comparison of the observed age distribution of the population with that estimated by the two procedures is shown in Figure 2. The age distribution which was determined assuming stability bears little resemblance to that which has been observed. However, using the present method, we have almost perfectly reproduced the observed age distribution of the population from the age distribution of deaths and the intercensal age-specific rates of growth.

Table 1: Male deaths, population, and rates of growth by age and values of $\hat{N}(a)$, ${}_5\hat{N}_a$, and ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ for Sweden, 1965-1970.

Age	${}_5r_a$	${}_5D_a$	${}_5N_a$	$\hat{N}(a)$	${}_5\hat{N}_a$	${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$
0	.0136	5,177	1,443,446	301,368	1,444,915	---
5	.0051	612	1,373,964	276,598	1,363,911	.997
10	-.0060	468	1,372,847	268,966	1,364,109	.993
15	-.0305	1,339	1,499,523	276,677	1,493,876	.995
20	.0196	1,766	1,540,850	320,873	1,525,295	.993
25	.0388	1,496	1,314,538	289,245	1,315,186	.995
30	.0098	1,620	1,147,401	236,830	1,151,890	1.002
35	-.0166	2,213	1,155,364	223,926	1,162,312	1.005
40	-.0317	3,360	1,287,952	240,998	1,299,393	1.008
45	.0126	5,552	1,326,431	278,759	1,337,772	1.009
50	-.0118	8,523	1,289,043	256,350	1,298,680	1.008
55	.0070	13,415	1,250,230	263,122	1,260,013	1.008
60	.0198	21,983	1,090,525	240,883	1,096,192	1.007
65	.0212	26,287	874,707	197,235	874,991	1.003
70	.0223	32,077	645,060	152,501	646,152	1.001
75	.0125	34,538	428,835	106,084	430,034	1.002
80	.0214	30,746	235,291	66,168	236,294	1.003
85	.0152	19,155	94,010	30,317	94,500	1.005
90	.0449	7,124	22,630	9,656	22,767	1.005
95	.0709	1,186	2,791	1,347	---	---

Source: Statistiska Centralbyrån (1966-).

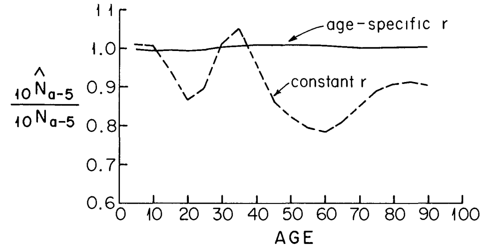

Figure 1: Sequence of estimates of the completeness of death registration for the total population with and without the assumption of stability, Swedish males, 1965-1970.

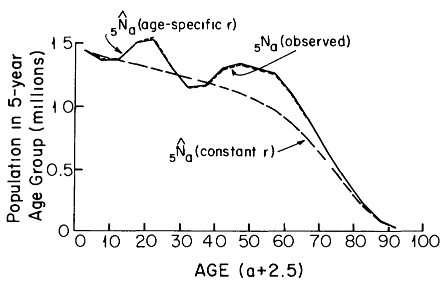

Figure 2: Comparison of observed and estimated age distributions, Swedish males, 1965-1970.

Korea, 1970-1975

Given evidence that the technique works well with a population for which data are virtually complete, we proceed to analyze a population whose data are severely defective We have chosen to study Korean females during the period 1970 to 1975 so that we may compare the results obtained using this method with those found in Preston et al. (1980).

Table 2 reveals that approximately 65 percent of female deaths in Korea were recorded during this period [8] This is in contrast to the figure of about 58 percent shown in Preston et al. (1980). The discrepancy in estimates is primarily due to the differing treatments of the open interval. An underestimate of N(60+) in the stable case had distorted the trend line of $\hat{N}(x+)/N(x+)$. It was then necessary to adjust r to correct for the downward slope, which in turn gave rise to a spuriously low estimate of the extent of completeness of registration. We have carried out the stable procedure using the present means of addressing the open interval problem and have compared these results with the results obtained from our non-stable procedure.

Figure 3 displays the values of $_{10}\hat{N}_{a-5}/_{10}N_{a-5}$ (over the range a=10, ..,55) derived in both analyses. [9] Although the median estimates of completeness are similar, note that the use of age-specific growth rates yields a far smoother set of ratios than that obtained under the assumption of stability. The slightly elevated values of $_{10}\hat{N}_{a-5}/_{10}N_{a-5}$ at ages 15, 20, and 25 (Table 2) are at age intervals where estimates show relatively large undercounts in the censuses (Coale, Cho, and Goldman, 1980, Figure A-5 and Table A-5).

Summary and Conclusions

Preston and his colleagues have developed a method to estimate the completeness of adult death registration in populations which are approximately stable. In this paper, it has been shown that, given a set of age-specific growth rates, minor modifications allow one to use the method even with populations which are far from stable.

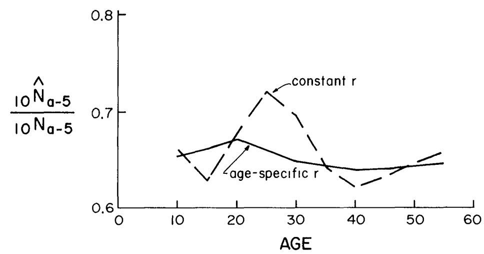

Figure 3: Sequence of estimates of the completeness of death registration for the total population with and without the assumption of stability, Korean females, 1970-1975.

Table 2: Female deaths, population, and rates of growth by age and values of $\hat{N}(a)$, $_5\hat{N}_a$, and $_{10}\hat{N}_{a-5}/_{10}N_{a-5}$ for Korea, 1970-1975.

Age	$_5r_a$	$_5D_a$	$_5N_a$	$\hat{N}(a)$	$_5\hat{N}_a$	$_{10}\hat{N}_{a-5}/_{10}N_{a-5}$
0	-.0048	28,533	10,313,278	1,416,494	7,096,297	---
5	-.0029	16,381	10,835,030	1,422,025	7,121,259	.672
10	.0055	12,141	10,744,253	1,426,479	7,004,890	.655
15	.0578	14,223	8,844,278	1,375,477	5,983,350	.663
20	.0421	14,284	6,839,605	1,017,863	4,574,162	.673
25	.0219	13,003	5,857,953	811,802	3,817,627	.661
30	.0015	12,958	5,442,945	715,249	3,530,349	.650
35	.0275	13,095	5,042,065	696,891	3,229,887	.645
40	.0343	14,435	4,214,360	595,064	2,707,947	.641
45	.0266	16,006	3,511,383	488,115	2,251,315	.642
50	.0362	18,915	2,846,740	412,411	1,848,215	.645
55	.0183	21,247	2,342,818	326,875	1,512,190	.648
60	.0273	178,708	5,388,703	278,002	---	---

Source: Coale, Cho, and Goldman (1980), Table 6; and National Bureau of Statistics (1972, 1977), Table 7

Another application of this technique is in its use with populations that are known to have rigorous registration systems but relatively poor enumeration. The impracticability of the method of extinct generations for use with data from current populations can now be circumvented. That is, we no longer must wait for cohorts to become extinct. Rather, we can now simply adjust cross-sectional death distribution data and subsequently obtain estimates of underenumeration by age.

The results of analysis of data from Sweden and Korea suggest that employing this refined version of the method results in greater accuracy of the estimated level of completeness of death registration in destabilized populations. There is a trade-off, however, involved between the original and modified versions of the method. The original stable method can be applied even if data from only one census are available. In addition, the stable r can be estimated as a by-product of the method. On the other hand, in the non-stable version, age-specific r's are used, in which case data from two censuses are nearly always required.

In comparison with other non-stable methods, especially the intercensal cohort survival method by Preston and Hill (1980) and the method of forward projection, it may be said that the computational process of the present method is slightly more complicated than those of already existing methods. However, these methods are inconvenient to apply when data are given in five-year age groups and the intercensal period is not close to a multiple of five years. This difficulty is not present in the method proposed here. In addition, the previous non-stable methods are based on the change in cohort size in two successive censuses. Given that the two censuses may suffer from age misreporting, the implied number of deaths (the difference in cohort sizes) may be subject to bias. This bias does not exist in the present method since we do not estimate deaths from cohort survivorship.

There are several limitations to this technique. First, the estimated completeness tends to be biased upward in the presence of net out-migration and downward by net in-migration. Therefore, if the extent of migration is significantly large compared with the number of deaths, prior adjustment of the data is necessary.

Second, we have assumed that there exists a reasonably high age (for practical purposes, at least 50) above which all age misstatement is found to occur. This condition may not be met in some populations.

Third, this method is based on the assumption that the underregistration of deaths is independent of age, at least among adults. However, as indicated in Preston et al. (1980), the completeness of registration sometimes differs between urban and rural areas, and this regional difference in underregistration may produce differential completeness by age in the total population. Similarly, the fact that completeness may be associated with the socioeconomic status of the families of the dead and that there exist socioeconomic differentials in mortality could lead to a violation of this assumption.

Fourth, this method may be sensitive to the differential enumeration of two successive censuses. Relative underenumeration in the first (second) census would raise (lower) age-specific growth rates and thereby bias the estimated completeness of death registration upward (downward). If there is no age misreporting and the completeness of death registration is constant over age, then ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ decreases (increases) with age if the first (second) census is relatively underenumerated. Therefore, we may inflate or deflate the set of observed ${}_5r_a$ by adding a constant such that the resulting sequence of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ becomes flat. [10]

Finally, underenumeration of the population is assumed to be constant over age. If, however, it is those at a young age who are enumerated to a different extent than the remainder of the population, then we can use the estimates of completeness of death registration obtained from data above that age without jeopardy. An estimate of completeness, ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$, is in no way contaminated by imperfections in the data below the age group to which it refers.

In spite of these limitations, most of which are shared with other methods for estimating underregistration, our analysis of data from Sweden and Korea seems to suggest that, if this method is used judiciously, then reliable estimates of completeness can be computed.

Notes

1 An open cohort refers to a group of persons born in a given year and before.

2 If the age group between ages a and a+5 is stable, then

$$N(x) \quad = \quad N(a)\exp[-(x-a) \cdot {}_5r_a]\,{}_{x-a}p_a \,, \qquad a \leq x \leq a+5$$

so that ${}_5\hat{N}_a$ is computed by

$${}_5\hat{N}_a \quad = \quad \int_a^{a+5} \hat{N}(x)dx \quad = \quad \hat{N}(a) \int_a^{a+5} \exp[-(x-a) \cdot {}_5r_a]\,{}_{x-a}p_a dx \,,$$

where ${}_{x-a}p_a$ is the survival probability from age a to x. In order to obtain ${}_{x-a}p_a$ we can utilize the fact that, as seen in populations that have accurate age data, the mortality in old ages is well approximated by the following exponential growth model, which is widely known as the Gompertz function:

$$\mu(x) \quad = \quad \mu(a)\exp[(x-a)\xi] \,, \qquad \text{for } x \geq a$$

where $\mu(x)$ is the death rate at age x. Based on the Gompertz mortality model, ${}_{x-a}p_a$ is estimated by

$${}_{x-a}\hat{p}_a \quad = \quad \exp\{[\hat{\mu}(a)-\hat{\mu}(x)]/\xi\} \,,$$

where $\hat{\mu}(a)$ and $\hat{\mu}(x)$ are given by

$$\hat{\mu}(a) \quad = \quad \frac{\ell n \left[\frac{\hat{N}(a+5)}{\hat{N}(a)}\right] + 5 \cdot {}_5r_a}{(1-\exp[5 \cdot \xi])/\xi}$$

and

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

Unfortunately, ξ cannot be directly estimated from defective mortality data. However, among most of the available population data in the world, ξ varies between .06 and .12, and the choice of a value between .08 and .10 seems to be an adequate approximation in computing ${}_5\hat{N}_a$. In the data analyses presented in this paper, we have used ξ=.10.

3 The way in which we treat the open interval is based on a suggestion by Ansley Coale. The derivation is as follows:

In a stable population, it holds that

$$N(a) = \int_0^\infty \exp[ry]D(a+y)dy \ .$$

In taking the first three terms of the Taylor expansion of exp[ry], we obtain

$$N(a) \doteq \int_0^\infty (1+ry+\frac{r^2y^2}{2})D(a+y)dy = D(a+)[1+r\bar{y}+\frac{r^2}{2}(\bar{y}^2+\sigma^2)] \ ,$$

where $\bar{y}$ and σ^2 are the mean and variance of age at death above a (less a), respectively.

Since $[d\bar{y}/dr]=-\sigma^2$ and σ^2 does not vary significantly with r, $\bar{y}$ can be approximated by

$$\bar{y} \doteq e(a) - r\sigma^2$$

Therefore, by substitution we obtain

$$N(a) \doteq D(a+)\left[1 + re(a) + \frac{r^2e(a)^2}{2} - \frac{r^2\sigma^2}{2}\right]$$

$$\doteq D(a+)\left\{\exp[re(a)] - \frac{r^2\sigma^2}{2}\right\} \ .$$

σ^2 is well approximated by $\sigma^2 \doteq [e(a)]^2/3$, for $a \geq 10$, for a wide array of existing life tables. Hence, it follows that

$$N(a) \doteq D(a+)\{\exp[re(a)] - [re(a)]^2/6\} \ .$$

4 Suppose, for example, that the mortality pattern of a population may be characterized by a Coale-Demeny model life table. In addition, let us say that we take e(0) to be 60 years, when in fact it is 55 years. The maximum error in e(70), 1 4 years, would occur in a situation where we had assumed the family to be "North" when it was actually "East". Had we correctly chosen a particular family of model life tables, the error would have been on the order of only half a year.

5 In order to approximate a closed population we remove migrants from the population and the deaths that would be associated with these migrants, assuming that they are subject to the same mortality schedule as the native population. An alternative means of handling the known migrant population is to treat out-migrants

as deaths and in-migrants as negative deaths (entries) at each age. The migrant population may then be interpreted as exiting a population by death (out-migrating) or entering a population as negative deaths (in-migrating).

6 We have used the value of e(95) given in the published Swedish life table for males (1.83 years) for the period 1966 to 1970 (Statistiska Centralbyrån, 1974).

7 Despite the fact that a lesser adjustment for migration was necessary when we applied the method to Swedish females, the resulting sequence of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$ showed slightly greater variation than that for the males (from 981 to 1.004).

8 Application of the intercensal cohort survival method by Preston and Hill (1980) results in a similar estimate of completeness.

9 Coale, Cho, and Goldman (1980) found that the mortality pattern of Korean females during this period approximately conforms to a model West life table, level 19. Hence, we have used a value of 17 years for e(60) in equation (7)

10 We may designate a corrected value of ${}_5r_a$ as ${}_5\tilde{r}_a$, where

$${}_5\tilde{r}_a = {}_5r_a + \delta .$$

If a particular value of δ results in a flat sequence of ${}_{10}\hat{N}_{a-5}/{}_{10}N_{a-5}$, then the differential completeness between the two censuses is estimated by

$$\frac{c_2}{c_1} = \exp[-\delta t] \doteq 1 - \delta t ,$$

where c_1 and c_2 have been previously defined, and t is the length of the intercensal period.

References

Brass, William. 1975. Methods for Estimating Fertility and Mortality from Limited and Defective Data: Based on Seminars Held 16-24 September 1971 at the Centro Latinoamericano de Demografía (CELADE) San José, Costa Rica. Chapel Hill, N.C.: University of North Carolina, International Program of Laboratories for Population Statistics. xii, 159 pp

Brass, William. 1979. A procedure for comparing mortality measures calculated from intercensal survival with the corresponding estimates from registered deaths. Asian and Pacific Census Forum (Honolulu) 6(2):5-7.

Coale, Ansley J.; and Demeny, Paul. 1966. Regional Model Life Tables and Stable Populations. Princeton, N.J.: Princeton University Press. xiii, 871 pp.

Coale, Ansley J.; Cho, Lee-Jay; and Goldman, Noreen. 1980 Estimation of Recent Trends in Fertility and Mortality in the Republic of Korea. Report No. 1. Washington: National Academy of Sciences [for] National Research Council, Committee on Population and Demography. xiv, 77 pp.

Korea. Republic. Economic Planning Board. National Bureau of Statistics. 1972. 1970 Population and Housing Census Report. Vol. 1. Complete Enumeration. [No.] 12-1 Republic of Korea. Seoul: Economic Planning Board, National Bureau of Statistics. 416 pp.

Korea Republic Economic Planning Board National Bureau of Statistics 1977 1975 Population and Housing Census Report Vol 1 Complete Enumeration [No.] 12-1 Whole Country Seoul Economic Planning Board, National Bureau of Statistics 368 pp

Martin, Linda G 1980. A modification for use in destabilized populations of Brass's technique for estimating completeness of death registration. Population Studies (London) 34(2).381-395

Preston, Samuel; and Hill, Kenneth 1980 Estimating the completeness of death registration Population Studies (London) 34(2):349-366

Preston, Samuel; Coale, Ansley J , Trussell, James; and Weinstein, Maxine 1980 Estimating the completeness of reporting of adult deaths in populations that are approximately stable Population Index (Princeton, N J) 46(2).179-202

Sweden Statistiska Centralbyrån 1966-1967 Folkmangdens Forandringar [Population Changes]. 1965 and 1966 editions Stockholm Statistiska Centralbyrån.

Sweden. Statistiska Centralbyrån 1969-1972. Befolkningsförändringar Del 3. Hela Riket och Länen m m [Population Changes ...Part 3 The Whole Country and the Counties etc.]. 1967, 1968, 1969, and 1970 editions. Stockholm. Statistiska Centralbyrån

Sweden. Statistiska Centralbyrån 1974 Livslangdstabeller for Artiondet 1961-1970 [Life Tables for the Decade 1961-1970]. Stockholm Statistiska Centralbyrån

Trussell, James, and Menken, Jane 1979. Estimating the completeness of deaths and relative underenumeration in two successive censuses. Asian and Pacific Census Forum (Honolulu) 6(2)·9-11

Vincent, Paul 1951 La mortalité des vieillards [The mortality of the aged]. Population (Paris) 6(2) 182-204.

Appendix: A Theoretical Basis for the Use of Age-Specific Growth Rates in Estimating the Completeness of Death Registration in a Closed Population

Theorem.

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

where N(a) is the number of persons aged a, $D^*(x)$ is the true number of deaths experienced by persons aged x and r(u) is the growth rate of the population aged u.

Proof: We have that

$$D^*(a) = -dN(a) = -\left[\frac{\partial N(a)}{\partial a} da + \frac{\partial N(a)}{\partial t} dt\right] \quad (A.1)$$

In addition,

$$\mu(a) = \frac{D^*(a)}{N(a)}$$

and

$$r(a) = \frac{1}{N(a)} \cdot \frac{\partial N(a)}{\partial t} dt .$$

After dividing both sides of equation (A.1) by N(a) and manipulating the terms, we obtain

$$\frac{1}{N(a)} \cdot \frac{\partial N(a)}{\partial a} da = - [\mu(a) + r(a)] . \qquad (A.2)$$

The integration of equation (A.2) from ages a to x yields

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

$$= N(a)_{x-a}p_a \exp\left[-\int_a^x r(u)du\right] , \qquad (A.3)$$

where the survival probability from ages a to x is

$$_{x-a}p_a = \exp\left[-\int_a^x \mu(u)du\right] .$$

Given that the probability of a person dying by age x after having survived to age a is $_{x-a}q_a$, we have that $_{\infty}q_a=1$. Hence,

$$N(a) = N(a)_{\infty}q_a = \int_a^{\infty} N(a)_{x-a}p_a \mu(x)dx$$

$$= \int_a^{\infty} N(a)_{x-a}p_a \exp\left[-\int_a^x r(u)du\right] \mu(x)\exp\left[\int_a^x r(u)du\right] dx .$$

Implementing equation (A.3), we have

$$N(a) = \int_a^{\infty} N(x)\mu(x) \exp\left[\int_a^x r(u) du\right] dx$$

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

Q.E.D.

For computational purposes, we note

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

$$= \left\{\int_{a+n}^{\infty} D^*(x)\exp\left[\int_{a+n}^{x} r(u)du\right]dx\right\}\cdot\exp\left[\int_a^{a+n} r(u)du\right]$$

$$+ \int_a^{a+n} D^*(x)\exp\left[\int_a^x r(u)du\right]dx$$

$$= N(a+n)\exp\left[\int_a^{a+n} r(u)du\right] + \int_a^{a+n} D^*(x)\exp\left[\int_a^x r(u)du\right]dx.$$

If $r(u) = {}_n r_a$, where $a \le u \le a+n$, and ${}_n D_a^* = \int_a^{a+n} D^*(x)dx$, then

$$N(a) = N(a+n)\exp[n \cdot {}_n r_a] + \int_a^{a+n} D^*(x)\exp[(x-a)\cdot {}_n r_a]dx$$

$$= N(a+n)\exp[n \cdot {}_n r_a] + {}_n D_a^* \exp[y \cdot {}_n r_a] ,$$

since there exists a y, where $a \le y \le x$, such that

$$\int_a^{a+n} D^*(x)\exp[(x-a)_n r_a]dx = \exp[y \cdot {}_n r_a]\int_a^{a+n} D^*(x)dx .$$

For n=5, we have that $y \doteq 2.5$ and $r(u) \doteq {}_5 r_a = \frac{1}{{}_5 N_a} \cdot \frac{d({}_5 N_a)}{dt}$,

where $a \le u \le a+5$, and hence,

$$N(a) \doteq N(a+5)\exp[5 \cdot {}_5 r_a] + {}_5 D_a^* \exp[2.5 \cdot {}_5 r_a] . \qquad (A.4)$$

If ${}_5 r_a = r$ for all a, it is clear that equation (A.4) reduces to

$$N(a) \doteq N(a+5)\exp[5r] + {}_5 D_a^* \exp[2.5r] ,$$

which is the computational formula found in Preston et al. (1980)

Acknowledgements

The text of this article was produced through the Interactive Computer Graphics Laboratory at Princeton University. The Editors of Population Index would particularly like to thank Professor Yehonathan Hazony, Walter Anderson, Alexandra Shulzycki, and Carol Nusser for their encouragement and assistance.