Chapter V

ESTIMATION OF ADULT MORTALITY FROM INFORMATION ON THE DISTRIBUTION OF DEATHS BY AGE

A. BACKGROUND OF METHOD

1. *Use of information on deaths by age and sex*

The straightforward way of calculating mortality rates is by using the information on deaths by age produced by a vital registration system. However, even though many countries possess such a system of registration it is often the case that not all the deaths are registered. As a result, the death rate implied by the reported deaths is usually an underestimate of the true death rate prevalent in the population in question, and some method of adjustment is required to transform the reported death rate into a better estimate of true mortality conditions. Of course, the same statements apply to the calculation of age-specific death rates; if a life table is calculated in the ordinary manner, and if deaths are underreported, the survival function, $l(x)$, will fall too slowly as age increases and estimates of life expectancy will be biased upward.

Over the years, demographers have suggested several methods of adjustment. Two methods were selected for inclusion here: one proposed by Preston and Coale;[1] and the other proposed by Brass.[2] A third method that essentially compares registered deaths between two censuses with the deaths implied by the census age distributions is presented in chapter IX.

The two methods described in this chapter are based on the assumption that the population being studied is stable (see chapter I, section C). They are both fairly robust to violations of the assumption of stability, particularly to recent changes in fertility and to gradual changes in mortality. Nevertheless, the methods hold strictly only for stable populations, and it is therefore helpful to recall the main characteristics of such populations. A stable population is one that has been subject to constant fertility and mortality for a long time. A feature of such a population is that the rate of exponential growth is constant at all ages, from which it follows that the same rate of growth must apply to births, deaths, deaths in any particular age group and the population of any age group.

The methods presented are also based on the assumption that the completeness of reporting of deaths is the same at all ages. That is, it is not assumed that reporting is 100 per cent complete; but if it is, for example, only 80 per cent complete, this figure applies to all age groups. In practice, because childhood and adult deaths are very often underreported to different extents, these methods are used only to estimate adult mortality (robust procedures for estimating childhood mortality are discussed in chapter III). Child mortality estimates obtained from the procedures described in chapter III and age-specific death rates for adults adjusted by the methods described in this chapter may be spliced together in the conventional manner or by using model mortality schedules as described in chapter VI.

2. *Organization of this chapter*

The two methods described in this chapter use information on the age distribution of deaths and that of the population. They both assume that the population is stable and that the degree of completeness of death registration is more or less the same at all ages after childhood (over age 5 or 10). The basic data required may be obtained from a vital registration system (deaths by age) and a census (age distribution of the population), or from surveys, whether retrospective or prospective in nature. In all cases, it is important to make sure that the deaths refer to the population whose age distribution is being used in the analysis. To aid the user in selecting the method best suited for a particular application, brief descriptions of the methods presented are given below (table 122 indicates their data requirements and the parameters they estimate):

Section B. Preston and Coale method. An equation is derived from stable-population theory that relates the population of age x to the deaths over age x expanded by a series of factors incorporating the stable growth rate. The ratios of the estimated population of age x derived from deaths over age x to the reported population of age x, denoted by $\hat{N}(x)/N(x)$, indicate the relative completeness of death registration. Although it is necessary to assume a growth rate, r, in order to calculate $\hat{N}(x)$, a "best" value of r can be selected as being that which produces the most consistent set of $\hat{N}(x)/N(x)$ ratios for different values of x. This method is more robust to departures from stability than the Brass method, but it is more sensitive to certain types of age misreporting;

Section C. Brass growth balance method. In a stable population, the rate of entry into the population aged x and over by reaching age x is equal to the rate of depar-

[1] Samuel Preston, Ansley J. Coale, James Trussell and Maxine Weinstein, "Estimating the completeness of reporting of adult deaths in populations that are approximately stable," *Population Studies*, vol. 46, No. 2 (Summer 1980), pp. 179-202.

[2] William Brass, *Methods for Estimating Fertility and Mortality from Limited and Defective Data* (Chapel Hill, North Carolina, Carolina Population Center, Laboratories for Population Studies, 1975).

TABLE 122. SCHEMATIC GUIDE TO CONTENTS OF CHAPTER V

Section	*Type of input data*	*Estimated parameters*
B. Preston and Coale Method	Deaths in a year classified by five-year age group, and by sex Mid-year population classified by five-year age group, and by sex Provisional estimate of the growth rate	Completeness of death-reporting in relation to population coverage Revised estimate of the growth rate Revised estimate of the death rate over age 10
C. Brass growth balance method	Deaths in a year classified by five-year age group and by sex Mid-year population classified by five-year age group and by sex	Completeness of death-reporting in relation to population coverage An estimate of the growth rate Revised estimate of the death rate over age 10

ture from the same population segment through death plus the stable population growth rate, which is the same for all values of x. This method uses this relation to estimate the stable growth rate and the relative completeness of death registration. It is somewhat less vulnerable to age exaggeration than the Preston and Coale method, but it is more sensitive to the effects of destabilization resulting from a rapid mortality decline.

B. PRESTON AND COALE METHOD

1. *Basis of method and its rationale*

In any population, the number of persons in a particular age group, say 25-29, at a particular time t will be equal to the total number of deaths to those persons from time t, when its members are 25-29, until the last survivor has died. If only 50 per cent of the deaths occurring every year are registered, then the ratio of the total number of deaths reported to the actual population will be 0.5, the value of the completeness of registration. The Preston-Coale procedure is based on this simple idea. Of course, in the example given above, one would have to wait a long time to obtain an estimate of the completeness of registration. However, if the number of deaths that will occur after time t can be estimated from the number of deaths reported for a particular year or calendar period, the comparison of the reported number of persons in a particular age group with the estimated total number of future deaths to the age group should provide an estimate of the completeness with which deaths are registered. In a stable population, there is a precise relationship between the numbers of current deaths and the numbers of persons in the population. The persons now aged x, to whom deaths are currently occurring, are the survivors of births x years ago, which, by the properties of stable populations, must have been smaller in number than current births by a factor of $\exp(-rx)$. Hence, the number of deaths that will occur to the current number of births when they are aged x will be larger than the current number of deaths to persons aged x by a factor of $\exp(rx)$.

It follows that the number of deaths that will be experienced by persons currently aged x (theoretically equal to the number of such persons) can be estimated from the current number of deaths recorded at each age above x. Specifically, if $N(x)$ is the number of persons at age x in a stable population with growth rate r, and $D(x)$ is the number of deaths at age x, then an estimate of $N(x)$, $\hat{N}(x)$ can be expressed as

$$\hat{N}(x) = \sum_{a=x}^{\omega} D(a)\exp(r(a-x)). \qquad \text{(B.1)}$$

If the population is genuinely stable, the rate of growth correctly specified, and deaths and population accurately reported, then $\hat{N}(x)$ will equal $N(x)$. If, however, deaths are underreported by some fixed proportion (say, 20 per cent), $\hat{N}(x)/N(x)$ will be less than 1.0—in this instance, 0.8. Since $\hat{N}(x)$ is an estimate of the population at exact age x, it cannot be compared directly to the reported population, which is normally tabulated by five-year age group. One can estimate $N(x)$ as $({}_5N_{x-5}+{}_5N_x)/10$, where ${}_nN_x$ is the reported population between ages x and $x+n$; or, alternatively, ${}_5\hat{N}_x$ can be estimated as $2.5(\hat{N}(x)+\hat{N}(x+5))$.

Thus, the ratio of $\hat{N}(x)$ to $N(x)$ is an estimate of the completeness of death registration in relation to population enumeration, but it would be unwise to estimate completeness from a single ratio $\hat{N}(x)/N(x)$ or ${}_5\hat{N}_x/{}_5N_x$, since the number of persons reported at a particular age (even if determined as the average of the number over an interval centred on x) is subject to overstatement or understatement because of age-misreporting, or because of differential omission of persons in a particular span of ages. A better estimate of the completeness of death registration in relation to population enumeration can be obtained from erratic $\hat{N}(x)/N(x)$ ratios by considering some type of representative value of individual ratios over an extensive range of ages. For example, the median can be used, or the ratio of the sum of estimates of ${}_5\hat{N}_x$ over an extensive range of ages to ${}_5N_x$ over the same range. The obvious possibility is the ratio

$$C = \sum_{x=0}^{\omega} {}_5\hat{N}_x \Big/ \sum_{x=0}^{\omega} {}_5N_x, \qquad \text{(B.2)}$$

which is a comparison of the total population estimated from the number of deaths to the total enumerated population. However, this possibility is not necessarily the best because estimates of $\hat{N}(0)$ and ${}_5\hat{N}_0$ are based in part on the reported numbers of infant and child deaths, which usually constitute a large proportion of the total number of deaths and are often subject to a completeness of registration quite different from that of deaths at older ages.

Another problem arising in the estimation of $\hat{N}(x)$ is the determination of the population estimated from reported deaths at the upper ages. At the upper end of the age range, there is always an "open interval", in which the number of deaths and the number of persons are tabulated in an undivided age category, such as 90+, 85+, 80+ or 75+. If the lower boundary of the open interval is denoted by A, the estimation of $\hat{N}(A)$ and of $\hat{N}(A+)$ from $D(A+)$ and the rate of growth requires special procedures because the distribution of deaths within the open interval is not available. It can be proved that the estimation of $\hat{N}(A)$ is less sensitive to uncertainties about the distribution of deaths in the open interval than is the estimation of $\hat{N}(A+)$.

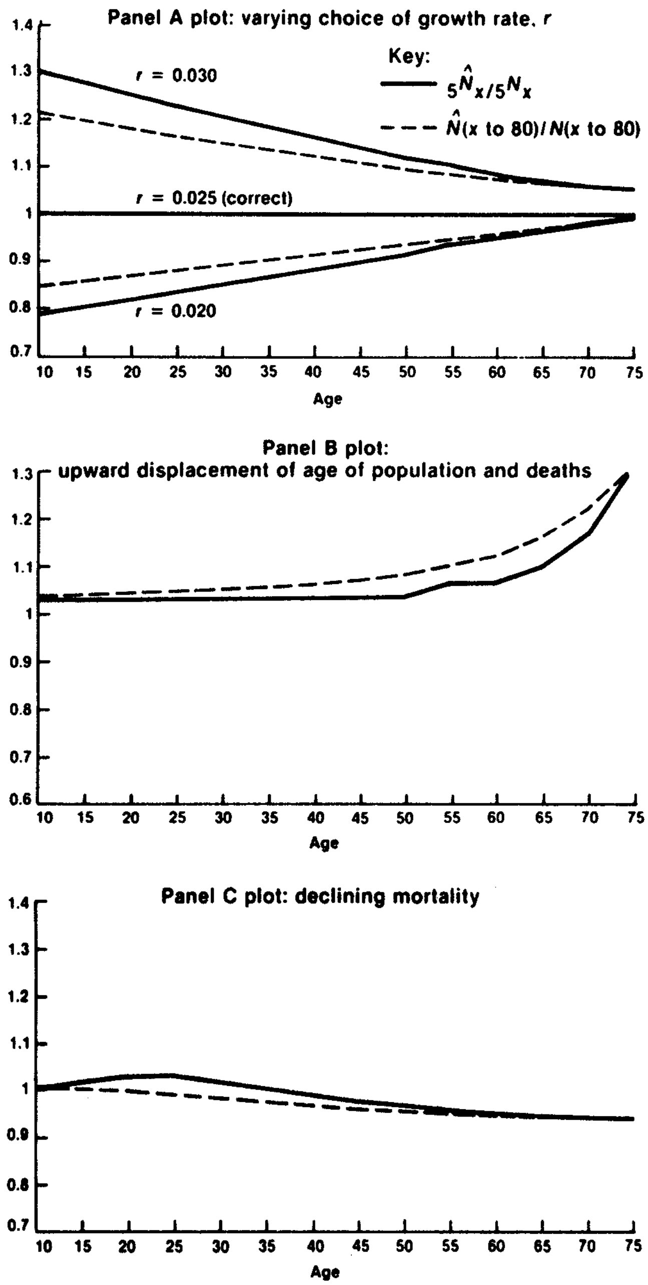
Because of the special difficulties involved in estimating the values of $\hat{N}$ in early childhood and in the open interval, a practical option to obtain the least erratic values of the ratios of estimated to reported population is to divide the sum of the estimated ${}_5\hat{N}_x$ values over certain ranges of age x (from age 5 up to age $A-5$, five years before the beginning of the open interval) by the sum of the reported ${}_5N_x$ values over the same range.

It is therefore suggested that two sets of estimates be considered in assessing the completeness with which deaths are recorded in relation to the population—the sequence of $10\hat{N}(x)/{}_{10}N_{x-5}$, and the sequence of $\hat{N}(x$ to $A)/N(x$ to $A)$ ratios for different values of x. In an ideal situation of accurately reported age, a genuinely stable age distribution of known growth rate, and the same completeness of registration of deaths at different ages, both sequences of ratios would be constant with respect to age and both would yield the same fixed estimate of completeness. In actuality (as stated earlier), the first set of estimates is erratic because of misreporting of age in the population or because of differential completeness of enumeration. In addition, it is frequently the case that the estimates both of $\hat{N}(x)$ and of $\hat{N}(x$ to $A)$ are affected by violations of the assumptions upon which the method is based. It turns out, however, that the most typical violations of the assumptions produce characteristic deviations from the expected horizontal plot of the $\hat{N}/N$ sequence, deviations that are fairly distinctive and therefore, in certain circumstances, interpretable. In order to examine such effects, the estimation method proposed was applied to data from a true stable population with a growth rate of 0.025, which have been artificially distorted. The conclusions reached from these experiments are summarized below.

(a) *Incorrect choice of growth rate.* In panel A of figure 11, plots of $\hat{N}/N$ are shown for the correct choice of the growth rate, r (0.025), and for values that are too high or too low. It can be seen that if r is too high, the sequence of $\hat{N}/N$ falls with age; the opposite occurs if the value of r is too low;

(b) *Upward displacement of age at death.* Very often the age of older persons is displaced upward, perhaps because social status is enhanced by advanced age. Typically, reported ages at death are displaced upward to a greater extent than the reported ages of the living. A common situation in which age at death is exaggerated in relation to the age of the living produces the rising sequence of $\hat{N}/N$ ratios shown in panel B of figure 11. Overstatement of age at death increases the estimate of $\hat{N}(x)$ in two ways. Recall that $\hat{N}(x)$ is the sum of the products of deaths over age x and exponen-

Figure 11. Illustration of the effects of deviations from the assumptions on the plot of estimated to reported population ratios, $\hat{N}/N$, for model cases



tial terms $\exp(r(a-x))$ (equation (B.1). Overstatement of age at death in the form of a transfer of reported deaths from under to over age x increases $\hat{N}(x)$ by simple exaggeration of the relative number of deaths over age x. Moreover, overstatement of age at death confined within the span of ages over x (that is, with no transfer from under x to over x) also increases $\hat{N}(x)$ for the following reason. When the sum

$$\sum_{a=x}^{\omega} D(a)\exp(r(a-x))$$

that determines $\hat{N}(x)$ is calculated, a transfer of deaths from lower to higher ages within the span over x results in multiplication of the transferred deaths by a larger than warranted exponential factor.

The first or direct effect of an upward transfer of deaths across some age boundary x is substantially greater than the second or indirect effect of age exaggeration above x, resulting in too old an age distribution of deaths over age x, though the indirect effect continues to distort the results for all values of x. The size of the indirect effect declines as x decreases, however, so that by age 10 it will be only about half as large as at age 70, say, if net upward transfers occur only over ages 75, 80 and 85. The overall effect, therefore, of exaggeration of age at death on the plot of $\hat{N}(x)/N(x)$ is likely to be an imperceptible upward trend with x up to the age at which exaggeration begins, followed by a sharp upward kink reflecting the direct effects of exaggeration. The values of the points up to the kink should not be severely distorted, unless exaggeration of age at death begins at fairly young ages;

(c) *Departures from stability*. If fertility has recently declined or if there was a birth deficit at some time in the past that created a small cohort in the population, the age distribution would not in fact be stable. The match between the true number of persons at age x and the number constructed from the number of deaths, even if deaths were completely reported, would be impaired. Specifically, small cohorts caused by low fertility either at the younger ages because of a recent decline, or for another cohort because of the effect of a war, would be reflected in high values of $\hat{N}(x)/N(x)$ for these age intervals. However, if the estimate of completeness of registration of adult deaths is derived by analysing deaths and population only over age 15 or 20, a deviation from stability caused by recent changes in fertility will have no effect on the results. A more frequent situation in contemporary populations in developing countries is a decline in mortality that has occurred over the past 20 or 30 years. A history of declining mortality causes a departure from the stable age distribution, but usually a rather limited departure. The sequence of ratios of $\hat{N}/N$ produced in a typical population in which mortality has been declining is shown in panel C of figure 11. The sequence of $\hat{N}(x)/N(x)$ rises and then falls, while that of $\hat{N}(x \text{ to } A)/N(x \text{ to } A)$ falls slowly at first and then more rapidly. The actual pattern depends upon the particular trend followed by the mortality decline and the time period during which the decline has taken place, but an inverted U pattern is fairly typical.

In any real population, of course, misstatement of reported age by the population will tend to produce a zigzag pattern in the sequence of the $\hat{N}(x)/N(x)$ ratios. Such distortions are lessened when the cumulated version of these ratios, $\hat{N}(x \text{ to } A)/N(x \text{ to } A)$, is considered, that is, when the estimated population whose age ranges from x to the lower limit of the open interval is compared with the number of persons in the enumerated population over the same age interval from x to A. Unfortunately, it is not always easy to distinguish between the effects of the different violations of the underlying assumptions.

Both overstatement of age at death and choice of a growth rate that is too small cause a rising sequence of $\hat{N}(x)/N(x)$. However, too small a value of r causes the sequence to rise linearly at all ages; while overstatement of age at death, concentrated as it often is at the highest ages, causes an upward bend at the first age at which a systematic upward transfer of reported age at death occurs. If the chosen r is too small, one simply tries a larger value. If $\hat{N}(x)/N(x)$ rises monotonically only over age 75, for example, overstatement of age at death is the likely cause. An estimate can be made of the extent to which the number of reported deaths above each age has been increased by age overstatement, and a rough correction can be made by reassigning the excess reported deaths to the next lower age interval. The $\hat{N}(x)/N(x)$ ratios can then be recalculated. Usually this correction only slightly reduces the $\hat{N}(x)/N(x)$ ratios at younger ages. An alternative to redistribution may be the selection of a lower beginning age for the open interval, so that age overstatement is largely confined to that interval.

A more intractable problem is to distinguish a declining set of $\hat{N}(x)/N(x)$ ratios that is associated with a population that has a history of declining mortality from that produced when the growth rate used is too high. Moreover, adjusting the estimates by assigning a smaller value to the rate of growth would lower the estimate of completeness; yet, the estimate derived from the population with a history of declining mortality may already be too low. When faced with a declining sequence of $\hat{N}(x)/N(x)$, one must therefore decide on other grounds whether the population in question has recently experienced extensive declines in mortality. If such a decline has occurred, the median of the ratios of cumulated populations with beginning ages from 10 to 35 or from 10 to 45 can be accepted as an estimate of completeness. On the other hand, if a prolonged and substantial decline in mortality appears unlikely, a lower value of the rate of increase can be employed to produce a more level sequence of $\hat{N}(x)/N(x)$ ratios.

Unfortunately, the generally optimistic tone of this discussion of the diagnostic value of the $\hat{N}(x)/N(x)$ sequence is not invariably justified in practice. Errors are not always as orderly as one would wish, and combinations of different types of errors can give rise to patterns in the $\hat{N}(x)/N(x)$ ratios that cannot be interpreted according to the simple rules presented above. Hence, it will not always be possible to obtain one single, unambiguous estimate of completeness, though by the judicious use of different growth rates and open intervals, it should at least be possible to obtain some idea of the range within which the true figure lies.

2. *Data required*

The following data are required for this method:

(a) The deaths occurring in a specific time period, normally in a given year (but the average over a two-year or a five-year period may be used as well), classified by age. Five-year age groups are adequate. It is useful, though not essential, to classify by sex;

(b) The population by age group (and sex) corresponding to the mid-period for which deaths are given;

(c) An estimate of the growth rate during the period. One need not have a precise estimate, since an incorrect choice of r should be evident in a diagnostic plot of $\hat{N}(x)/N(x)$. Indeed, the final choice of r may be determined by that value which produces the most level trend over the central age groups.

3. *Computational procedure*

The steps of the computational procedure are described below.

Step 1: estimation of growth rate. Since the method is applied only to adults, the provisional estimate of the growth rate could be obtained as the intercensal rate of growth of the population over age 10 or 15. However, a better estimate can usually be found by taking the median of the growth rates of the population over ages 10, 15, 20, ..., 60. If censuses are taken at times t_1 and t_2, and the population in question at time t_1 is N_1 and at time t_2 is N_2, then the growth rate is calculated as $r = \ln(N_1/N_2)/(t_1 - t_2)$. If it is not possible to calculate intercensal growth rates, then an iterative procedure that would begin with a reasonable value of r and modify it until the sequence of $\hat{N}(x)/N(x)$ is as close to being horizontal as possible would be necessary.

Step 2: adjustment of reported population to mid-point of period. Strictly speaking, when the data on deaths are for a given year, the age distribution of the mid-year population should be used. If a census was taken on or near the middle of the year, its data can be used without adjustment. If the population data refer to the beginning or the end of the year, one possible course of action is to use the average of the data on deaths for the two adjacent years. Another, more general solution is to adjust the population figures so that they conform in time to the death data. The simplest way to accomplish this task is to assume that the population is growing at the growth rate estimated in step 1. Thus, if t_m is the date corresponding to the middle of the period (or year) to which the death data refer and t_c is the reference date of the census, the adjusted population values are computed as

$$_5N_x^{adj} = {_5N_x} \exp(r(t_m - t_c)).$$

Note that the values of $_5N_x$ and $N(x+)$ would all be multiplied by the same adjustment factor $\exp(r(t_m - t_c))$. Thus, in practice, it is much simpler to apply this method without adjusting the data at this stage, but making sure to adjust the estimated completeness of death registration, C, by the factor $\exp(-r(t_m - t_c))$. Furthermore, if one is interested in adjusted age-specific mortality rates and not in the level of completeness *per se*, this adjustment is not necessary. When no adjustment is performed, C will be an estimate of completeness of death registration in relation to the size of the unadjusted population, so multiplying the reported deaths by $1.0/C$ and dividing by the unadjusted population will yield the same estimated mortality rates as those obtained when the population is adjusted as explained above.

Another common situation that requires adjustment of the $_5N_x$ values arises when the deaths correspond to an intercensal period. In this case, the population figures should be the average of those recorded by the two censuses and the death figures should be the average annual number of deaths.

Step 3: estimation of population from registered deaths. It can be shown that equation (B.1), from which the estimated population at exact age x, $\hat{N}(x)$, is calculated, can be expressed in recursive fashion as $\hat{N}(x) = \hat{N}(x+5)\exp(5r) + {_5D_x}\exp(2.5r)$. Hence, the calculation of $\hat{N}(x)$ is straightforward, except for the open-ended age interval. In general, since the estimation for this interval requires a certain amount of approximation, it is advisable to use as narrow an open interval as the data permit, taking into consideration the possible effects of age exaggeration.

For each of the five-year age intervals from x to $x+4$, it is assumed that the length of time from x to the mid-point of the interval provides an adequate estimate of the average length of time over which deaths need to be inflated by the stable growth rate. In the case of the open interval (deaths at age A and over), there is no mid-point, so some alternative procedure is required to include these heavily weighted deaths in the analysis. There are two different problems involved, the first being to decide what age A to select as the lower boundary of the open interval, and the second being to estimate the age point at which the deaths in the interval should be assumed to be concentrated in order to apply the correct growth rate adjustment to them.

The first problem is sometimes resolved by the form in which the data are available. If the tabulations of deaths by age stop at age groups 55-59, 60-64 or 65-69, leaving open-ended intervals beginning at values of A of 60, 65 or 70, respectively, then nothing better can be done than to use the value of A as given. If more age detail is available, the question arises whether to use the A of the basic data or to use a lower value. The advantage of using the highest possible value according to the data tabulations is that the higher the value of A, the smaller the approximation made in estimating the age point at which deaths in the open interval are assumed to be acting. The advantage of using a lower value of A

is that the age errors in death reporting, which probably increase with age, will be reduced. In general, a value of A of 75 may be adopted except in cases where age reporting is extremely good, when values of 80 or 85 can be used. Age 75 is selected because it offers the most advantageous balance between gains obtained by reducing the effects of age errors at very high ages and losses related to the approximation involved in estimating the weight of the deaths in the open interval.

The second problem, that of estimating the weight for the deaths in the open interval, can be solved by using models. If A is the lower boundary of the open interval, the number of people aged A is given by

$$\hat{N}(A)=\sum_{x=A}^{\omega} D(x)\exp(r(x-A)),$$

an expression that cannot be evaluated from the data because the values of $D(x)$ are not available. However, there exists a length of time, $z(A)$, such that

$$D(A+)\exp(rz(A))=\sum_{x=A}^{\omega} D(x)\exp(r(x-A))$$

so that the number of people aged A can be calculated from the number of deaths over age A, $D(A+)$, as

$$\hat{N}(A)=D(A+)\exp(rz(A)). \quad \text{(B.3)}$$

Values of $z(A)$ have been calculated for a range of model cases having different growth rates, mortality levels, mortality patterns and values of A. Least-squares regression was used to relate the values of $z(A)$ to two parameters available in any application, namely, the growth rate r and the exponential value of the ratio of deaths over age 45 to deaths over age 10. The coefficients shown in table 123 make it possible to estimate $z(A)$ for values of A ranging from 45 to 85 for each family of the Coale-Demeny model life tables by using the following equation:

$$z(A)=a(A)+b(A)r+c(A)\exp\left[\frac{D(45+)}{D(10+)}\right]. \quad \text{(B.4)}$$

In any particular application, therefore, the value of $z(A)$ may be estimated from the growth rate, r, as estimated in step 1, and the ratio of deaths at age 45 and over to deaths at age 10 and over, a ratio that can be readily calculated from the distribution of reported deaths by age group. The population aged A can then be estimated using equation (B.3). If no indications exist as to which family of life tables should be used, the West family is a satisfactory default.

Once the value of $\hat{N}(A)$ has been calculated from deaths in the open age interval, the remaining calculations are straightforward. One begins with the open interval and continues downward, using the following recursive equations to calculate $\hat{N}(A-5)$ from $\hat{N}(A)$, $\hat{N}(A-10)$ from $\hat{N}(A-5)$ and so on:

$$\hat{N}(x)=\hat{N}(x+5)\exp(5r)+{}_5D_x\exp(2.5r), \quad \text{(B.5)}$$

$${}_5\hat{N}_x=2.5(\hat{N}(x)+\hat{N}(x+5)) \quad \text{(B.6)}$$

where ${}_5D_x$ is the number of reported deaths in the interval from x to $x+4$.

TABLE 123. COEFFICIENTS[a] FOR ESTIMATION OF THE AGE FACTOR FOR THE OPEN INTERVAL, $z(A)$, FROM THE RATIO OF DEATHS OVER AGE 45 TO DEATHS OVER AGE 10 AND THE POPULATION GROWTH RATE

Regional family (1)	Age A (2)	Coefficients a(A) (3)	b(A) (4)	c(A) (5)
North	45	−11.42	185.2	17.02
	50	−10.63	167.2	14.99
	55	− 9.78	147.8	12.96
	60	− 8.57	126.1	10.85
	65	− 6.83	101.6	8.62
	70	− 4.53	74.6	6.28
	75	− 1.91	47.1	3.98
	80	0.46	22.7	2.00
	85	1.82	6.4	0.67
South	45	−15.26	183.4	18.23
	50	−14.91	168.4	16.36
	55	−14.22	151.2	14.38
	60	−12.89	130.8	12.22
	65	−10.67	106.4	9.80
	70	− 7.53	78.4	7.15
	75	− 3.84	48.8	4.47
	80	− 0.47	22.6	2.14
	85	1.47	5.6	0.63
East	45	−15.87	174.3	18.06
	50	−15.14	158.5	16.06
	55	−13.97	140.4	13.93
	60	−12.10	118.8	11.60
	65	− 9.43	93.9	9.05
	70	− 6.07	66.5	6.38
	75	− 2.52	39.3	3.81
	80	0.37	16.8	1.73
	85	1.79	3.5	0.48
West	45	−13.43	181.4	17.57
	50	−12.49	163.6	15.49
	55	−11.24	143.7	13.34
	60	− 9.50	121.2	11.07
	65	− 7.21	96.1	8.67
	70	− 4.48	69.2	6.23
	75	− 1.64	42.9	3.91
	80	0.72	20.5	1.98
	85	2.03	5.9	0.70

Note: $z(A)$ is the age that satisfies the relationship

$$D(A+)\exp(rz(A))=\sum_{A}^{\omega} D(x)\exp(r(x-A))$$

and it is estimated from the equation

$$z(A)=a(A)+b(A)r+c(A)\exp\left[\frac{D(45+)}{D(10+)}\right].$$

[a] Based on 11 levels of mortality, with e_0 ranging from about 40 years to about 75 years.

Step 4: estimation of completeness of death registration. The completeness of death registration, C, may be taken as the average level of the sequence of $\hat{N}(x)/N(x)$ values, or of the sequence $\hat{N}(x \text{ to } A)/N(x \text{ to } A)$. Use of the median or the mean of the values over ages where the latter sequence is approximately flat, say, from age 10 to age 45, is suggested. If the plot of the $\hat{N}(x)/N(x)$

ratios shows a generally rising trend with x, suggesting that the growth rate used was too low, or a falling trend with x, suggesting that the growth rate used was too high, the growth rate should be modified accordingly and the calculations repeated. If some test is required to decide which set of points provides the closest approximation to a horizontal line, the absolute deviations of the points, excluding obvious outliers, from their grand mean (excluding again the obvious outliers) could be summed; and that growth rate giving the minimum sum could be accepted. Another procedure would be to divide the points, again excluding outliers, into two groups of equal size (for instance, points for ages from 10 to 35 and from 40 to 65); to calculate the mean value of $\hat{N}(x)/N(x)$ for each group, and to accept the growth rate that minimizes the difference between the two means. If the slope of $\hat{N}(x)/N(x)$ indicates that the growth rate has been misspecified, the following rule of thumb can be used to estimate an adjusted value of r. The effect of an estimate of r that incorporates an error Δr on $\hat{N}(x)$ is to increase $\hat{N}(x)$ by a multiplier $1+(\Delta r)e_x$, where e_x is the expectation of life at age x. Given that e_x does not vary rapidly with the overall level of mortality for advanced values of x, the estimation of e_x for particular applications can be avoided while still obtaining a reasonable first approximation to Δr. If the values of $\hat{N}(x)/N(x)$ for values of x from 20 to 40 and from 40 to 60 are averaged, and the difference between them (a rough indicator of slope) is obtained by subtracting the 20-40 average from the 40-60 average, a first approximation to Δr can be found by dividing the result by 14.5. This value of 14.5 is suitable for an e_{10} of about 55 years; if the true value of e_{10} were 45 years, the value of the denominator should be about 12.7, whereas for 65 it should be 18.4.

Step 5: adjustment of reported death rates for underregistration and calculation of a life table for adults. If the age intervals are, for example, 10-14,..., 75-79, 80+, then adjusted death rates are calculated in a straightforward manner as follows:

$$_5m_x^{adj} = \frac{_5D_x}{C} \cdot \frac{1.0}{_5N_x} \quad \text{for } x = 10,\ldots,75 \qquad \text{(B.7)}$$

and

$$_{\omega-80}m_{80}^{adj} = \frac{D(80+)}{C} \cdot \frac{1.0}{N(80+)} \quad \text{for } x = 80. \qquad \text{(B.8)}$$

Lastly, the $_5m_x^{adj}$ values are converted into life-table $_5q_x$ values in the usual manner for the age groups from 10-14 to 75-79:

$$_5q_x = \frac{(5.0)_5m_x^{adj}}{1.0+(2.5)_5m_x^{adj}} \qquad \text{(B.9)}$$

Then, using the relation $l(x+5) = l(x)(1.0 - {}_5q_x)$, values of $l(x)$ for ages from 15 to 80 can be determined with a radix $l(10)$ of 1.0 at age 10. If an estimate of $l(10)$ is available from the procedures described in chapter III, it can be employed directly in calculating the full life table. However, some irregularities generally will remain in the sequence of adjusted $_5m_x$ values, because of age-misreporting or other errors; and further adjustment and smoothing, for example, by comparison with mortality models, will be desirable. For further details on these types of adjustments, see chapter VI.

4. *First detailed example*

The case of the female population of El Salvador in 1961 is considered first. The deaths registered in that year, classified by age, are shown in table 124, along with the population enumerated at the time of the 1961 census, with a reference date of 5 May 1961.

TABLE 124. FEMALE POPULATION AND DEATHS BY AGE, EL SALVADOR, 1961

Age (1)	*Registered female deaths* $_5D_x$ (2)	*Reported female population* $_5N_x$ (3)	*Female population aged from x to 75* $N(x \text{ to } 75)$ (4)
0	6 909	214 089	1 258 060
5	610	190 234	1 043 971
10	214	149 538	853 737
15	266	125 040	704 199
20	291	113 490	579 159
25	271	91 663	465 669
30	315	77 711	374 006
35	349	72 936	296 395
40	338	56 942	223 859
45	357	46 205	166 417
50	385	38 616	120 212
55	387	26 154	81 596
60	647	29 273	55 442
65	449	14 964	26 169
70	504	11 205	11 205
75	1 360[a]	16 193[a]	-
TOTAL	13 652	1 274 253	

[a] For the open interval 75+.

TABLE 125. INTERCENSAL GROWTH RATES FOR FEMALES, EL SALVADOR, 1950-1971

Age range (1)	*Intercensal growth rate* 1950-1961 (2)	1961-1970 (3)	Average (4)
10+	0.0236	0.0330	0.0283
15+	0.0224	0.0307	0.0266
20+	0.0232	0.0289	0.0260
25+	0.0246	0.0287	0.0266
30+	0.0258	0.0290	0.0274
35+	0.0253	0.0298	0.0276
40+	0.0266	0.0307	0.0287
45+	0.0284	0.0309	0.0296
50+	0.0293	0.0311	0.0302
55+	0.0351	0.0326	0.0338
60+	0.0366	0.0327	0.0346

Step 1: estimation of growth rate. Censuses were taken in El Salvador in 1950, 1961 and 1971. From the cumulated population figures, $N(x+)$, growth rates for the population over ages 10, 15, 20,..., 60 were calculated as $r = \ln[N_2(x+)/N_1(x+)]/(t_2-t_1)$ for each of the two intercensal periods. Results are shown in table 125, along with the averages of the two. If the population

were really stable and all data were accurately recorded, then the growth rates would all be the same. Examination of table 125 reveals that they are not, so some choice must be made. The median of the average values (which roughly correspond to the growth rates in 1961) for ages ranging from 20 to 60, 0.0287, was selected. If this choice is a poor one, the diagnostic plot of $\hat{N}/N$ should reveal the error and an improved choice can be made.

Step 2: adjustment of reported population to mid-point of period. The population figures correspond to the census reference date of 5 May, while the deaths are centred on 1 July. Thus, the ${}_5N_x$ and $N(x \text{ to } A)$ values are too small by a tiny fraction. The difference in time between the mid-year and 5 May is 56 days, or 0.1534 of a year. Thus, the adjustment factor that should be applied to the population data is $\exp[(0.1534)(0.0287)] = 1.0044$. However, as stated before, the population data need not be adjusted at this stage, since it is simpler to adjust the resulting completeness of death registration (see step 4 below).

Step 3: estimation of population from registered deaths. In this case, the open interval is the age group 75 and over, so A is taken equal to 75. Thus, in order to begin the recursive calculation of $\hat{N}$, an estimate of $z(75)$ is needed.

The rate of growth has been estimated as 0.0287, and the mortality pattern regarded as suitable is West, so all that is required to estimate $z(75)$ is the ratio of deaths at ages 45 and over to deaths at ages 10 and over. Cumulating from the bottom of column (2) of table 124,

$$D(45+) = {}_{\omega-75}D_{75} + {}_5D_{70} + \cdots + {}_5D_{45}$$

$$= 1{,}360 + 504 + \cdots + 357 = 4{,}089.$$

The value of $D(10+)$ can be obtained most readily by subtracting deaths under age 10 from all deaths:

$$D(10+) = D(0+) - {}_5D_0 - {}_5D_5$$

$$= 13{,}652 - 6{,}909 - 610 = 6{,}133.$$

The ratio is then calculated:

$$D(45+)/D(10+) = 4{,}089/6{,}133 = 0.6667;$$

and its exponential value found:

$$\exp[D(45+)/D(10+)] = \exp(0.6667) = 1.948.$$

Using the coefficients from table 123 in equation (B.4),

$$z(75) = a(75) + b(75)(0.0287) + c(75)(1.948)$$

$$= -1.64 + (42.9)(0.0287) + (3.91)(1.948) = 7.21.$$

$\hat{N}(75)$ is then obtained by applying equation (B.3):

$$\hat{N}(75) = D(75+)\exp[rz(75)] =$$

$$(1{,}360)\exp[(0.0287)(7.21)]$$

$$= (1{,}360)(1.2299) = 1{,}672.9$$

Then $\hat{N}(70)$ is computed using the recursive equation (B.5):

$$\hat{N}(70) = \hat{N}(75)\exp(5r) + {}_5D_{70}\exp(2.5r)$$

$$= (1{,}672.9)\exp[(5.0)(0.0287)]$$

$$+ (504)\exp[(2.5)(0.0287)]$$

$$= 1{,}931.0 + 541.5 = 2{,}472.5.$$

TABLE 126. VALUES OF REPORTED DEATHS, ESTIMATED POPULATION FOR DIFFERENT AGES AND RATIOS OF ESTIMATED TO REPORTED POPULATION, FEMALES, EL SALVADOR, 1961

Age x (1)	Reported deaths ${}_5D_x$ (2)	Estimated population: $\hat{N}(x)$ (3)	${}_5\hat{N}_x$ (4)	$\hat{N}(x \text{ to } 75)$ (5)	Ratio of estimated to reported population: ${}_5\hat{N}_x/{}_5N_x$ (6)	$\hat{N}(x \text{ to } 75)/N(x \text{ to } 75)$ (7)	$\hat{N}(x \text{ to } 60)/N(x \text{ to } 60)$ (8)
75	1 360[a]	1 672.9	-	-	-	-	-
70	504	2 472.5	10 363.5	10 364	0.925	0.925	-
65	449	3 336.4	14 522.3	24 886	0.970	0.951	-
60	647	4 546.4	19 707.0	44 593	0.673	0.804	-
55	387	5 663.7	25 525.3	70 118	0.976	0.908	0.992
50	385	6 951.3	31 537.5	101 656	0.817	0.847	0.895
45	357	8 407.5	38 397.0	140 053	0.831	0.842	0.873
40	338	10 068.0	46 188.8	186 241	0.811	0.832	0.856
35	349	11 996.5	55 161.3	241 403	0.756	0.815	0,829
30	315	14 186.1	65 456.5	306 860	0.842	0.820	0.835
25	271	16 666.3	77 131.0	383 991	0.841	0.825	0.839
20	291	19 550.7	90 542.5	479 533	0.798	0.819	0.832
15	266	22 853.3	106 010.0	580 543	0.848	0.824	0.837
10	214	26 609.6	123 657.3	704 200	0.827	0.825	0.837
5	610	31 371.0	144 951.5	849 152	0.762	0.813	0.824

[a] For the open-ended age group 75 and over.

Estimates of $\hat{N}(x)$ for all the required values of x are shown in column (3) of table 126. Next, ${}_5\hat{N}_{70}$ is calculated as

$${}_5\hat{N}_{70} = 2.5(\hat{N}(75) + \hat{N}(70)) = 2.5(1{,}672.9 + 2{,}472.5) = 10{,}363.5$$

and $\hat{N}(x \text{ to } 75)$ is calculated by cumulation. The complete sets of estimated ${}_5\hat{N}_x$ and $\hat{N}(x \text{ to } 75)$ values are shown in columns (4) and (5) of table 126.

Step 4: estimation of completeness of death registration. Values of ${}_5\hat{N}_x / {}_5N_x$ and $\hat{N}(x \text{ to } 75)/N(x \text{ to } 75)$ are shown in columns (6) and (7) of table 126 (the denominators of the ratios are found in columns (3) and (4) of table 124); both sets of values are displayed in figure 12. To illustrate the stability of the ratios $\hat{N}(x \text{ to } A)/N(x \text{ to } A)$ with respect to A, their values using an open interval of 60+ are also shown in table 126. Although they are slightly higher than those obtained using as open interval 75+, the differences observed are small. Because, as figure 12 shows, the sequence of $\hat{N}(x \text{ to } 75)/N(x \text{ to } 75)$ ratios is approximately flat over the age range from 5 to 60, their median, 0.825, is selected as an unadjusted estimate of C, the completeness of death registration. Adjustment of C is necessary to make allowance for the difference existing between the date of enumeration of the population and the mid-year (see step 2). Thus, the adjusted completeness of death registration, $\hat{C}$, is

$$\hat{C} = 0.825/1.0044 = 0.821$$

Because in this case the difference in the dates mentioned above is relatively small, the difference between the adjusted and unadjusted C is minimal, and for all practical purposes, adjustment can by bypassed. However, it is important to remember its theoretical necessity in cases where the date difference is large.

Step 5: adjustment of reported death rates. The age-specific mortality rates can be adjusted in the straightforward manner by using equation (B.7) with the unadjusted estimate of C. For example, the adjusted death rate for age group 50-54 is given by

$${}_5m_{50}^{adj} = \frac{{}_5D_{50}}{C({}_5N_{50})} = \frac{385}{(0.825)(38{,}616)} = 0.0121.$$

5. *Second detailed example*

The second example, using data from Andra Pradesh State in India, illustrates how sets of $\hat{N}/N$ ratios can be used to estimate the completeness of death registration when the sequence of ratios is more irregular than in the case of El Salvador. The details of the calculations are not presented, however, since they are similar to those of the first example.

Step 1: estimation of growth rate. In Andra Pradesh, the male population aged 10 and over increased from 13,137,000 on 1 March 1961 to 15,806,000 on 1 April 1971. The intercensal period was thus 10 years and 31 days, or 10.085 years, so the intercensal growth rate of the population aged 10 and over, $r(10+)$, is given by

$$r(10+) = [\ln(15.806) - \ln(13.137)]/10.085 = 0.0183.$$

Step 2: adjustments to basic data. The basic data (for males) on deaths by age and population by age for 1970 and 1971 come from the Indian Sample Registration

Figure 12. Plot of ratios of estimated to reported female population, El Salvador, 1961

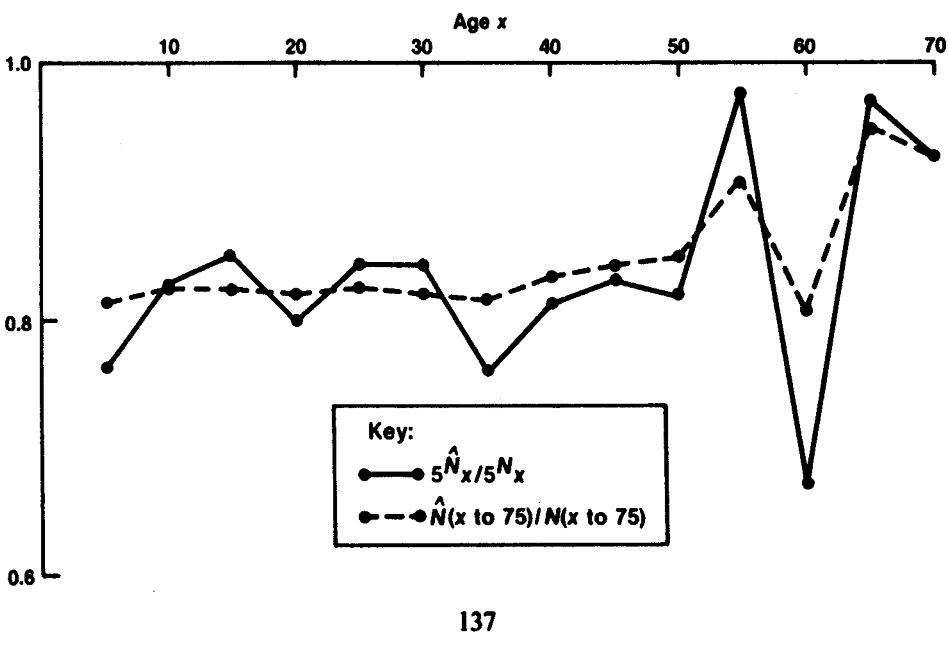


TABLE 127. AGE DISTRIBUTION OF MALE POPULATION, CENTRAL DEATH RATES AND DERIVED NUMBERS OF DEATHS, BY AGE GROUP, ANDRA PRADESH STATE, INDIA, 1970 AND 1971

	1970			1971				
Age group (1)	Male population (2)	Central death rate (3)	Number of deaths (4)	Male population (5)	Central death rate (6)	Number of deaths (7)	Total number of deaths (8)	Total population (9)
0-4	1 534	0.0462	70.87	1 355	0.0447	60.57	131.44	2 889
5-9	1 388	0.0049	6.80	1 416	0.0050	7.08	13.88	2 804
10-14	1 211	0.0020	2.42	1 230	0.0023	2.83	5.25	2 441
15-19	871	0.0027	2.35	1 008	0.0021	2.12	4.47	1 879
20-24	735	0.0043	3.16	728	0.0035	2.55	5.71	1 463
25-29	696	0.0060	4.18	708	0.0069	4.89	9.07	1 404
30-34	682	0.0044	3.00	673	0.0041	2.76	5.76	1 355
35-39	620	0.0070	4.34	627	0.0064	4.01	8.35	1 247
40-44	608	0.0105	6.38	591	0.0100	5.91	12.29	1 199
45-49	458	0.0125	5.73	467	0.0133	6.21	11.94	925
50-54	425	0.0201	8.54	423	0.0266	11.25	19.79	848
55-59	245	0.0285	6.98	240	0.0302	7.25	14.23	485
60-64	258	0.0493	12.72	247	0.0403	9.95	22.67	505
65-69	111	0.0607	6.74	143	0.0646	9.24	15.98	254
70+	158	0.1221	19.29	144	0.1279	18.42	37.71	302
TOTAL	10 000			10 000			318.52	20 000
Deaths at age 45+							122.32	
Deaths at age 10+							173.22	

System and are published in the form of the age distribution of the sampled population per 10,000 population of all ages and of central death rates, ${}_5m_x$. Data by five-year age group are available up to age group 65-69, the open interval being 70 and over. No adjustments are required for the population data, which are intended to represent person-years of exposure, but deaths by age group must be obtained by multiplying the central death rates, ${}_5m_x$, by the population exposed to the risk of dying, ${}_5N_x$. The calculations are carried out separately for 1970 and 1971 in table 127. Total deaths and total population for the period 1970-1971 are then obtained by summing the figures for individual years, the results being shown in columns (8) and (9) of table 127.

Step 3: estimation of population from reported deaths. The ratio of $D(45+)/D(10+)$ is calculated from table 127 as (122.32)/(173.22), or 0.7062. The value of its exponential is therefore exp (0.7062) = 2.0263. Assuming that mortality in India conforms to a South pattern, $z(70)$ is estimated as

$$z(70) = -7.53 + (78.4)(0.0183) + (7.15)(2.0263)$$

$$= 8.393.$$

The value of $\hat{N}(70)$ is then obtained as

$$\hat{N}(70) = D(70+)\exp[(8.393)(0.0183)]$$

$$= (37.71)(1.1660) = 43.97.$$

Full results are shown in table 128. The ${}_5\hat{N}_x/{}_5N_x$ and $\hat{N}(x \text{ to } 70)/N(x \text{ to } 70)$ ratios are plotted in figure 13.

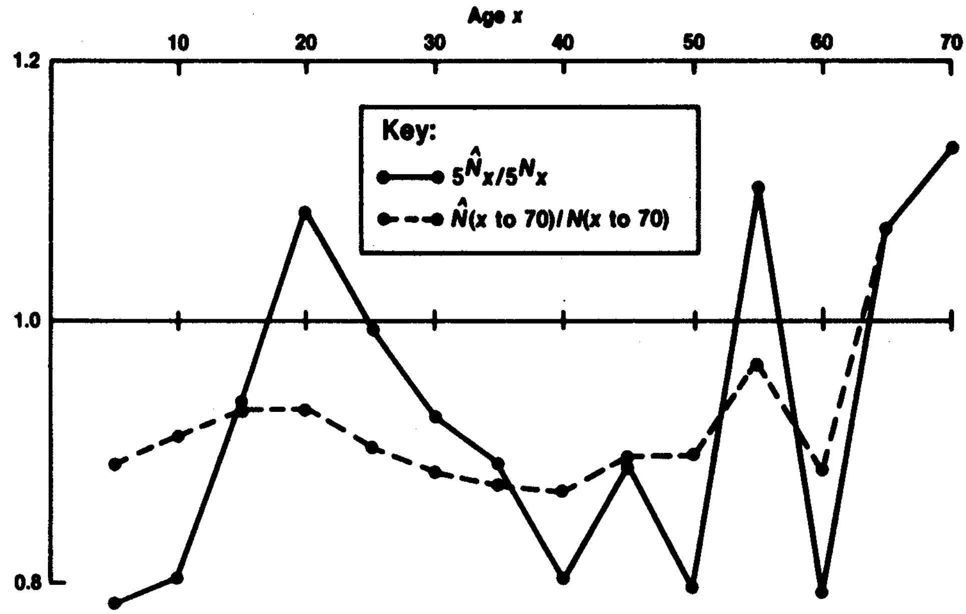
The plot of ${}_5\hat{N}_x/{}_5N_x$ rises from a value below 0.8 at

TABLE 128. VALUES OF REPORTED DEATHS, ESTIMATED POPULATION FOR DIFFERENT AGES AND RATIOS OF ESTIMATED TO REPORTED POPULATION, MALES, ANDRA PRADESH STATE, INDIA, 1970-1971

	Reported deaths	Reported population		Estimated population			Ratio of estimated to reported population	
Age x (1)	${}_5D_x$ (2)	${}_5N_x$ (3)	$N(x \text{ to } 70)$ (4)	$\hat{N}(x)$ (5)	${}_5\hat{N}_x$ (6)	$\hat{N}(x \text{ to } 70)$ (7)	${}_5\hat{N}_x/{}_5N_x$ (8)	$\hat{N}(x \text{ to } 70)/N(x \text{ to } 70)$ (9)
70	37.71	302	-	43.97	-	-	-	-
65	15.98	254	254	64.91	272.20	272.2	1.072	1.072
60	22.67	505	759	94.86	399.42	671.6	0.791	0.885
55	14.23	485	1 244	118.85	534.28	1 205.9	1.102	0.969
50	19.79	848	2 092	150.95	674.50	1 880.4	0.795	0.899
45	11.94	925	3 017	177.91	822.15	2 702.6	0.889	0.896
40	12.29	1 199	4 216	207.82	964.33	3 666.9	0.804	0.870
35	8.35	1 247	5 463	236.47	1 110.73	4 777.6	0.891	0.875
30	5.76	1 355	6 818	265.16	1 254.08	6 031.7	0.926	0.885
25	9.07	1 404	8 222	300.06	1 413.05	7 444.7	0.994	0.905
20	5.71	1 463	9 685	334.79	1 587.13	9 031.9	1.085	0.933
15	4.47	1 879	11 564	371.55	1 765.85	10 797.7	0.940	0.934
10	5.25	2 441	14 005	412.65	1 960.50	12 758.2	0.803	0.911
5	13.88	2 804	16 809	466.72	2 198.43	14 956.6	0.784	0.890
0	131.44	2 889	19 698	649.03	(2 789.38)	(17 746.0)	(0.966)	(0.901)

age 5 to a value over 1.0 at age 20, then falls to 0.80 at age 40, remaining between 0.79 and 0.89 from ages 35 to 60, except for age 55. The probable explanation of this up-and-down movement is the deviation of the reported age distribution from a stable form, which, although attributable to a true lack of stability is more likely to be the result of differential omission, or of the tendency to misstate age. It should be borne in mind that the construction of the estimated population ${}_5\hat{N}_x$ from the recorded distribution of deaths and the growth rate, being cumulative in form, tends to follow a stable age distribution quite closely. Thus, if there are segments of the age distribution that are undercounted in the census or survey, the value of ${}_5\hat{N}_x/{}_5N_x$ will reach a peak at these points; and where, through age-misreporting, the population is over-enumerated, ${}_5\hat{N}_x/{}_5N_x$ will show a minimum. The high point of ${}_5\hat{N}_x/{}_5N_x$ at age 20 is almost certainly the result of an undercount of the cohort aged 20-24. The cumulated values, $\hat{N}(x$ to $70)/N(x$ to $70)$, show no general trend with age; and the median value of 0.896 can be accepted as a sensible estimate of completeness of death registration.

Figure 13. Plot of ratios of estimated to reported male population, Andra Pradesh State, India, 1970-1971



It is likely that the age-misreporting that appears to have distorted the reported age distribution of the population also produced an inexact reported age distribution of deaths. Thus, the life table obtained from the central death rates, ${}_5m_x$, adjusted for underreporting of deaths by a factor of about 1.0/0.9, might represent the overall level of mortality moderately well, but the detailed structure of the resultant schedule of mortality rates and the accompanying life table would probably be distorted. It is therefore preferable to calculate the expectation of life at age 10 in this life table and to utilize a model life table with the same life expectancy as perhaps a more trustworthy representation of mortality.

C. Brass growth balance method

1. *Basis of method and its rationale*

In 1975, Brass[3] proposed a method to estimate the completeness of registration of adult deaths. This method is based on the equation:

$$N(x)/N(x+)=r+D^*(x+)/N(x+) \qquad (C.1)$$

where $N(x)$ is the number of persons of exact age x; $N(x+)$ is the total number of persons aged x and over; $D^*(x+)$ is the total number of deaths occurring to persons aged x and over; and r is the growth rate.

Brass proved that equation (C.1) is exact for stable, closed populations. Its validity can be explained in common-sense terms by the following argument: since $N(x)$ may be thought of as being the number of persons in a year entering the group of those aged x and over, the ratio $N(x)/N(x+)$ can be interpreted as a "birth rate" for the population aged x and over. $D^*(x+)/N(x+)$ is the death rate corresponding to the same population; and, if one denotes by $r(x+)$ the growth rate for the population aged x and over, the equation

$$N(x)/N(x+)=r(x+)+D^*(x+)/N(x+) \quad (C.2)$$

[3] *Ibid.*

simply states the familiar truism that in a closed population the birth rate is equal to the sum of the growth rate and the death rate. Now invoking the assumption of stability, it is the case that in a stable population the growth rate $r(x+)$ is, by definition, the same for every x; and, therefore, $r(x+)$ can be replaced by r, and equation (C.2) becomes equation (C.1), that proposed by Brass.

Once the validity of equation (C.1) is established by assuming stability, a second assumption is made. Suppose that instead of observing $D^*(x+)$ (the total number of deaths over age x), only a proportion of them was recorded, say $D(x+)$, where

$$D(x+)=C(x)D^*(x+), \tag{C.3}$$

$C(x)$ being a factor representing the completeness of registration of deaths at age x and over.

If it is then assumed that the completeness of death registration does not vary with age, at least over age 5 or age 10, $C(x)$ can be replaced by a constant C that does not change with age. Letting $K=1/C$ and substituting equation (C.3) in equation (C.1), the following relationship is obtained:

$$N(x)/N(x+)=r+K(D(x+)/N(x+)). \tag{C.4}$$

For a closed, stable population, where the completeness of death registration is the same at every age and where age-reporting is accurate, equation (C.4) provides a method by which to estimate the completeness of death registration. According to equation (C.4), the relationship between $D(x+)/N(x+)$ and $N(x)/N(x+)$ is linear; and the slope of the line defined by the points $[D(x+)/N(x+),\ N(x)/N(x+)]$ is the value of the adjustment factor K. Hence, to estimate K, one needs only to find the slope of the line defined by the points $[D(x+)/N(x+),\ N(x)/N(x+)]$. Note, however, that no explicit allowance is being made for age-misreporting, whether of age at death or of age of the living, though the use of cumulation is likely to smooth out some of the effects of age errors. It is fair to say, therefore, that whenever the age structure of deaths or of the population is distorted by poor reporting, the estimates yielded by this method may be biased, although the use of cumulation and judicious elimination of points from the fitting procedure will reduce the effects of such errors.

In practice, the points $[D(x+)/N(x+),\ N(x+)/N(x+)]$ seldom fall exactly on a straight line, and K is obtained by selecting the line that best fits the observed points. In some cases, however, the deviations of these points from a linear trend are so marked that the use of this method of estimation is unwarranted. Large deviations from linearity may be due to several causes. The most common cause is the inaccuracy of the data (usually the existence of age-misreporting), but lack of linearity may also be due to differential completeness of death registration by age or to the lack of stability of the population considered (in this case, $r(x+)$ would not be constant for every x). So, in general, a plot of the observed values $D(x+)/N(x+)$ against $N(x)/N(x+)$ provides a good diagnostic tool to assess the validity of using this method to estimate K. Unfortunately, when the points are markedly non-linear it is not generally possible to identify a unique cause for this lack of linearity, nor to adjust the data in such a way as to make the use of this method of estimation possible. If, however, the deviation from approximate linearity is confined to points for the elderly, a straight line can be fitted excluding the deviant points and therefore avoiding serious distortion.

It is important to notice that the relationship expressed by equation (C.4) allows the estimation both of K, the adjustment factor for registered deaths, and of r, the growth rate of the population. This estimate of r is not robust to deviations from the hypotheses under which equation (C.4) was derived, and the user should not regard it as necessarily sound. However, the estimate of r may be compared with estimates obtained from other sources, and reasonable agreement may be interpreted as some confirmation that the assumptions being made are justified.

When a solid, independent estimate of r is available for a stable population where age-misreporting is not prevalent, equation (C.4) may be modified in the following way:

$$N(x)/N(x+)=r+K(x)D(x+)/N(x+) \tag{C.5}$$

where the adjustment factor K is no longer a constant, but is allowed to change with age. The assumption of equal underenumeration of deaths at all ages is, therefore, being dropped; and, since r is assumed to be known, equation (C.5) can be rewritten in the following form to allow the estimation of an adjustment factor, $K(x)$, for each open-ended age group:

$$K(x)=(N(x)-rN(x+))/D(x+). \tag{C.6}$$

Sadly, equation (C.6) can seldom be used in practice because it is uncommon to find cases of stable populations where age-reporting is fairly accurate and where $K(x)$ would therefore measure differential underregistration of deaths by age rather than differential age-misreporting.

To conclude this discussion, some comments need to be made on the validity of estimates obtained by this method when the population being studied is not stable. A stable population is established when fertility and mortality have both remained constant for a fairly long period of time. In today's world, few, if any, populations are truly stable. In most cases, destabilization has been brought about by changes in mortality. Most human populations have seen their chances of survival improve during the past 40 years and cannot, therefore, be stable. However, simulation has shown that when stable populations are destabilized by prolonged mortality changes that occur slowly, the bias introduced in the estimation of K by this lack of stability is relatively small.[4] Only when abrupt changes occur does the bias become large.[5] Of course, these simulations have been carried out using data that, apart from representing a non-stable population, are otherwise perfect. It is possible, therefore, that in reality even the biases introduced by rapid mortality changes may be small compared with those caused by the poor quality of the available data. However, it is important to remember that declining mortality usually biases the value of K upward, thus producing an underestimate of the completeness of death registration. Therefore, when this method is used to estimate completeness in a population where mortality has been declining and where age is reported accurately, the death-rate estimate obtained may safely be interpreted as an upper bound for the true rate.

The effect that changes in fertility have on the estimation of K has not been as extensively studied as the effects of changing mortality. The reason is that fertility changes have only recently become common in countries with poor data; and because these changes affect mainly the youngest age groups, they have little impact on the performance of this method of estimation.

2. *Data required*

The data required for this method are described below:

(a) The number of deaths occurring during a period (normally one year) classified by age and sex (though classification by sex is not necessary). Five-year age groups are adequate. The last age group must be open-ended; that is, it must include all the deaths at or over some age A;

(b) The number of persons in each age group at the mid-point of the period being considered, classified by sex if the deaths are so classified. The age groups used must correspond to those used for deaths. The last age group must once again be open, including all the population aged A and over.

3. *Computational procedure*

Two variants of the Brass method have been proposed, the difference lying essentially in the way in which the left-hand side of equation (C.4) is evaluated. One procedure, and that originally proposed, is to attempt to estimate $N(x)$, the number of persons reaching age x in a year, from the number of persons in the five-year age groups adjacent to x; the second is to estimate $N(x)$ for the mid-point of a five-year age group. The latter procedure is theoretically preferable, because the non-linearity of the age distribution is less marked over a five-year age span than over a 10-year span. However, for a population in which there is more than a trivial amount of age-heaping, the theoretical advantage of the second procedure is outweighed by the practical advantage of the first, that is, that the effects of heaping are reduced when $N(x)$ is estimated from two adjacent five-year age groups, one of which is likely to be inflated (the one including an age ending in zero) and one of which is likely to be deflated. Numerous applications have shown that the first procedure yields results that are more easily interpreted. Hence, only this procedure to estimate $N(x)$ is described here. The steps of the complete computational procedure are given below.

Step 1: the person-years lived by the population subject to the risk of dying. This step, which consists of adjusting the reported population to the mid-point of the period for which the death data are available, is exactly the same as step 2 of the Preston-Coale procedure. It need not, therefore, be repeated, especially since it can generally be omitted, in accordance with the discussion under step 2 in subsection B.3.

Step 2: calculation of population at an exact age. By definition, $N(x)$ is the number of persons who reach age x during the course of the year under consideration. When the total population for the mid-year is classified by single year of age, $N(x)$ can be estimated from the number of persons in the two contiguous age groups, ${}_1N_{x-1}$ and ${}_1N_x$. Thus, if one lets ${}_1N_x$ be the number of persons of age x enumerated by the census or survey,

$$N(x)=({}_1N_{x-1}+{}_1N_x)/2.0. \qquad \text{(C.7)}$$

If the classification by age is made in five-year age groups and ${}_5N_x$ is the number of persons in the age group from x to $x+4$ at the time of the census or survey, $N(x)$ can be estimated as

$$N(x)=({}_5N_{x-5}+{}_5N_x)/10.0. \qquad \text{(C.8)}$$

In general, equation (C.8) is preferable because it contributes to the reduction of the effects of age-heaping.

Step 3: calculation of population over an exact age. The total number of persons aged x and over is denoted by $N(x+)$. Thus, when the data are available for five-year age groups,

$$N(x+)=\sum_{j=x}^{A-5} {}_5N_j+N(A+) \qquad \text{(C.9)}$$

where $N(A+)$ represents the number of persons in the last, open-ended age interval. If the data are available only by single year of age, then

$$N(x+)=\sum_{y=x}^{A-1} {}_1N_y+N(A+). \qquad \text{(C.10)}$$

Note that there is no advantage in calculating $N(x+)$ from single-year rather than grouped data, unless the entire analysis is to be carried out by single years of age, a procedure that is likely to be excessively tedious and more unreliable when any degree of age-heaping is present.

[4] Hoda M. Roshad, "The estimation of adult mortality from defective registration", unpublished doctoral dissertation, University of London, 1978.

[5] Linda G. Martin, "A modification for use in destabilized populations of Brass's technique for estimating completeness of birth registration", *Population Studies*, vol. XXXIV, No. 2 (July 1980), pp. 381-395.

Step 4: calculation of number of deaths after an exact age. The calculation of the cumulated number of deaths from age x onward is very similar to that of $N(x+)$. $D(x+)$ is just the total number of deaths recorded as having occurred to persons aged x and over. Thus, if the deaths are classified by five-year age group,

$$N(x+)=\sum_{j=x}^{A-5} {}_5D_j + D(A+) \qquad \text{(C.11)}$$

where $D(A+)$ denotes the deaths in the open-ended age interval A and over. If the data are only available by single year, or it is wished for some reason to carry out the analysis by single years of age, then

$$D(x+)=\sum_{y=x}^{A-1} {}_1D_y + D(A+). \qquad \text{(C.12)}$$

Step 5: points defined by partial death and birth rates. Using the quantities calculated in the previous steps, the calculation of the values $D(x+)/N(x+)$ and $N(x)/N(x+)$ is straightforward. Once calculated, they should be displayed graphically, plotting the values of $D(x+)/N(x+)$ on the x-axis and those of $N(x)/N(x+)$ on the y-axis. The same scale should be used on both axes. The plotted points should ideally follow a linear trend. If it is clear that the points do lie roughly on a straight line, the next step can be performed with some confidence. Otherwise, the use of this estimation method may have to be abandoned.

Step 6: selection of a best fitting line. There are several ways of fitting a straight line to a series of points. Probably the most widely used method is that known by the name of "least-squares". The least-squares line is that which minimizes the sum of the squared distances between its points and the points to be approximated. In the sense that it approximates all these points as closely as possible, it provides a best fit. If the data points derived in step 5 were affected only by random errors, this line would certainly be that chosen to estimate the adjustment factor K. However, given the inaccuracies typically present in the data sets considered here, the use of other methods of fitting appears to be preferable. Two of the simplest are suggested below:

(a) *The "mean" line.* This line is defined by the means of the abscissae (horizontal axis values) and ordinates (vertical axis values) of the derived points when those points are divided, according to age, into two groups of approximately equal size. For example, if 15 age groups are used, the values of $D(x+)/N(x+)$ for the first eight values of x are averaged, as are the values for the last eight x values. (The middle point is included in both averages.) Similarly, mean values are obtained for the first eight and last eight values of $N(x)/N(x+)$. The desired line is that which passes through the two points defined by the two pairs of mean co-ordinates. This line will not be as near as possible to all the points derived in step 5, but it will closely approximate their general trend when this trend is linear. One possible disadvantage of this fitting procedure is that it gives equal weight to all the derived points. In practice, it is often the case that one or two of these points (usually at the two extremes of the age range) depart markedly from the linear trend followed by the rest. In such cases, it may be better to ignore them altogether than to take them into account in selecting a representative straight line;

(b) *The "robust" line.* To deal with such cases as that described above, where linearity is apparent in the middle of the age range but where distortions are present at the extremes, it is recommended that trimmed means be used in fitting the mean line, instead of using the usual means. The line fitted by using trimmed means is called a "robust" line because trimming makes the fit less sensitive to large deviations from linearity at the extremes of the age range. Trimmed means are a generalization of the usual mean or average, which is calculated by giving equal weight to each entry. When trimming is performed, the weights applied to different observations are not constant for all observations; the sum of weights times observations is standardized by dividing it by the sum of the weights used. The procedure for computing trimmed means is relatively simple and is fully described in the detailed examples.

Step 7: adjustment of death rate. Once the straight line that best fits the points derived in step 5 is selected, the values of the adjustment factor, K, and of the growth rate, r, are given by its slope and y-intercept, respectively. The value of the completeness of death registration, C, is just the reciprocal of K, that is, $C = 1.0/K$, provided all the necessary data adjustments were carried out in step 1. If the value of r is plausible, one can attach greater confidence to C, the estimate of completeness. In such a case, an adjusted death rate can be obtained by multiplying the reported death rate by K. Note, however, that the trend of the points $[N(x)/N(x+), D(x+)/N(x+)]$ is determined mainly by those corresponding to fairly advanced ages, so that effectively the level of completeness estimated refers only to adult deaths. Therefore, the value of K should be regarded as an adjustment factor only for the death rate over age 5 or 10, and may not be applicable to the death rate of the population as a whole.

The value of K may also be used to adjust the derived age-specific death rates, ${}_5m_x$. This procedure is valid only if the assumption of constant completeness of death registration by age holds true. The adjusted ${}_5m_x$ values can then be used, together with independent estimates of child mortality (see chapter III), to construct a life table for the population considered. In practice, this life table is seldom constructed directly from the adjusted ${}_5m_x$ values since age-misreporting makes them too unreliable; however, they can be used as a basis for fitting a mortality model from which the final life table will be derived. For further details of this procedure, see chapter VI.

4. *First detailed example*

In this example, the case of the female population of El Salvador in 1961 is again analysed. The numbers of female deaths and women classified by age have already been presented in table 124 in the previous detailed example. Step 1 is omitted here, as adjustment for the census date is not necessary at this point.

Step 2: calculation of population at an exact age. The value of $N(x)$, the number of persons of exact age x, is obtained by adding the number of persons in two contiguous age groups and dividing by the number of years spanned by those age groups (usually 10). In this case, $N(10)$, for example, is obtained as follows:

$$N(10) = (190{,}234 + 149{,}538)/10 = 33{,}977.2.$$

Column (2) of table 129 shows the values obtained for all values of x up to 70; no value can be computed for 75, because the open age interval begins at that age.

TABLE 129. ELEMENTS NEEDED TO ESTIMATE COMPLETENESS OF DEATH REGISTRATION AMONG ADULT FEMALES, EL SALVADOR, 1961

Age x (1)	Population at exact age x $N(x)$ (2)	Population aged x and over $N(x+)$ (3)	Reported deaths over age x $D(x+)$ (4)
5	40 432.3	1 060 164	6 743
10	33 977.2	869 930	6 133
15	27 457.8	720 392	5 919
20	23 853.0	595 352	5 653
25	20 515.3	481 862	5 362
30	16 937.4	390 199	5 091
35	15 064.7	312 488	4 776
40	12 987.8	239 552	4 427
45	10 314.7	182 610	4 089
50	8 482.1	136 405	3 732
55	6 477.0	97 789	3 347
60	5 542.7	71 635	2 960
65	4 423.7	42 362	2 313
70	2 616.9	27 398	1 864

Step 3: calculation of population over an exact age. The number of persons aged x and over, denoted by $N(x+)$, is computed by cumulating the numbers reported in each age group (shown in table 124) from that beginning with age x upward. For example, $N(50+)$ is

$$N(50+) = (38{,}616 + 26{,}154 + 29{,}273 + 14{,}964$$
$$+ 11{,}205 + 16{,}193) = 136{,}405.$$

All values of $N(x+)$ are given in column (3) of table 129. Note that in practice one calculates $N(x+)$ by cumulating downward, each time adding the population of one more five-year age group to $N((x+5)+)$, so $N(50+) = N(55+) + {}_5N_{50}$, for example.

Step 4: calculation of number of deaths after an exact age. The calculation of $D(x+)$ is exactly analogous to that of $N(x+)$. Thus, for example, $D(50+)$ is

$$D(50+) = (385 + 387 + 647 + 449 +$$
$$504 + 1{,}360) = 3{,}732$$

where the deaths reported in each age group have been taken from table 124. Also, it is important to point out that to calculate both $N(x+)$ and $D(x+)$, the number of persons whose age was recorded as unknown is ignored. Since these numbers are generally small, serious biases are not likely to be introduced by disregarding them.

Step 5: points defined by partial birth and death rates. Using the values obtained in the previous steps, the ratios $D(x+)/N(x+)$ and $N(x)/N(x+)$ are calculated. Their values are shown in table 130. The calculation of the partial birth rate, $N(x)/N(x+)$, involves simply dividing the entries listed in column (2) of table 129 by the corresponding entries in column (3) of the same table. For example,

$$N(40)/N(40+) = 12{,}987.8/239{,}552 = 0.0542.$$

Similarly, the calculation of the partial death rate, $D(x+)/N(x+)$, involves the division of the entries in column (4) by those in column (3) of table 129. Thus, for example,

$$D(40+)/N(40+) = 4{,}427/239{,}552 = 0.0185.$$

Table 130 gives results for other ages and figure 14 shows a plot of the points. Note that the values of $D(x+)/N(x+)$ are plotted on the x-axis (abscissa), while those of $N(x)/N(x+)$ appear on the y-axis (ordinate). The same scale has been used to plot both values. The use of equal scales for both axes aids in the assessment of linearity of the plotted points.

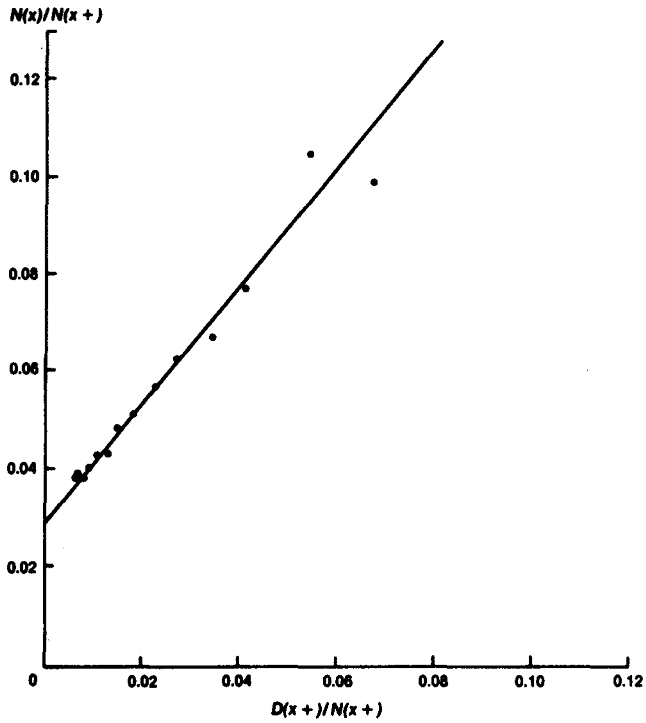
TABLE 130. PARTIAL DEATH AND BIRTH RATES FOR FEMALES, EL SALVADOR, 1961

Age x (1)	Partial death rate $D(x+)/N(x+)$ (2)	Partial birth rate $N(x)/N(x+)$ (3)
5	0.0064	0.0381
10	0.0070	0.0391
15	0.0082	0.0381
20	0.0095	0.0401
25	0.0111	0.0426
30	0.0130	0.0434
35	0.0153	0.0482
40	0.0185	0.0542
45	0.0224	0.0565
50	0.0274	0.0622
55	0.0342	0.0662
60	0.0413	0.0774
65	0.0546	0.1044
70	0.0680	0.0955

It is important at this point to examine figure 14 with some care. Ideally, as explained in the general introduction to this method, the points should lie approximately on a straight line. Deviations from linearity are, however, inevitably present in empirical data. The possible causes of such deviations are varied and the investigator should consider all the relevant ones. In the case of females in El Salvador in 1961, the points show a definite linear trend, though the last point (for age 70) deviates markedly from the linear trend suggested by the others. As for the strange pattern displayed by the first three points, it is frequently encountered in empirical data and is associated with distortions of the values of $N(x)$ brought about by age-misreporting.

Examination reveals that the data at hand are far from perfect. Aside from the possible lack of stability of the population being analysed, errors in age-reporting are undoubtedly present. Yet, the relative linearity of the central segment of the plotted points *still* warrants an attempt to obtain a rough estimate of the completeness of death registration. This parameter is calculated in the next step.

Figure 14. Plot of partial birth rates, $N(x)/N(x+)$, against partial death rates, $D(x+)/N(x+)$, for females, El Salvador, 1961



Step 6: selection of a best fitting line. An examination of the points shown in figure 14 suggests that all of them, except the last two (those for age 65 and over and age 70 and over), lie around an approximate straight line. A straight line is therefore fitted to the remaining 12 points, from ages 5 to 60. The points are divided into two equally sized groups, one comprising points for ages ranging from 5 to 30, the other from 35 to 60. The mean abscissa and ordinate values for each group are calculated by summing the observations for each group and dividing the sum by the number of observations in each group. Thus, the mean of the first group of partial death rates $D(x+)/N(x+)$, denoted by X_1, is

$$X_1 = (0.0064 + 0.0070 + 0.0082 + 0.0095 + 0.0111 + 0.0130)/6.0 = 0.0092.$$

The groups and their means are shown in table 131.

TABLE 131. FITTING OF A STRAIGHT LINE BY GROUP MEANS, FEMALES, EL SALVADOR, 1961

Age x (1)	Partial death rate $D(x+)/N(x+)$ (2)	Partial birth rate $N(x)/N(x+)$ (3)
	(a) *Group 1*	
5	0.0064	0.0381
10	0.0070	0.0391
15	0.0082	0.0381
20	0.0095	0.0401
25	0.0111	0.0426
30	0.0130	0.0434
TOTAL	0.0552	0.2414
MEAN	$X_1 = 0.0092$	$Y_1 = 0.0402$
	(b) *Group 2*	
35	0.0153	0.0482
40	0.0185	0.0542
45	0.0224	0.0565
50	0.0274	0.0622
55	0.0342	0.0662
60	0.0413	0.0774
TOTAL	0.1591	0.3647
MEAN	$X_2 = 0.0265$	$Y_2 = 0.0608$

Once the means of each group have been computed, the slope of the fitted line is calculated according to the equation

$$K = (Y_2 - Y_1)/(X_2 - X_1) = (0.0608 - 0.0402)/(0.0265 - 0.0092) = 1.191.$$

The value of K implies that the completeness of death registration is $C = 1.0/K = 0.840$, or 84 per cent, and taking into account the slight adjustment necessary for the difference between the census date and the mid-year (see step 2 of subsection B.4), the final estimate of completeness is $\hat{C} = 0.840/1.0044 = 0.836$. Once more, in this case the necessary adjustment is so small that for all practical purposes it can be ignored.

From the estimate of K, the corresponding value of the growth rate r can be obtained as follows:

$$r' = Y_1 - KX_1 = 0.0402 - 1.191(0.0092) = 0.0292.$$

Note that this rate of growth is almost identical to that of 0.0287 obtained in subsection B.4 from census data.

Even though the estimate of K obtained above is probably the best possible, as an example another is computed by fitting a robust line. This exercise will serve to show that the use of trimmed means reduces the influence of extreme points on the estimated value of K.

Table 132 illustrates the way in which trimmed means are computed. All the observations have been used, but three have been "trimmed" at each end of the age range. The weights associated with each observation are listed in column (2). Note that the data set has been divided into groups of equal size and that the weights are symmetrical with respect to the centre. The line labelled "Weighted total" shows under column (2) the sum of the weights used, and under columns (3) and (4), the sum of the weighted observations. Thus, for example,

$$0.0602 = (0.25)(0.0064) + (0.50)(0.0070) + (0.75)(0.0082)$$

$$0.0095 + 0.0111 + 0.0130 + 0.0153,$$

and

$$0.2320 = (0.25)(0.0381) + (0.50)(0.0391) + (0.75)(0.0381)$$

$$+0.0401 + 0.0426 + 0.0434 + 0.0482.$$

TABLE 132. ESTIMATION OF AN ADJUSTMENT FACTOR, K, BY USE OF TRIMMED MEANS, FEMALES, EL SALVADOR, 1961

Age x (1)	Weights (2)	Partial death rate $D(x+)/N(x+)$ (3)	Partial birth rate $N(x)/N(x+)$ (4)
		(a) *Group 1*	
5.............	0.25	0.0064	0.0381
10.............	0.50	0.0070	0.0391
15.............	0.75	0.0082	0.0381
20.............	1.00	0.0095	0.0401
25.............	1.00	0.0111	0.0426
30.............	1.00	0.0130	0.0434
35.............	1.00	0.0153	0.0482
Weighted total..........	5.50	0.0602	0.2320
Weighted mean........		$X_1^* = 0.0109$	$Y_1^* = 0.0422$
		(b) *Group 2*	
40.............	1.00	0.0185	0.0542
45.............	1.00	0.0224	0.0565
50.............	1.00	0.0274	0.0622
55.............	1.00	0.0342	0.0662
60.............	0.75	0.0413	0.0774
65.............	0.50	0.0546	0.1044
70.............	0.25	0.0680	0.0955
Weighted total..........	5.50	0.1778	0.3732
Weighted mean........		$X_2^* = 0.0323$	$Y_2^* = 0.0679$

The "totals" for the second group are obtained in exactly the same way. Once these totals are computed, the means are obtained by dividing the sum of the weighted observations by the sum of the weights, that is, the totals under columns (3) and (4) by the total under column (2). For example,

$$X_1^* = 0.0602/5.5 = 0.0109.$$

The X^* and Y^* values are the trimmed means of the grouped observations. From them, the calculation of estimates of K, C and r is straightforward:

$$K^* = (Y_2^* - Y_1^*)/(X_2^* - X_1^*)$$

$$= (0.0679 - 0.0422)/(0.0323 - 0.0109) = 1.201;$$

$$C^* = 1.0/K^* = 0.0833;$$

$$r^* = Y_1^* - K^* X_1^* = 0.0422 - 1.201(0.0109) = 0.0291.$$

Note that the estimates obtained for each of these three parameters by using trimmed means are very similar to those obtained earlier using simple means. The rounded value of C^* still implies that adult female death registration in El Salvador was only 83 per cent complete in 1961 (disregarding for the moment the adjustment for census date). The value of r^* is virtually identical to r'. It is also interesting to recall that the use of the Preston-Coale procedure described in subsection B.3 led to an estimate of coverage of 0.825 using a 75+ open-ended interval and a growth rate estimate of 0.0287, again disregarding the adjustment for census date (see step 4 of subsection B.4). The similarity of these estimates is reassuring.

Step 7: adjustment of death rate. To obtain an adjusted death rate or adjusted age-specific mortality rates, one simply multiplies the observed rates by K or by K^*. Because the trimmed-mean estimate is somewhat more robust, it has been preferred. For example, for the death rate over age 5,

$$d(5+) = K^*[D(5+)/N(5+)] = 1.201(0.0064) = 0.0077.$$

TABLE 133. VALUES OF POPULATION AT DIFFERENT AGES, REPORTED DEATHS AND PARTIAL BIRTH AND DEATH RATES FOR MALES, ANDRA PRADESH STATE, INDIA, 1970-1971

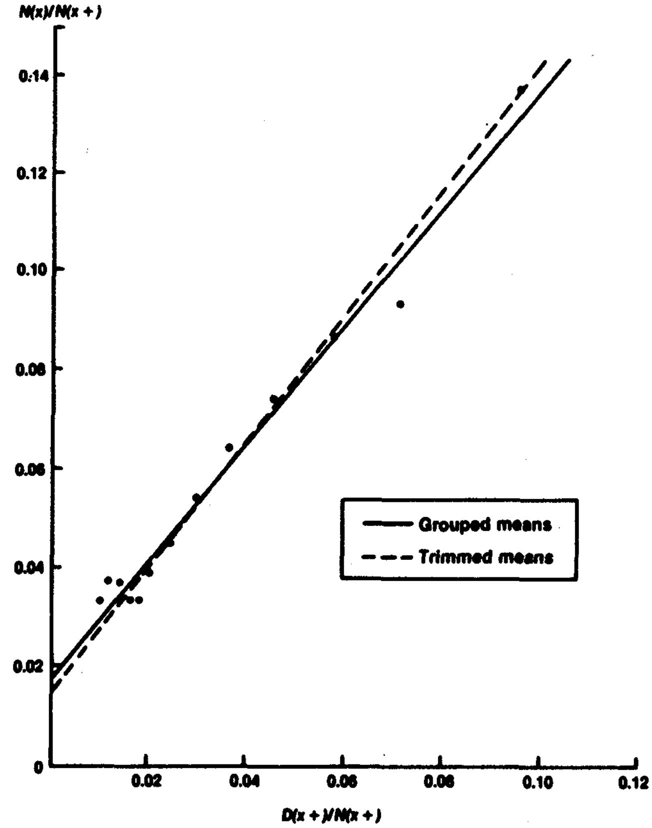
Age x (1)	Population at exact age x $N(x)$ (2)	Population aged x and over $N(x+)$ (3)	Reported deaths over age x $D(x+)$ (4)	Partial birth rate $N(x)/N(x+)$ (5)	Partial death rate $D(x+)/N(x+)$ (6)
5.............	569.3	17 111	187.10	0.0333	0.0109
10.............	524.5	14 307	173.22	0.0367	0.0121
15.............	432.0	11 866	167.97	0.0364	0.0142
20.............	334.2	9 987	163.50	0.0335	0.0164
25.............	286.7	8 524	157.79	0.0336	0.0185
30.............	275.9	7 120	148.72	0.0388	0.0209
35.............	260.2	5 765	142.96	0.0451	0.0248
40.............	244.6	4 518	134.61	0.0541	0.0298
45.............	212.4	3 319	122.32	0.0640	0.0369
50.............	177.3	2 394	110.38	0.0741	0.0461
55.............	133.3	1 546	90.59	0.0862	0.0586
60.............	99.0	1 061	76.36	0.0933	0.0720
65.............	75.9	556	53.69	0.1365	0.0966

5. *Second detailed example*

Andra Pradesh provides an interesting example of what happens to the application of the Brass growth balance method when age is poorly reported. The basic data have already been presented in table 127. The values of $N(x)$, $N(x+)$, $D(x+)$, $N(x)/N(x+)$ and $D(x+)/N(x+)$ are shown in table 133. The values of $N(x)/N(x+)$ and $D(x+)/N(x+)$ are plotted in figure 15.

The first five points are a trendless jumble, but the points from age 30 upward follow something like a straight line, although if the last point were excluded, the points for 30-60 would look as if they were following a curve. A straight line can be fitted to the points from 30 on; the line, fitted by group means, has a slope of 1.169 and an intercept of 0.0177, indicating that adult deaths were underreported by nearly 14 per cent. However, such a line is strongly affected by the last point and also by the exclusion of the first five points; it might, therefore, represent a case where trimming would be very effective. Following the previous example, using all available points, but weighting the first and last by 0.25, the second and second from last by 0.50, and the third and third from last by 0.75, gives a rather different straight line with a slope of 1.257 and an intercept of 0.0144; thus, the robust line indicates a rather lower growth rate and a level of omission of some 20 per cent.

Figure 15. Plot of partial birth rates, $N(x)/N(x+)$, against partial death rates, $D(x+)/N(x+)$, for males, Andra Pradesh State, India, 1970-1971



Of the two, the robust line should probably be preferred, but not a great deal of confidence can be attached to the results of this method when the derived points do not approximate a straight line. One feature that is not prominent in figure 15 is evidence of exaggeration of age at death; such exaggeration should lead to a tendency for the plotted points to curve off to the right, because the partial death rates over each age would increase more rapidly than the partial birth rates. The points from 30 to 60 do show something akin to such a pattern, but the point for age 65 and over fails to confirm the continuation of the pattern to higher ages. In general, exaggeration of age at death is likely to result in overestimates of completeness, since the points for higher ages would tend to pull the slope down.

In conclusion, it is difficult to interpret the results of applying the Brass method in this case. It may be mentioned in passing that in this application, the results of the Brass method are rather different from the ones obtained by applying the Preston-Coale method, which in subsection B.5 gave an estimate of coverage of nearly 90 per cent and a growth rate of 0.0183. Thus, this example should be taken as a warning against the mechanical application of any method. In applying the Brass procedure, careful examination of the behaviour of the plotted points is essential.