Estimating the Completeness of Death Registration

S. PRESTON AND K. HILL*

INTRODUCTION

There are increasing indications that mortality is declining more slowly in many less developed countries and in some cases declines have even ceased. In most cases they seem at least to have fallen short of expectations. Such tendencies are only dimly perceived, however, because of well-known inadequacies in statistical reporting systems. For many countries, all we know is child mortality, estimated from reports of mothers regarding surviving and dead children. Adult mortality estimates can sometimes be based upon registered deaths, but often the degree of registration completeness is unknown.

Brass[1] has proposed a useful procedure to estimate the degree of underregistration of adult deaths in populations where assumptions of stability can be reasonably maintained. The approach relies upon a comparison of the age distribution of the population with that of deaths. In this paper, we present two alternative methods for estimating the extent of the underregistration of deaths. One is, like Brass's approach, based upon assumptions of stability. However, it uses an estimate of the growth rate, rather than the age distribution of the population, as the principal basis of estimation. In logic it is similar to Carrier's procedure for estimating mortality from skeletal remains.[2] In application it is more straightforward than Brass's approach. The second technique is more tedious to apply but does not depend on the assumption of stability. It is based upon intercensal comparisons of successive cohorts, with an accounting for registered intercensal cohort deaths. A by-product of this second technique is an estimate of the relative completeness of enumeration of the two censuses.

The two approaches will be applied to statistics for Thailand. It will be shown that they produce results that are consistent between themselves and with an independent estimate of registration completeness.

In both of these new approaches, as well as in the earlier method proposed by Brass, the age distribution of deaths serves the function more conventionally provided by a set of model life tables. It specifies the age pattern of mortality; once this is known, it is necessary only to solve analytically for the level of mortality. When the 'model' is taken directly from registered deaths, the level of mortality can be interpreted in terms of the completeness of death registration, so long as this completeness is constant at all ages.

ESTIMATING THE COMPLETENESS OF DEATH REGISTRATION FOR A STABLE POPULATION OF KNOWN GROWTH RATE

Notation

$D(a)$ = recorded deaths at age a

D = total recorded deaths

* Kenneth Hill is Senior Research Associate with the Committee on Population and Demography of the U.S. National Academy of Sciences. Samuel Preston is Professor of Sociology, Population Studies Center, University of Pennsylvania.

[1] W. Brass, *Methods for Establishing Fertility and Mortality from Limited and Defective Data.* Laboratories for Population Statistics. University of North Carolina. Chapel Hill, 1975. Cf. also Y. Courbage and P. Fargues. 'A Method for Deriving Mortality Estimates from Incomplete Vital Statistics' *Population Studies* 33 (1). 1979 pp. 165–180.

[2] N.H. Carrier, 'A Note on the Estimation of Mortality and Other Population Characteristics, Given Deaths by Age.' *Population Studies* 21 (2) November, 1958 pp. 149–163.

$\hat{B}$ = true births in the year or period for which $D(a)$ and D are recorded
r = true annual rate of population growth
K = factor by which recorded deaths should be inflated to equal true deaths (assumed constant for all ages)
C = completeness of death registration = $1/K$
N = population size in the year or period for which $D(a)$ and D are recorded
$N(a)$ = number of persons aged a in that year
$p(a)$ = probability of surviving from birth to age a under prevailing true mortality rates
$m(a)$ = true annual death rate of persons at exact age a
$d(a)$ = deaths at age a in the true life table (with unit radix) to which the population is subject $= p(a)m(a)$
$\delta(a)$ = proportion of recorded deaths that occurred at age a = $D(a)/D$
$\hat{D}(a)$ = true deaths at age a
e_0 = life expectancy at birth in the true life table
d = recorded crude death rate = D/N
$\hat{b}$ = true birth rate = B/N

DERIVATION

The number of persons at age a in the population, under the assumption of stability, will be equal to

$$N(a) = N \cdot \hat{b} \cdot e^{-ra} p(a).$$

The true number of deaths at age a will therefore equal

$$\hat{D}(a) = N \cdot \hat{b} \cdot e^{-ra} p(a) \cdot m(a). \tag{1}$$

We can simplify Equation (1) by noting that $p(a)m(a) = d(a)$ and that $N\hat{b} = \hat{B}$. Furthermore, if the degree of registration completeness is invariant with age, $\hat{D}(a) = K \cdot D(a)$. Therefore,

$$K \cdot D(a) = \hat{B} \cdot e^{-ra} d(a).$$

Re-writing, we have

$$d(a) = (K/\hat{B})(a) \cdot e^{ra}. \tag{2}$$

Since $d(a)$ is a probability density function,

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

Therefore,

$$K/\hat{B} \int_0^\infty D(a) \cdot e^{ra} da = 1,$$

or

$$K/\hat{B} = 1 \bigg/ \left(\int_0^\infty D(a) \cdot e^{ra} da \right) \tag{3}$$

We now write the rate of natural increase, r, as

$$r = (\hat{B} - \hat{D})/N = ((\hat{B}/K) \cdot K - KD)/N = [K((\hat{B}/K) - D)]/N$$

Rewriting,

$$K = rN/((\hat{B}/K) - D).$$

Substituting from (3), we have

$$K = rN \bigg/ \left(\int_0^\infty D(a) \cdot e^{ra} da - D \right),$$

or

$$K = (r/d) \Big/ \left(\int_0^\infty \delta(a) \cdot e^{ra} da - 1 \right) \tag{4}$$

or

$$C = (d/r) \left(\int_0^\infty \delta(a) \cdot e^{ra} da - 1 \right) \tag{5}$$

Equation (4) provides a very simple and direct way to estimate the extent of death under-registration. All elements required for the calculation are assumed to be known: the growth rate, usually measured on an intercensal basis, the recorded death rate, and the age distribution of recorded deaths. It can be shown that the quantity $\int_0^\infty d(a) \cdot e^{ra} da$ in a stable population is the ratio of the birth rate to the death rate. So it is easy to see upon simplification of (4) that when registration is complete and all statistics accurate, $K = 1$.

The estimated completeness of registration is exactly proportional to the recorded death rate. This characteristic assures that if r and the age distribution of deaths are estimated accurately, the completeness of coverage will be known precisely. There is, therefore, no need to show how the method works in a stable population where only a constant error in death registration completeness is simulated.

The more interesting question relates to the sensitivity of the estimates to errors in the estimate of r. The most straightforward way to evaluate this is by differentiating Equation (4) with respect to r, yielding

$$dK/K = (dr/r) \cdot (1 - b \cdot e_0) \tag{6}$$

The proportionate error in K is equal to the proportionate error in r multiplied by a factor which may be interpreted as the negative value of the momentum factor of population growth: the factor by which the population would grow if births remained constant at the current level, with constant mortality. This factor, $[b \cdot e_0 - 1]$, has been computed for major regions of the world as 1.162 in Africa, 0.584 in East Asia, 1.092 in South Asia, 1.292 in Latin America, and 0.952 for less developed countries as a whole.[3] Because this factor is nearly unity for a typical developing country, the proportionate error in the estimate of K should be nearly equal (but opposite in sign) to the proportionate error in the estimate of r.

It is not surprising that knowledge of the age distribution of deaths and of the population growth rate is sufficient to estimate the completeness of death registration. The actual age distribution of deaths in a stable population simply compounds the age distribution of deaths in the stationary or life table, population ($d(a)$) and the population growth rate which transforms the stationary age distribution into a stable one. The problem is solved once the growth rate can be introduced to transform the actual age distribution of deaths back to the distribution in the corresponding stationary population, since all other mortality functions can be derived from $d(a)$.

It is interesting to note that the birth rate estimate that can be derived from estimates of r, d, and the completeness of death registration is not nearly as sensitive to error in the assumed intercensal growth rate as is the estimated growth rate or the death rate. The reason is in part that the birth rate is the sum of these two components whose proportionate errors are approximately equal but opposite in sign, with the result that the two errors tend to cancel. If we designate the proportionate error in the estimate of r as ϵ, the effect on the estimated birth rate ($\hat{b}$) is given approximately by

$$\begin{aligned} \hat{b} - b &= [r(1 + \epsilon) + d(1 - \epsilon)] - (r + d) \\ &= (r - d)\epsilon \\ (\hat{b} - b)/b &= [1 - (2d/b)]\,\epsilon \end{aligned}$$

[3] S. Preston, 'An Elementary Approach to Population Momentum.' *Intercom*, Population Reference Bureau, 5 (3) March, 1977 pp. 8–9.

Since the factor multiplying a in this last formula must be less than unity, the proportionate error in the estimate of the birth rate must be less than that in the growth rate or death rate. In countries where the birth rate is about double the death rate, as tends to be the case in African countries and the Indian sub-continent, the estimated birth rate will be very insensitive to error in the estimated growth rate, so that the birth rate can be estimated within quite narrow bounds from the age distribution of deaths alone.

Finally, it should be noted that this approach makes absolutely no use of the recorded age distribution of the population, so that techniques for estimating birth and death rates that rely primarily upon age distributional analysis can be regarded as providing information that is partially or even completely independent. Estimates of the completeless of registration do depend on the accuracy of the population total N, since the reported death rate is used in both Equations 4 and 5. If the population is underenumerated, the completeness of death registration will be over-estimated by the reciprocal of enumeration completeness, but the estimate of the death rate will be unaffected, since the estimate of completeness of registration is relative to the enumerated population. Of course, it should be borne in mind that changes in enumeration completeness are likely to affect the stable growth rate estimate used in the method, thus introducing another source of error that may be quantitatively more important. It should also be stressed that the performance of the method depends both on the assumption of stability and upon the assumption that the extent of underregistration of deaths does not vary with age.

APPLICATION

In its discrete form Equation (5) may be written as

$$C = 1/K = (d/r)\left\{\sum_{a}^{\infty} \delta(a)e^{ra} - 1\right\} \tag{7}$$

where $\delta(a)$ is the proportion of deaths that occur to persons in the interval in which the average age at death is a. For all age groups except the first and the last, which will, of course, be open-ended, a may be taken as the mid-point of the age at death interval. Equation (7) applies also to the population above *any* age, and not simply to the population above age zero, so it also provides a method of estimating the completeness of adult death registration. Indeed, since the ultimate weighting of child deaths is rather low, and since the completeness of registration of infant and child deaths is often very different from that of adult deaths, it is preferable to interpret the results as estimates of the completeness of adult death registration. The value of a must in all cases refer to distance from the lowest age in the analysis, and $\delta(a)$ must refer only to the distribution of deaths above this age, so that in all cases $\sum_{a}^{\infty} \delta(a) = 1.00$.

The country chosen for a test of the technique is Thailand, in part because an independent estimate of the completeness of death registration is available. We will attempt to estimate death registration completeness for the period between the censuses of 1960 (April 25) and 1970 (April 1), a period of 9.934 years. Figures used in the final computations are shown in Tables 1 and 2.

It was pointed out above that a in the exponential term of Equation (7) should be the average age at which deaths in the age group occur. For the first age group 0 to 1, this age will lie below the mid-point of the group, 0.5, since such deaths are heavily concentrated in the first months of life. The exact figure used is not important, however, to the functioning of the method, and 0.3 may be taken as a standard value. The last, open-ended age group is more difficult to deal with, since the lower limit of the interval may vary for different data, and the numerical value chosen is more important for the results obtained since it can vary much more than in the case of the first interval. Models have been used in order to calculate average ages at death for a range of open intervals, the results are shown in Table 3. For any particular lower boundary of the open

Table 1. *Estimating the Completeness of Death Registration from the Age Structure of Deaths; Thailand*

	Enumerated Population Size		Average Annual Intercensal Growth Rate	Registered Intercensal Deaths*	Average Annual Intercensal Registered Death Rate**	$\sum_n \delta(a) \cdot e^{r(a+(n/2))} - 1$	Estimated Completeness of Death Registration
	April 25, '60 (1)	April 1, '70 (2)	(3)	(4)	(5)	(6)***	(6) × (5) ÷ (3)
Total	26,257,916	34,397,374	0.027181	2,277,260	0.007605	2.42104	0.677
Males	13,154,149	17,123,862	0.026549	1,244,673	0.008324	2.22277	0.697
Females	13,103,767	17,273,512	0.027811	1,032,587	0.006887	2.65573	0.658

Source: United Nations, *Demographic Yearbooks,* 1962, 1966, 1974.
* January 1, 1960–December 31, 1969
** Denominator is intercensal person-years derived by assuming a constant annual growth rate:

$$PY = \frac{N(1970) - N(1960)}{r}$$

*** Developed as shown in Table 2, based on deaths in 1965 only.

Table 2. *Calculations required to compute Column (6) in Table 1 for total population of Thailand*

Age interval	Proportion of recorded deaths that occurred in interval* $\delta(a)$ (1)	Mid-point of interval (a) (2)	$\delta(a) \cdot e^{ra}$
0	0.17420	0.3	0.17563
1–4	0.14740	3.0	0.15992
5–9	0.05148	7.5	0.06312
10–14	0.02935	12.5	0.04123
15–19	0.03026	17.5	0.04869
20–24	0.03233	22.5	0.05959
25–29	0.03333	27.5	0.07038
30–34	0.03683	32.5	0.08909
35–39	0.03906	37.5	0.10824
40–44	0.03967	42.5	0.12594
45–49	0.03993	47.5	0.14522
50–54	0.04434	52.5	0.18473
55–59	0.04757	57.5	0.22704
60–64	0.05323	62.5	0.29103
65–69	0.04977	67.5	0.31173
70–74	0.04873	72.5	0.34964
75–79	0.04007	77.5	0.32936
80–84	0.03080	82.5	0.29002
85 +	0.03165	88.46	0.35043
			$\Sigma = 3.42104$

* The distribution of deaths in 1965 is used from United Nations *Demographic Yearbook,* 1974. Deaths of unknown age are ignored.

Table 3. *Average Age at Death in the Open Age Interval by Lower Boundary, Rate of Growth, and Expectation of Life at Birth*

Lower Boundary of Open Interval	Rate of Population Growth (r)	Expectation of Life at Birth, e_0 70	60	50	40
85	0.01	88.8	88.6	88.5	88.5
	0.02	88.7	88.5	88.4	88.4
	0.03	88.6	88.4	88.3	88.3
80	0.01	85.7	85.2	84.9	84.8
	0.02	85.5	85.0	84.8	84.6
	0.03	85.4	84.9	84.6	84.5
75	0.01	83.2	82.2	81.7	81.4
	0.02	83.0	82.0	81.4	81.1
	0.03	82.7	81.7	81.2	80.9
70	0.01	81.3	79.7	78.8	78.2
	0.02	80.9	79.3	78.4	77.9
	0.03	80.5	79.0	78.1	77.6
65	0.01	79.7	77.5	76.2	75.4
	0.02	79.2	77.0	75.7	74.9
	0.03	78.7	76.6	75.3	74.5
60	0.01	78.5	75.8	74.0	72.8
	0.02	77.8	75.1	73.3	72.2
	0.03	77.1	74.4	72.7	71.5

interval in the data being analysed, the average age at death in the open interval can be obtained from Table 3 by using the observed rate of population growth, r, and an approximate value for the expectation of life at birth, e_0. The ages in Table 3 are more sensitive to e_0 than for r. In general, of course, e_0 will not be known precisely, although the approximate level of mortality of the population studied will normally be known. This approximate level can be used to enter the table, and its value can later be revised if the resultant life table shows a very different e_0.

The results in Table 1 suggest that, during the intercensal period, 69.7 per cent of male deaths and 65.8 per cent of female deaths were registered. It should be noted that the calculations for men and women are performed completely independently. Aggregating the deaths and populations for the two sexes and recomputing the coverage rate for the whole population from scratch gives an estimate of 67.7 per cent.

An independent estimate of underregistration of deaths is available for 1964–1965 from the Thailand Survey of Population Change. This is a somewhat frustrating source to use because it is replete with inconsistencies and errors. In adjacent paragraphs the completeness of death registration is given as 71 per cent and 63 per cent. The figure of 63 per cent is supported by examination of raw data presented in an Appendix (Table 5) for the survey area itself.[5] Completeness of registration nationally was probably somewhat higher than this, since the metropolitan area of Bangkok – Thonburi, in which death registration may have been of above average completeness, was excluded from the survey area. A figure of around 65 per cent may be guessed at. The 71 per cent estimate is apparently an error.[6]

Because our results (67.7 per cent) match those of the *Survey* (65 per cent) fairly closely, it is tempting to stop here and conclude that the technique works quite well. Indeed, even its estimate of the sex difference in completeness of coverage of four percentage points is close to the five percentage point difference in the *Survey* shown in Table 4. However, this congruence of estimates turns out to result from two compensating errors in our estimates: child deaths are almost certainly differentially underregistered; and the rate of population growth is almost certainly too low. Both these factors are suggested by the *Survey* itself. The estimated rate of natural increase was 0.031 in the *Survey* rather than the intercensal growth rate of 0.027. The difference between the two rates, as will be demonstrated below, seems largely attributable to differential underenumeration in the 1970 Census. The likelihood of differential underregistration with age is suggested in Table 4, reproduced after correction from the Survey Report.

Table 4. *Estimated completeness of death registration from the Thailand Survey of Population Change, 1964–65*

	All ages	Under 1	1–9	10–59	60+
Males	64.5*	50.4	69.6	70.8*	74.5*
Females	60.0	47.4	58.8	69.1	68.6

Source: Table K *op. cit.* in footnote 4. (Corrected from basic data where marked*)

There is also a tendency in Thailand to report age at next birthday rather than age at last birthday, at least for the living.[7] If age at death is distorted in the same way in the vital registration system, the values of a used will be half a year too large throughout, except in the first interval, in which ages are typically reported in months rather than years. Allowance for this error will result in a reduction in the estimated completeness of registration.

In the circumstances, it is desirable to re-apply the procedure using $r = 0.031$ and restrict analysis to the population above age 10, a range where age differentials in coverage appear minor. In this way, the problem of systematic age error is also avoided, since the new origin of age 10 is affected by the same half-year bias as the ages over 10. Results are shown in Table 5 for males.

[4] Thailand National Statistical Office, *The Survey of Population Change Report.* Bangkok. No Date (approx. 1969).

[5] Here, and in the remainder of this paragraph, table numbers refer to those of the *Report* cited in the previous footnote.

[6] If the deaths in Table 5 that were expected to have occurred (using the Chandrasekaran-Deming approach) but were recorded in neither system are omitted from the denominator, the ratio of registered to total recorded deaths is 0.713. Elsewhere in the Survey Report, however, these deaths are included. The "registered" crude death rate of 7.7 (cited in Table J) likewise appears to be an error, and should be 7.2.

[7] A. Chamratrithirong, N. Debavalya & J. Knodel, *Age Reporting in Thailand: Age at last Birthday versus Age at Next Birthday* Institute of Population Studies: Chulalongkorn University Bangkok. Paper No. 25, 1978.

Table 5. *Estimating the Completeness of Death Registration for Males Above Age 10 in Thailand*

Age Interval	Distance of Mid-point from Age 10	Proportion of Deaths in Interval $\delta(a)$	$\delta(a) \cdot e^{0.031 \cdot a}$
10–14	2.5	0.04804	0.05191
15–19	7.5	0.04995	0.06302
20–24	12.5	0.05056	0.07449
25–29	17.5	0.05051	0.08689
30–34	22.5	0.05602	0.11253
35–39	27.5	0.05990	0.14050
40–44	32.5	0.06381	0.17476
45–49	37.5	0.06930	0.22162
50–54	42.5	0.07824	0.29215
55–59	47.5	0.08298	0.36180
60–64	52.5	0.09183	0.46752
65–69	57.5	0.08325	0.49490
70–74	62.5	0.07618	0.52879
75–79	67.5	0.06038	0.48939
80–84	72.5	0.04018	0.38027
85 +	78.46	0.03886	0.44240
			$\Sigma = 4.38294$

$$C = \frac{d}{r} \cdot \left\{ \sum_{a}^{\infty} (a) \cdot e^{ra} - 1 \right\}$$

$$= \frac{0.007208^{*}}{0.0310} \times 3.38294 = 0.787$$

* Death rate 1960–69 for persons aged ten and over, obtained by taking base population of 1960, moving it to January 1, 1960, and allowing it to grow at annual rate of 0.031.

According to this new procedure, the completeness of registration above age ten is 0.787 for males and 0.700 for females. These figures are not very different from those for ages over ten of 0.72 for males and 0.69 for females given in the *Survey* and shown in Table 4, particularly when it is realized that these latter figures need to be increased somewhat in order to adjust the survey results to be nationally representative. It should be borne in mind that the sampling error in the *Survey* may have been quite large, since only 666 male and 501 female deaths were registered.

Thus, there are good reasons to conclude that the first technique performs reasonably well in estimating the extent of adult death underregistration, at least where such underregistration is invariant with age. The analysis presented below tends to confirm the range of estimated registration completeness. When applied to all ages, however, its effectiveness disappears. The estimated completeness based on $r = 0.031$ and the $\delta(a)$ function for all ages is 0.84 for males and 0.76 for females, figures that are probably too high by about ten percentage points.[8] The reason is simply that because of selective underregistration of child deaths the appearance is given that deaths are occurring at an older age than they really are. As a result, the inferred life table ($d(a)$) yields too high a life expectancy and implies that a higher fraction of deaths are being registered than is really the case. The technique thus should not be applied to deaths at all ages unless there are reasons to believe that underregistration of child deaths is approximately the same as that of adult deaths. It should also be pointed out that the age structure of deaths for 1965 was used throughout, whereas the death rates were calculated for the intercensal period. It would be more satisfactory, though more tedious, to use the age structure of all intercensal registered deaths instead.

Because no use is made of the recorded age distribution of the population, this distribution can be used to check the plausibility of results. Just as a growth rate estimate 'falls out' of Brass's approach, an age distribution 'falls out' of the present approach. If the growth rate estimate used is too high, the implied age distribution will be too young compared with the actual age distribution.

[8] Note that the proportionate difference in estimated completeness of coverage when $r = 0.031$ is used in place of the intercensal growth rate is close to the proportionate difference in r itself, as implied by Equation 6.

That an age distribution is implied by the approach is evident from the formula for the proportion of a stable population that is aged a:

$$c(a) = e^{-ra}p(a) \Big/ \int_0^\infty e^{-ra}p(a)da$$

If we substitute the equivalent expression $\int_a^\infty d(a)da$ for $p(a)$, simplify, and convert to discrete notation, the proportion of the implied stable population in the age interval centered on a^* will be

$$c(a^*) = \left[e^{-ra^*}\sum_{a^*}^{\infty}\delta(a)e^{ra}\right] \Big/ \int_a^\infty \left[e^{-ra}\sum_0^\infty \delta(a)e^{ra}\right] da$$

Figure 1 compares the actual age distribution of Thai males, 1970, with that implied by our procedures using $r = 0.031$ and the $\delta(a)$ function as recorded during 1960–69. It is clear that the correspondence is very close, especially after allowance for census age misreporting, the possibility of small birth cohorts during World World II, or selective underenumeration of men in their 20's.

The growth rate that falls out of Brass's procedure is the value of a linear function at the origin, or its intercept.[9] The slope of this line is the reciprocal of estimated registration completeness. If the implied growth rate is the same as that used in the present technique, the methods should yield identical results. This correspondence may not occur, however, because several lines drawn from the same intercept may fit the data almost equally well, or may represent the best fit, using different criteria of goodness of fit. That is, there is an arbitrary element in the fitting procedure that is not present when Equation (4) is used. Brass's technique *could* be used to identify a value of r that is then substituted into Equation (4); this approach would be indicated when the estimate of the intercept is believed to be more reliable than that of the slope, which may well be common because of the bunching of data points towards the origin in Brass's approach. Because of this bunching, Brass's slope (the reciprocal of estimated registration completeness) is highly dependent on observations at older ages, a dependence that appears to be somewhat reduced in the present approach. However, it would certainly be premature to conclude that the present approach is in general preferable to that proposed by Brass. Much will depend upon whether the growth rate is known more reliably (in the sense of sensitivity of results to error) than the age distribution. Some combination of the two approaches, perhaps of the kind just outlined, may prove to work better than either approach taken separately.

It is worth emphasizing that the technique can be applied with various initial ages, so that for any particular growth rate selected, a series of estimates of registration completeness is produced, one for each starting age. There is much to be gained from critically examining this series, since an erratic set of estimates will indicate the violation of underlying assumption or basic flaws in the input. For example, experience suggests that estimated completeness is usually greater when starting with age zero than with age five or ten. As noted above, this disparity is often plausibly ascribed to more deficient registration below the ages of five than at higher ages. This results in the distribution of deaths which begins at age zero being too old, implying better mortality conditions for a given r and hence that a higher fraction of actual deaths has been captured by the registration system. Similarly, large disparities above ages five or ten may lead to doubt about the assumptions of stability or equal completeness with age. Particularly troublesome would be a series that exhibits a distinct age trend. Such a trend is most likely to result either from an inappropriate choice of growth rate or a very serious violation of the stability assumption.

Use of too low a growth rate will lead to an underestimation of completeness, as was shown above. It was not obvious from the example, however, is that the sensitivity of estimated completeness to errors in r varies considerably with age. Estimated completeness is more sensitive

[9] Brass, *op. cit.*, in footnote 1, pp. 117–123.

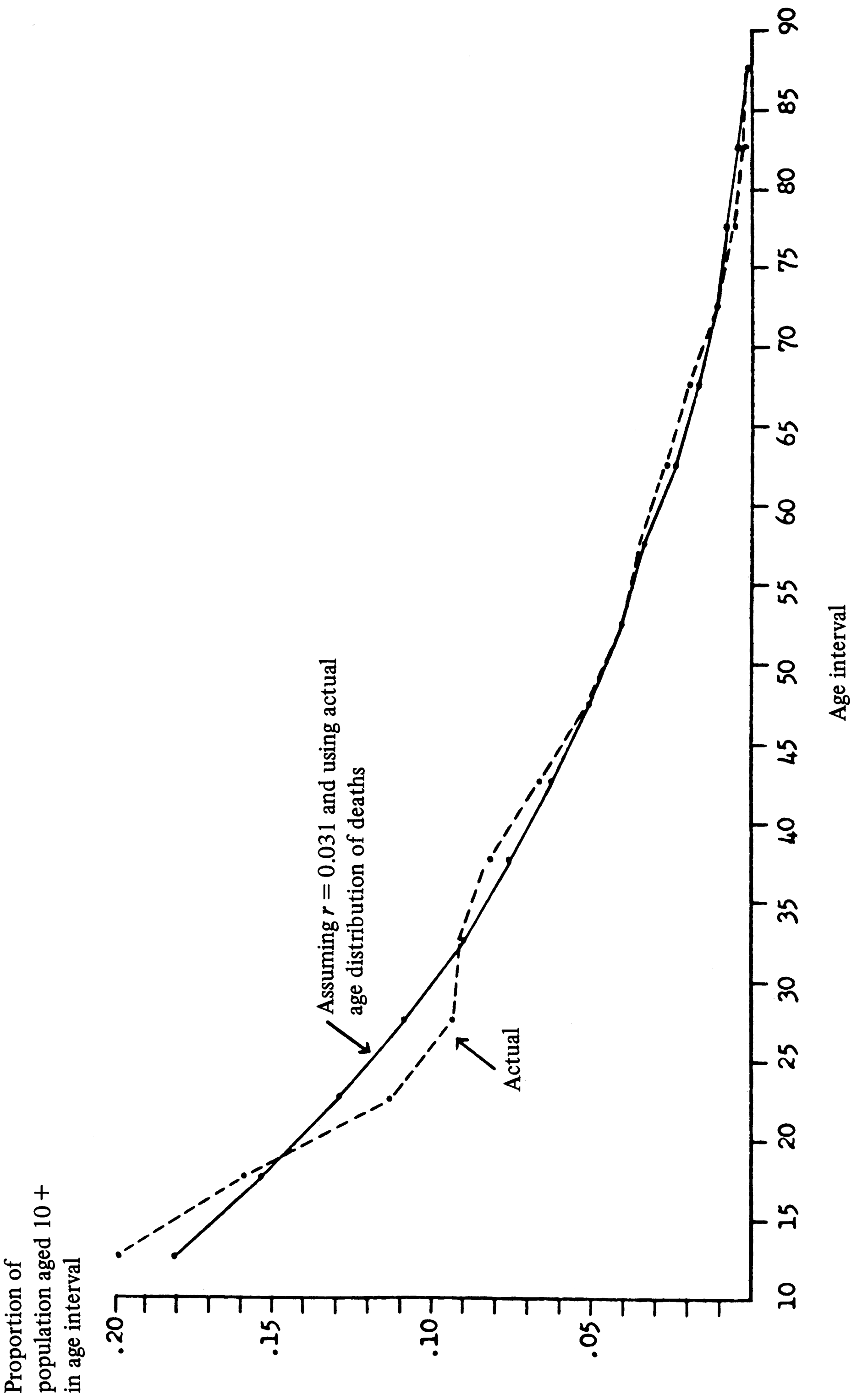

Figure 1. Comparison of actual and hypothetical age distributions of Thailand males, 1970

to errors in r at younger than at older ages. Operationally, the reason for this may be seen in the formula for C, in which r appears in two places. It forms the denominator of the entire expression, in which role its errors have the same proportional effect, regardless of the initial ages. In its other role it appears in the numerator in the e^{ra} series, where 'a' is the distance from the starting age. As this starting age increases, the distance of any particular age from it declines, and this errors in r are weighted less heavily. In terms of Equation (6) above, expressing the sensitivity of $K = 1/C$ to error in r, this sensitivity is shown to depend on the quantity $(1 - b_x \cdot e_x)$, where b_x is the birth rate for persons over age x (number of persons at age x divided by the number of persons above age x) and e_x is life expectancy at age x. This series declines sharply with age, as can be illustrated with figures from model stable populations:

Age (x)	$(1 - b_x \cdot e_x)$*
10	−0.917
30	−0.617
60	−0.253

* Estimates drawn from Coale-Demeny 'West' model male stable populations with growth rate of 0.025 and life expectancy at birth of 56.47 years.

A. J. Coale and P. Demeny, *Regional Model Life Tables and Stable Populations.* Princeton, 1966.

In other words, in this not extreme stable population, the sensitivity of completeness estimates to errors in r is 50 per cent greater age ten than at age 30, and more than 300 per cent greater at age ten than at age 60.

This differential sensitivity by age to errors in r can be used in a manner that adds considerable scope to the technique. Not only does it serve to indicate an error in the initial value of r, but it can be used to determine the appropriate value of r itself. That is, the user can start with different trial values of r and select the value (together with the implied completeness level) that results in the most level age-sequence of completeness estimates. Just such a procedure has been implemented for El Salvador by John Hobcraft. Because of their pertinence to the present discussion, the unpublished results are reproduced below:

Age range to compute completeness	Estimated completeness of death registration for El Salvador females, 1961, based upon three different assumed growth rates and using different initial ages*		
	$r = 0.025$	$r = 0.030$	$r = 0.035$
0 +	0.700	0.877	1.113
1 +	0.701	0.875	1.108
5 +	0.706	0.871	1.087
10 +	0.724	0.879	1.078
15 +	0.732	0.875	1.054
20 +	0.736	0.866	1.026
25 +	0.749	0.867	1.012
30 +	0.754	0.860	0.988
35 +	0.756	0.851	0.963
40 +	0.780	0.867	0.967
45 +	0.794	0.871	0.958
50 +	0.803	0.870	0.945
55 +	0.819	0.877	0.941
60 +	0.780	0.827	0.877

* United Nations Population Division, *Model Life Tables for Developing Countries: An Interim Report:* Working Paper No. 63. ESA/P/WP.63, January 1979, p. 38.

As suggested above, completeness estimates in El Salvador rise for each age as the value of r used in the computations increases, and their sensitivity to changes in r is greatest at the youngest ages. What is more striking about this example is the very nearly level age sequence of completeness estimates that is produced when using a value of $r = 0.030$. At this growth rate, the estimate of completeness is invariably between 0.85 and 0.88 for all starting ages between 0 and 55. When r is lower by 0.005, there is a persistent upward trend in estimated completeness between ages 0 and 55 (spanning a range of about 12 percentage points), and when r is higher by 0.005 there is a persistent downtrend of about 17 percentage points. Obviously, in this case the technique provides some rather decisive internal clues about the appropriate value of the growth rate for use in making the completeness estimates.

Thus, the technique may be used to solve simultaneously for a level of completeness and a level of the growth rate, exactly as Brass's technique is operated (though in the latter information on the age distribution of the population is used explicitly, whereas in our technique it is only used for purposes of verification). The solution pair would occur at a value of r that minimizes the age-trend in completeness estimates (as measured, for example, by the standard deviation of completeness estimates as the initial age varies from 5 to 40). If the growth rate is known with fairly high reliability, of course, this simultaneous-solution approach is not required, although it would still be wise to examine the age sequence of estimates for trend.

ESTIMATING THE COMPLETENESS OF DEATH REGISTRATION FROM INTERCENSAL COHORT SURVIVAL

Intercensal survival has long been recognized as a potentially useful technique by which to estimate adult mortality risks. In a closed population, survivors of a cohort at one census can be identified at the next, and ratios of standard life table functions representing the intercensal period may be derived. In practice, the technique has proved disappointing in applications to many developing countries, partly because of the distorting effects of migration and age misreporting, but also because of variations in completeness of enumeration between one census and another. However, the availability of information on registered deaths, though incomplete, allows the use of a procedure that can cope with changes in the completeness of enumeration and omission of deaths from the vital registration system at the same time. The procedure is based on the simple notion that, in a closed population, experiencing perfect reporting, a cohort at the first census will be equal to the sum of the survivors of the same cohort at the second census and the deaths occurring to cohort members during the intercensal period.

Notation

$P_1(a,n)$ = true population of age group $a, a+n$ at time of first census

$P_2(a+i,n)$ = true population of age group $a+i, a+i+n$ at time of second census

$D(a,n)$ = true deaths occurring during the intercensal period in the true cohort aged a, $a+n$ at time of first census

$\bar{P}_1(a,n)$ = enumerated population of age group $a, a+n$ at first census

$\bar{P}_2(a+i,n)$ = enumerated population of age group $a+i, a+i+n$ at second census

$\bar{D}(a,n)$ = registered deaths during intercensal period to cohort aged a, $a+n$ at time of first census

i = intercensal period in years

$K_1(a,n)$ = completeness of enumeration of population aged a, $a+n$ at first census = $\bar{P}_1(a,n)/P_1(a,n)$

$K_2(a+i,n)$ = completeness of enumeration of population aged $a+i, a+i+n$ at second census $=\bar{P}_2(a+i,n)/P_2(a+i,n)$

$C(a,n)$ = completeness of intercensal death registration for cohort aged a, $a+n$ at first census $=\bar{D}(a,n)/D(a,n)$

In a closed population, the size of a cohort at the first census will equal the survivors at the second census plus the intercensal cohort deaths. Thus

$$P_1(a,n) = P_2(a+i,n) + D(a,n)$$

In terms of observed numbers, this may be rewritten as

$$\frac{\bar{P}_1(a,n)}{K_1(a,n)} = \frac{\bar{P}_2(a+i,n)}{K_2(a+i,n)} + \frac{D(a,n)}{C(a,n)} \tag{8}$$

All the numerators in Equation (8) are known, but none of the denominators. However, multiplying out by $K_1(a,n)$ and dividing through by $\bar{P}_2(a+i,n)$ gives

$$\frac{\bar{P}_1(a,n)}{\bar{P}_2(a+i,n)} = \frac{K_1(a,n)}{K_2(a+i,n)} + \frac{K_1(a,n)\bar{D}(a,n)}{C(a,n)\bar{P}_2(a+i,n)} \tag{9}$$

Thus, the ratio of the enumerated cohort size at the first census to its size at the second census is equal to the ratio of the completeness of enumeration plus the ratio of registered deaths to cohort size at the second census multiplied by a factor equal to the completeness of enumeration at the first census and divided by the completeness of cohort death registration. If it can be assumed that completeness of enumeration and of death registration are invariant with age, values of $\bar{P}_1(a,n)/\bar{P}_2(a+i,n)$ and $\bar{D}(a,n)/\bar{P}_2(a+i,n)$ for successive cohorts will lie on the straight line with an intercept equal to the completeness of enumeration of the first census relative to that of the second and slope equal to the completeness of enumeration of the first census relative to that of intercensal death registration. Solving for this slope and intercept by examining observations at different ages will thus provide estimates both of the completeness of death registration and differences in completeness of census enumeration.

APPLICATION

Equation (9) can refer to any initial cohort whatever, and, therefore, the starting point may be a population divided into five-year age groups, or the population over given ages, or the populations over successive ages but under a given age. Each possibility has its own strengths and weaknesses. Working with individual five-year age groups will tend to show up deviations from the assumption of constant underregistration and underenumeration by age, but will also be affected by any net transfers across age group boundaries as a result of age misreporting. The use of cumulated data, starting with successive initial cohorts aged 0+, 5+, 10+, and so on, limits the effects of age misreporting to net transfers across the lower boundary, and will clearly result in a much smoother set of values, but will give little warning of deviations from the assumptions of constant underenumeration and underregistration by age. The use of semi-cumulated data, starting with successive cohorts aged five to 65, ten to 65, 15 to 65, and so on, concentrates on a section of the age distribution for which the assumption of constancy is probably strongest, and limits some of the fluctuations that may occur as a result of age misreporting. Since there seem to be no *a priori* reasons for preferring one way to another, all three have been applied to the case of Thailand males between 1960 and 1970.

The basic figures, in the form of five-year age distributions for 1960 and 1970, and deaths for five-year age cohorts in 1960 registered in the years 1960 to 1969, are shown in Table 6. Both the 1960 and 1970 age distributions have been adjusted slightly so as to refer to 1st January of those years; this adjustment was made by applying the observed intercensal growth rate of 2.65 per cent to the reported age distributions, after distributing the small number of those of not stated age proportionately. The 1970 age distribution was further adjusted to allow for the fact that it was based on a question about date of birth, unlike the 1960 census and death registration, for which age was obtained in completed years. Following Chamratritnirong *et al.*,[10] ages in 1970

[10] *op. cit.* in footnote 7.

Table 6. *Male Population by Age Group, 1960 and 1970, and Intercensal Cohort Deaths; Thailand*

Age Group in 1960	Male Population 1/1/1960[1]	1/1/1970[2]	Cohort Deaths 1960–1969[3]
0–4	2,120,461	2,336,986	157,076
5–9	1,996,402	1,869,905	44,109
10–14	1,550,462	1,351,468	35,774
15–19	1,252,688	1,105,860	34,969
20–24	1,202,393	1,047,916	37,321
25–29	1,016,540	960,281	40,104
30–34	876,955	787,195	44,246
35–39	686,616	610,710	47,026
40–44	564,136	479,987	53,119
45–49	489,817	394,452	60,603
50–54	398,460	307,676	66,257
55–59	319,083	220,202	69,503
60–64	226,762	135,647	65,675
65 +	323,766	130,484	167,180

[1] From 1960 age distribution, 'not stateds' distributed, moved back to 1 January 1960 using fixed observed intercensal growth rate.
[2] From 1970 age distribution, not stateds distributed, moved back to 1 January, 1970 using fixed intercensal growth rate, and adjusted for half-year change in age reporting.
[3] From registered deaths, 1960 to 1969, distributed not stateds, and weights given in text.

were taken as being half a year younger than they would have been had they been based on a question asking age in completed years, and new age groupings corresponding to 0 to $4\frac{1}{2}$, $4\frac{1}{2}$ to $9\frac{1}{2}$, $9\frac{1}{2}$ to $14\frac{1}{2}$, and so on were constructed in order to make the 1970 age classification more consistent with that used in the other sources.

Cohort deaths were obtained from deaths by calendar year and five-year age group on the assumption that in each age group deaths were distributed evenly by age. Weights were applied in order to apportion deaths between cohorts. For example, the deaths for the cohort aged 5 to 9 at the beginning of 1960 were obtained as

$$\begin{aligned} {}_5D_5 &= 0.9\,{}_5d_5^{1960} + 0.1\,{}_5d_{10}^{1960} + 0.7\,{}_5d_5^{1961} + 0.3\,{}_5d_{10}^{1961} \\ &\quad + 0.5\,{}_5d_5^{1962} + 0.5\,{}_5d_{10}^{1962} + \ldots \\ &\quad + 0.1\,{}_5d_{10}^{1969} + 0.9\,{}_5d_{15}^{1969} \end{aligned}$$

It is at this stage that the method becomes rather tedious, though as compensation the calculations are very simple.

Table 7 shows the required ratios of $\bar{P}_1(a, n)/\bar{P}_2(a+i, n)$ and $\bar{D}(a, n)/\bar{P}_2(a+i, n)$ for cohorts formed from initial five-year age groups, from the initial population over successive ages, and from the initial population over successive ages, but under 65. The ratios are plotted against each other in the three parts of Figure 2. Straight lines have been fitted to the points using a trimmed means procedure by which the points were divided into two equal groups, the mean coordinates of each group were found after progressively underweighting the extreme three points in each direction, and the slope and intercept of the straight line passing through the two mean points were calculated.[11] Table 7 shows the slopes and intercepts obtained, and Figure 2 shows the lines. The intercept represents the ratio of completeness of enumeration at the first census to that at the second, whereas the slope represents the ratio of completeness of enumeration at the first census to the completeness of intercensal death registration.

Figure 2 shows that in all three cases, the ratios fall close to a straight line even in the first case when the data are classified by simple five-year age groups; in both the cumulated

[11] The procedure is described by D. F. Andrews et al., *Robust Estimates of Location: Survey and Advances.* Princeton, Princeton University Press, 1972.

Table 7. *Ratios of Initial Cohort to Cohort Survivors and Cohort Deaths to Cohort Survivors; Males, Thailand, 1960 to 1970*

Lower Age Boundary of Initial Cohort in 1960 (a)	Cohort Width (= n): 5		$w-a$		$65-a$	
	$\frac{P_1(a,n)}{P_2(a+10,n)}$	$\frac{P_1(a,n)}{D(a,n)}$	$\frac{P_1(a,n)}{P_2(a+10,n)}$	$\frac{P_1(a,n)}{D(a,n)}$	$\frac{P_1(a,n)}{P_2(a+10,n)}$	$\frac{P_1(a,n)}{D(a,n)}$
0	0.907	0.0672	1.110	0.0786	–	–
5	1.068	0.0236	1.160	0.0815	1.141	0.0646
10	1.147	0.0265	1.183	0.0958	1.160	0.0749
15	1.133	0.0316	1.190	0.1110	1.163	0.0858
20	1.147	0.0356	1.203	0.1283	1.169	0.0979
25	1.059	0.0418	1.217	0.1524	1.175	0.1146
30	1.114	0.0562	1.267	0.1871	1.213	0.1384
35	1.124	0.0770	1.320	0.2323	1.250	0.1686
40	1.175	0.1107	1.392	0.2891	1.299	0.2049
45	1.242	0.1536	1.479	0.3612	1.356	0.2477
50	1.295	0.2153	1.597	0.4642	1.423	0.3036
55	1.449	0.3156	1.788	0.6217	1.534	0.3799
60	1.672	0.4842	2.069	0.8750	1.672	0.4842
65	2.481	1.2812	2.481	1.2812	–	–
Fitted straight line:						
Intercept	1.060		1.054		1.047	
Slope	1.143		1.163		1.248	

cases, the points are very close to a straight line. The figures at the bottom of Table 7, however, show that the straight lines fitted have rather different implications. The estimates of relative completeness of enumeration of the two censuses do not differ greatly, showing that the 1960 Census is more complete than that of 1970, the relative undernumeration of males in the latter ranging from four per cent to six per cent. The estimates of completeness of registration of intercensal deaths show much more variation however, ranging from 80 per cent, relative to the completeness of the 1960 Census, to 87 per cent. The former figure, however, includes only events to the cohort aged 5 to 65 in 1960, whereas the latter includes events to the entire 1960 population.

The three ways of manipulating the data yield somewhat different results, and the question remains how to make a choice between them. A look at the points for individual are groups may be expected to provide information about the suitability of the assumptions of constant rates of omission by age. One point stands out clearly, that for the age groups 0 to 4 in 1960, which is found to be smaller than the adjusted cohort aged 10 to 14 in 1970; this is a result of the habit of reporting age at nearest birthday rather than at last birthday, and perhaps also a result of relative underenumeration of young children in 1960. The adjustment of the 1970 age distribution by half a year downwards did not eliminate the problem for the first age group, since in 1960 the first age group is only 4½ years wide, whereas the corresponding cohort group is five years wide in 1970.

It may then be noticed that the points for three early cohorts are clearly above the line, indicating surprisingly small cohorts in 1970, corresponding to cohorts aged 10 to 24 in 1960, or 20 to 34 in 1970; these are prime age groups for undernumeration of males, and the supposed underenumeration in 1970 may have been relatively more serious at those ages. The point for the cohort aged 25 to 29 in 1960, 35 to 39 in 1970, is below the line, indicating a larger than expected 1970 cohort. Since death rates are low at these ages, almost all deviations are likely to arise from differences in enumeration completeness or age reporting errors rather than death registration completeness.

For cohorts aged 30 and over in 1960, the points all lie close to the line, tending, however,

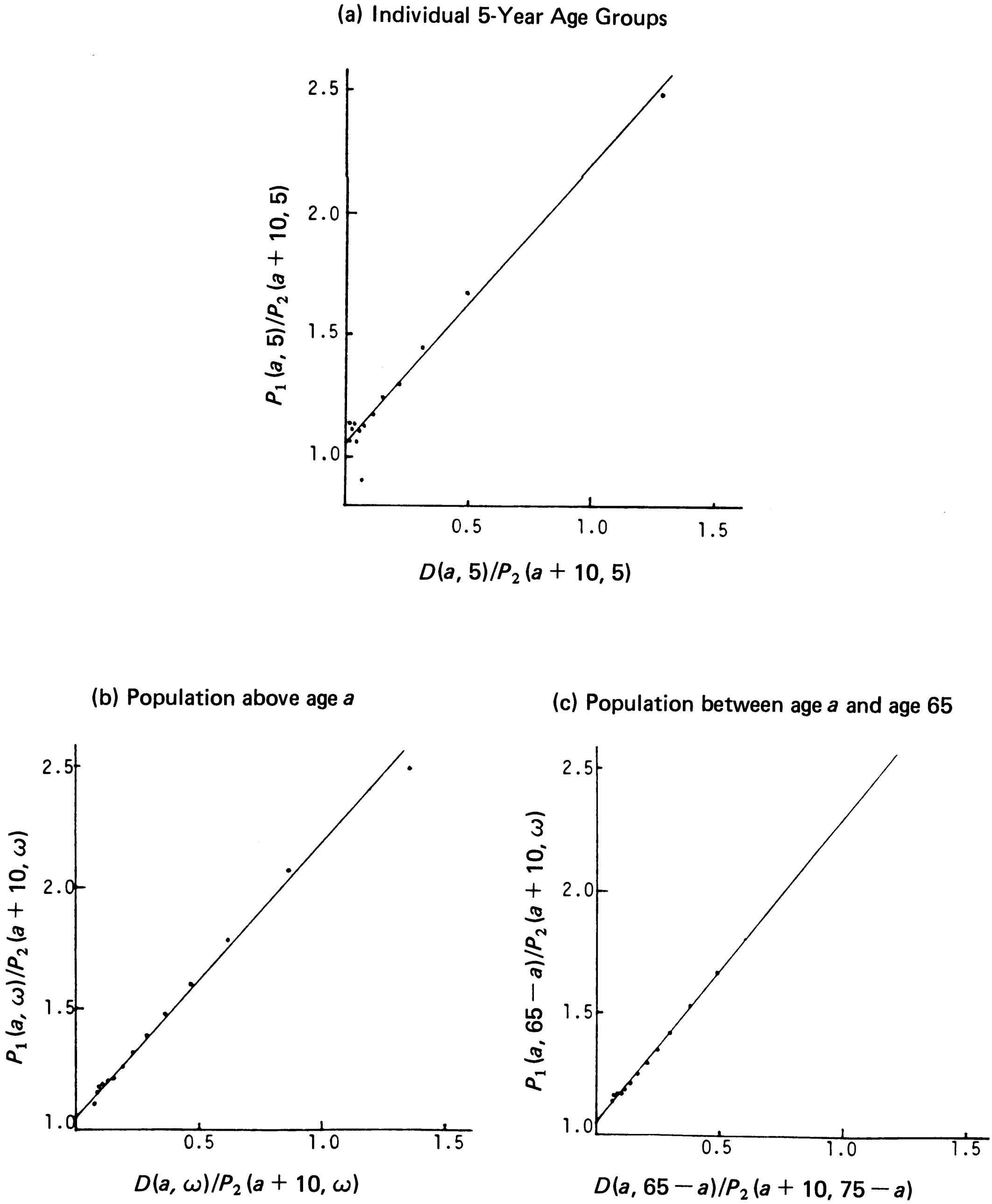

Figure 2. Plots of Cohort Size Ratios, $P_1(a, n)/P_2(a + i, n)$, Against Cohort Death Ratios, $D(a, n)/P_2(a + 10, n)$, for Different Types of Cohorts

to indicate a steeper slope up to age 65 than later, with the last point, for the cohort 65 and over, falling below the line. For these later points, completeness of both deaths and enumeration is important. The point for 65 and over may be lower than expected as a result of more complete reporting of deaths of old people, or systematic exaggeration of the age at death, or simply a tendency to shove difficult cases into an open-ended age category.

There is obviously good reason to exclude the first point from analysis, because of problems with the 1960 age distribution, and it may be that, because of the apparently lower omission of deaths over 65, the open-ended interval should also be omitted. The results from the limited cumulation from 5 to 65 are, therefore, taken as the most reliable for the age range they cover. We, therefore, conclude from the application of this method that males were underenumerated in 1970 relatively to 1960 by some 4.5 per cent; it is reasonable to assume that the underenumeration was roughly constant at all ages except for the economically active age group 20 to 35, for which it was almost certainly higher. The registration of deaths between the ages of about 10 and about 70 is estimated as 80.1 per cent complete on average during the decade; underregistration of deaths under the age of 10, as is known from other sources, was undoubtedly higher, and underregistration of deaths over about 70 may have been lower. In fact, assuming that underenumeration in 1970 for those aged 75 and over was 4.5 per cent relative to the enumeration of those over 65 in 1960, that is, the general average, the completeness of registration of deaths over about age 70 may be estimated directly from Equation (8) as 89.3 per cent, giving a compound value of 82.0 per cent for deaths of around 10 and over.

These estimates may be compared with those from other sources. For registration of deaths of adult males, the findings of the Survey of Population Change 1964 to 1965, shown in Table 4, estimated completeness at ages 10 to 60 at 71 per cent, with a somewhat higher level of completeness, 75 per cent, for those over age 60. The method presented in the first part of this paper gave an estimate of completeness of registration over age 10, based on an annual growth rate of 3.1 per cent, of 79 per cent. As far as underenumeration in 1970 is concerned, Fulton[12] has estimated an omission rate of 5.3 per cent for both sexes relative to 1960, and Arnold and Wanglee arrived at a figure for males of 5.5 per cent.[13] Both figures are somewhat higher than that obtained here, though the difference is not great. Not only do the estimates derived from the new method agree reasonably well with those available from other sources or obtained by other methods, but, even more encouragingly, they agree well with comparable estimates derived in the first half of this paper through the application of stable population theory. If the 1970 male population is adjusted for the relative omission of 4.5 per cent, the population growth rate between 1960 and 1970 becomes 3.12 per cent, very close to the rate of natural increase obtained from the 1964 to 65 Survey of Population Change.

One of this method's limitations should be emphasized in passing; the estimates of completeness obtained are never absolute, they are always relative to the completeness of enumeration of one or other of the two censuses. For the estimation of mortality, this is not important, since both deaths and the second census (or the first) can be rendered consistent with the completeness of enumeration of the first census (or the second), and both numerators and denominators of conventional mortality rates will be liable to equal proportionate errors. Consistent census populations may also be useful for the analysis of other statistics, such as intercensal registered births, even though numerator and denominator errors may not exactly cancel in such cases. It may also be mentioned that, analogous to the idea that Brass's Growth Balance Equation may provide a better estimate of the rate of population growth than of the completeness of death registration

[12] Economic and Social Commission for Asia and the Pacific. *Population of Thailand* ST/ESCAP/18. Bangkok, 1976.

[13] F. Arnold and A. Wanglee. Demographic Evaluation of the 1970 Census of Population and Housing in Thailand. Paper presented at the Fourth Population Census Conference. East-West Population Center Hawaii 1975.

because of the bunching of points close to the origin, the present approach may for the same reason provide a better estimate of relative enumeration completeness than of the completeness of death registration.

A more serious problem with the technique is its vulnerability to systematic overstatement of age at the higher ages. The technique in effect draws its inferences about registration completeness from developments at the older ages, where deaths are concentrated, and about relative census completeness from developments at younger ages, where population is concentrated. Systematic overstatement of age that worsens with age will affect the estimated death registration completeness in two ways:

(1) Because ages are overstated in censuses, too many people will be found at the second census at ages (x + 10) and above relative to the population aged x and above ten years earlier. This "in-migration" into cohorts will create the appearance that mortality conditions are better than they in fact are, so that recorded deaths will appear to represent a higher fraction of true deaths.

(2) As a result of overstatement of age at death, too many deaths will be recorded for cohorts starting the intercensal period at older ages.

Both these tendencies serve to inflate the estimated completeness of death registration (and to a much smaller extent to reduce the estimated completeness of the second census relative to the first). The only practicable way of minimizing the effects of age overstatement is to truncate the calculation of completeness at an age, across which the population is unlikely to be transferred by age overstatement. Should it be necessary to adopt a low upper limit for computation (say, age 50), then substantial information on registration completeness is sacrificed, although relatively little information on comparative enumeration completeness is lost.

SUMMARY

Two techniques are outlined for estimating the completeness of death registration. The first is based on assumptions of population stability and makes use of an estimate of the population growth rate. Except in unusual cases where there are reasons to believe child death registration to be no worse than that of adult deaths (e.g. where whole areas are excluded from the death registration area), it should be applied only to ages 5 or 10 and above. The second technique is based on intercensal cohort comparisons. It yields both an estimate of (primarily adult) death registration completeness and of differentials in census coverage. The two techniques are highly complementary when used to evaluate intercensal death registration, because the census completeness differential estimated in Technique II can be used to evaluate and modify the intercensal growth rate estimate that is used in Technique I. The two approaches are shown to provide reasonably consistent estimates of intercensal adult death registration (and growth rates) in Thailand, and are also validated by an independent source, the 1964 to 65 Survey of Population Change.