technical note . . .

A PROCEDURE FOR COMPARING MORTALITY MEASURES CALCULATED FROM INTERCENSAL SURVIVAL WITH THE CORRESPONDING ESTIMATES FROM REGISTERED DEATHS

by William Brass

The old and familiar procedure of mortality estimation from two successive censuses is based on two assumptions: that there is no migration, and that the census age distributions are accurate. The migration assumption gives trouble sometimes, but the real Achilles heel is the assumption that the age distributions are accurate, for it turns out that a relatively small error in these can throw the results very far off. This note presents a new technique which introduces intercensal death registration data into the census survival calculation.

The new procedure

The basic equation of the sectional growth balance procedure can be written as

$$\left(\frac{N_y}{P_y}\right) = r_y + \left(\frac{f_y D_y^*}{P_y}\right) \qquad (1a)$$

where N_y and P_y are the numbers living around the age point y and above y, respectively; r_y is the growth rate of the section of the population aged over y; D_y^* the recorded deaths in the current period at ages over y; and $f_y D_y^*$ the true deaths. This may also be written

$$b_y = r_y + f_y d_y^* \qquad (1b)$$

where $b_y \equiv N_y/P_y$ and $d_y^* \equiv D_y^*/P_y$, in which form the equation is seen to be the demographic equation in rate form for the section of the population aged over y. These equations are exact since r_y and f_y are allowed to vary with y. In application it is normally necessary to introduce the further assumption that r_y and f_y are constant with respect to y. Values of the growth rate r and the factor f representing underregistration of deaths may then be estimated by plotting b_y against d_y^* and fitting a straight line. A thorough discussion of this method appears in Rachad (1978).

Suppose now that intercensal survival calculations yield an independent source of data on deaths and write

$$b_y = r_y + d_y^\dagger + \delta_y \qquad (2)$$

where $d_y^\dagger$ is the death rate at over y from these data and δ_y its error. Again, this is perfectly general as long as no assumptions are made about the form of δ_y.

Obviously a third equation can be derived from the first two, namely

$$d_y^\dagger = -\delta_y + f_y d_y^* \qquad (3)$$

This could have been written down directly, since $d_y^\dagger + \delta_y$ and $f_y d_y^*$ are both expressions for the true death rate of the population over age y, but we wished to bring out the connection with the other equations.

In the abstract this formulation is trivial. It gains its significance from the nature of the errors in the estimation of deaths by different procedures. The errors for registered deaths tend to be proportional to the number of deaths in any age section. When deaths are found from intercensal survival, however, the errors tend to occur because of varying completeness of census enumeration or because of migration and are thus proportional to the numbers at risk in each age section. Exact proportionality would give a constant absolute error in the calculated death rate.

If the assumption of a proportional error in registered deaths and an absolute error in intercensal death rates holds, then the relation of $d_y^\dagger$ to d_y^* is a straight line from which f, δ, and the true death rate can be determined. From δ the relative completeness of coverage for the censuses can be calculated. Note that this does not require any specification of stability of the population.

Consider again the relation expressed by the equations

$$b_y = r_y + f_y d_y^* \qquad (1b)$$

and

$$d_y^\dagger = -\delta_y + f_y d_y^*. \qquad (3)$$

The plotting on the same graph of b_y and $d_y^\dagger$ against d_y^* provides a simple and useful summary of the relations among registered deaths, intercensal changes in numbers alive, and the age distribution of the living. In ideal circumstances the points will define two parallel straight lines with a slope of one. The line for the second equation will go through the origin and for the first through $(0,r)$. Deviations from this ideal pattern will provide evidence about the completeness of registration and census coverage by age. In favorable circumstances, the comparisons will suggest the appropriate hypotheses upon which corrections can be based.

Exhibit 1 Notation and formulas

Notation

N_y	-	number of persons living around the age point y
P_y	-	number of persons living above age y
r_y	-	growth rate of the section of the population aged over y
D_y^*	-	recorded deaths in the current period of ages over y
$f_y D_y^*$	-	true deaths in the current period of ages over y
$D_y^\dagger$	-	intercensal deaths in the current period at ages over y
b_y	-	"birth" rate of the section of the population aged over y - N_y/P_y
d_y^*	-	recorded death rate of the section of the population aged over y - D_y^*/P_y
$d_y^\dagger$	-	intercensal death rate of the section of the population aged over y - $D_y^\dagger/P_y$
δ	-	absolute error of $d_y^\dagger$

Basic equations

Growth balance: $b_y = r_y + f_y d_y^*$

Intercensal survival: $b_y = r_y + d_y^\dagger + \delta_y$

Derived equation: $d_y^\dagger = -\delta_y + f_y d_y^*$

Application to Thailand

The application to 1960–70 data from Thailand is very illuminating, for a good deal is known about the case, in particular that migration is probably negligible, and that the 1970 census was probably a less complete enumeration than the 1960 census.

The relevant data are shown in Table 1. The census dates were 25 April 1960 and 1 April 1970. The registered deaths are all for 1961 to 1969 plus 0.685 of those recorded for 1960 and 0.25 of those for 1970. The intercensal interval is 9.935 years. In the application, no adjustment is made to numbers by age group to allow for the slight deviation of the census interval from ten years.

The first stage of the calculations is given in Table 2. The numbers at the censuses in sections of the population over ages 0, 5, 10, ... are cumulated. The averages for the two censuses provide the populations at risk in the intercensal period, $P_0, P_5, \ldots$, and the differences with a ten-year age displacement the decrease in numbers for corresponding sections. Thus the number over 20 years of age in 1960 minus the number over 30 years in 1970 gives the deaths for this section in the census interval, provided recording is accurate and there is no migration. The lower age of the section changes over the interval from 20 to 30, but it is sufficient to regard the decrease as an estimate of deaths at above the mid-age of 25 years. More elaborate procedures are, of course, possible, but they do not seem to be justified here; the extra precision is small in relation to other uncertainties. In these calculations, the very small numbers of females recorded as "age not stated" are omitted. The estimated deaths at five years and over are less than those for ten years and over, contrary to logic, but the sequence falls

Table 1 Census age distributions and registered intercensal deaths: Thailand, 1960–70, females

Age group	Census age distributions (in thousands) 1960	1970	Intercensal registered deaths
0–4	2,101.9	2,796.2	308,379
5–9	1,979.8	2,605.7	52,640
10–14	1,525.4	2,252.6	25,825
15–19	1,236.3	1,885.4	25,418
20–24	1,204.2	1,361.7	30,752
25–29	1,046.5	1,143.4	31,165
30–34	869.9	1,077.1	34,580
35–39	679.9	957.6	36,291
40–44	563.8	766.3	35,422
45–49	483.0	597.5	33,535
50–54	410.4	489.8	36,734
55–59	329.0	401.7	38,682
60–64	245.0	324.2	45,235
65–69	163.6	238.9	44,527
70–74	244.7[a]	167.7	49,062
75–79	u	98.2	120,587[b]
80 and over	u	87.6	u
Not stated	20.4	21.8	76,404
Total	13,103.7	17,273.5	1,025,238

u—unavailable.

a All persons 70 and over were combined into a single age group in the published 1960 census data.

b All persons 75 and over were combined into a single age group in published registration data.

SOURCES: For 1960 age distribution, *Thailand Population Census: 1960 Whole Kingdom* (Bangkok: Central Statistical Office, 1962), Table 3, page 9. For 1970 age distribution, *1970 Population and Housing Census: Whole Kingdom* (Bangkok: National Statistical Office, no date), Table 4, page 12, to age 70; Table 3, pages 10–11 for 70 and over. Registered deaths calculated from *Statistical Yearbook: Thailand*, No. 30, 1972–73 (Bangkok: National Statistical Office, no date), Table 33, page 89, as sum of all registered deaths for the years 1961 through 1969 plus 0.685 of registered deaths for 1960 plus 0.250 of registered deaths for 1970.

Table 2 Cumulative population and deaths: Thailand, 1960–70, females

Age x	Population over age x (in thousands) 1960 (1)	1970 (2)	Average population over age x (in thousands) (3)	Cohort difference (in thousands) (4)	Registered deaths over age x (5)
0	13,083	17,252	15,168	na	948,834
5	10,981	14,455	12,718	1,234	640,455
10	9,002	11,850	10,426	1,384	587,815
15	7,476	9,597	8,537	1,290	561,990
20	6,240	7,712	6,976	1,126	536,572
25	5,036	6,350	5,693	1,033	505,820
30	3,989	5,207	4,598	906	474,655
35	3,119	4,130	3,624	817	440,075
40	2,440	3,172	2,806	714	403,784
45	1,876	2,406	2,141	631	368,362
50	1,393	1,808	1,600	557	334,827
55	982	1,318	1,150	476	298,093
60	653	917	785	390	259,411
65	408	592	500	300	214,176
70	245	354	299	222	169,649
75	129	186	157	157	120,587
80	u	88	u	u	u

u—unavailable.

NOTES: Columns (1) and (2) cumulated from Table 1. Number aged 70–74 in 1960 estimated by multiplying number 70 and over in 1960 by proportion of persons 70 and over in 1970 who were 70–74. Cohort difference by subtraction of column (2) values from column (1) values, e.g., 1,234 = 13,083 – 11,850. Last-digit discrepancies are due to rounding. Column (5) cumulated from Table 1.

steadily thereafter. The anomaly at five years is probably due to the underenumeration of young children at the 1960 census. No adjustment has been made for this.

The calculations of the sectional measures are completed in Table 3. If the age points y are taken as 5, 10, 15, ..., the P_y and D_y numbers would come directly from Table 2, and N_y would be formed as one-tenth of the numbers in the two five-year age groups bordering y (above and below). Hoda Rachad (1978) has provided evidence that slightly better results are obtained by taking y at the mid-points of the five-year age groups. N_y is then one-fifth of the number in the corresponding age group, and the P_y and D_y measures are formed by linear interpolation between the values at the end age points. The slight improvement comes because linear interpolation in the cumulated measures is a better approximation than taking N_y equal to one-tenth of the ten-year age group bracketing it.

In Table 3, $10N_y$ is the sum of the numbers in the five-year age groups at the 1960 and 1970 censuses. P_y is found from the average 1960–70 column of Table 2 by linear interpolation. $10D_y^*$ and $10D_y^\dagger$ are derived in the same way from columns (4) and (5), respectively, of Table 2. Note that the deaths of persons whose age was not recorded are omitted from the calculation. They could have been allocated in some way to ages, but it seems preferable to leave them aside at this stage of the comparison procedure. They must, of course, be brought into the final assessment. The multiplier 10 appears with N_y, D_y^*, and $D_y^\dagger$ in the table because the calculations are for a ten-year range and the symbols are by definition on an annual basis. The rates b_y, d_y^*, and $d_y^\dagger$ are finally found by dividing by P_y, the estimated population at risk.

Figure 1 gives a plot of b_y and $d_y^\dagger$ against d_y^*. In order to bring out the joint characteristics more clearly, the scale for b_y has been displaced by 27/1,000 relative to that for $d_y^\dagger$. This adjusts approximately for the difference between

Table 3 Sectional birth and death rate measures: Thailand, 1960–70, females

Age (y)	Numbers				Rates per thousand		
	$10N_y$	P_y	$10D_y^*$	$10D_y^\dagger$	b_y	d_y^*	$d_y^\dagger$
7.5	4,586	11,572	614	1,309	39.63	5.31	11.31
12.5	3,778	9,482	575	1,337	39.84	6.06	14.10
17.5	3,122	7,757	549	1,208	40.25	7.08	15.57
22.5	2,566	6,335	521	1,080	40.51	8.22	17.05
27.5	2,190	5,146	490	970	42.56	9.52	18.85
32.5	1,947	4,111	457	862	47.36	11.12	20.97
37.5	1,638	3,215	422	766	50.95	13.13	23.83
42.5	1,330	2,474	386	673	53.76	15.60	27.20
47.5	1,080	1,871	352	594	57.72	18.81	31.75
52.5	900	1,375	316	517	65.45	22.98	37.60
57.5	731	968	279	433	75.52	28.82	44.73
62.5	569	643	237	345	88.49	36.86	53.65
67.5	402	400	192	261	100.68	48.00	65.25
72.5	284	228	145	190	124.56	63.60	83.33

NOTES: See Exhibit 1 for notation. $10N_y$ calculated as sum of number in age group in 1960 and 1970 censuses. P_y calculated by linear interpolation between rows in column (3) of Table 2. $10D_y^*$ by linear interpolation between rows in column (5) of Table 2. $10D_y^\dagger$ by linear interpolation between rows in column (4) of Table 2.

$d_y^\dagger$ and δ_y at the lower age points. If the hypothesis of constancy in these measures held, the b_y and $d_y^\dagger$ points would be brought into close coincidence. Up to age 55 years or so this holds reasonably well, but there is a divergence at later ages. It seems too large to be explainable by variation in r_y or δ_y alone. Probably both contribute. Errors in b_y due to age misstatement effects on N_y may also be a factor.

The ideal line of slope one through the origin is clearly a grossly unsatisfactory description of the relation between $d_y^\dagger$ and d_y^*. A greater slope and a positive intercept on the $d_y^\dagger$ axis are required. A line has been fitted to the ten points from 12.5 through 57.5 years, which have a very regular trend. The value at 7.5 years is clearly affected by the underreporting of young children at the 1960 census, and the measures at 62.5 years and above diverge in a way inconsistent with the points on the graph of b_y against d_y^*. The resulting fitting equation is

$$d_y^\dagger = 5.72 + 1.38d_y^*$$

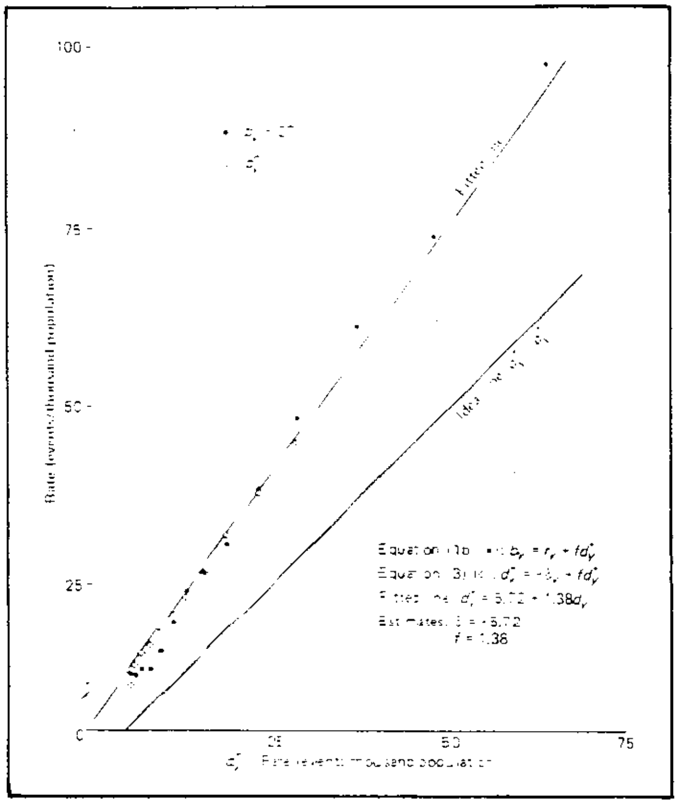
whence $\delta = -5.72$ and $f = 1.38$. This suggests that registered deaths have to be raised by about 40 percent to give the true number. If deaths at unknown ages are allocated proportionally, 8 percent of this deficit is accounted for. If they are all allocated to ages above 15 years, 14 percent of the deficit is explained. The 5.72 per thousand value of δ implies an error of 5.7 percent over the ten years in the relation between the coverage of the two censuses compared with closed populations. The percentage applies to the average coverage of the population at the two censuses, but for practical purposes the 1970 census can be said to have a coverage poorer than the 1960 census by about 5.5 percent. □

REFERENCE

Rachad, Hoda Mohamed. 1978. The estimation of adult mortality from defective registration data. Ph.D. dissertation, London School of Hygiene and Tropical Medicine, University of London.

William Brass is Professor of Medical Demography at the University of London and Director of the Centre for Overseas Population Studies at the London School of Hygiene and Tropical Medicine.

Figure 1 Plot of $b_y - 27$ and $d_y^\dagger$ against d_y^*: Thailand, 1960–70, females



SOURCE: Table 3.

DUALABS PLANS CENSUS REPORTS ON WOMEN

The status and roles of women in developing countries are the subject of an international study sponsored by Data Use and Access Laboratories (DUALabs) and the U.S. Agency for International Development (USAID). The goal of the project is to improve the usefulness of data on women from the 1980 round of censuses. An international working group was recently formed to prepare census reports on the status of women. Cooperating institutions, in addition to DUALabs and USAID, include national statistical offices from a number of countries and Data for Development International Association.

Meeting in France in June, delegates from Bangladesh, Costa Rica, Indonesia, Kenya, Mauritania, Panama, Peru, and the Philippines discussed issues affecting women for which data are needed, along with analytic approaches for presenting such data. DUALabs will provide assistance in the form of guidelines on data tabulation and analysis, literature resources, consultation on census data processing, installation and use of software, and guidance in the selection of computer hardware. To be considered for assistance, a national statistical office must commit itself to the preparation of an analytic report on women. It must also identify and support the persons assigned to write the report.

DUALabs is a nonprofit organization whose purpose is to help people gain access to and use public data. Data for Development is a private voluntary association concerned with the use of data in the context of development. To learn more about the project on census data on women, write to Mary Ellen Barrorey, Women in Development Project, Data Use and Access Laboratories, Inc., 1601 North Kent Street, Suite 900, Arlington, Virginia 22209, U.S.A.