DEMOGRAPHY© Volume 20, Number 2 May 1983

AN INTEGRATED SYSTEM FOR DEMOGRAPHIC ESTIMATION FROM TWO AGE DISTRIBUTIONS

Samuel H. Preston
Population Studies Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Abstract—This paper presents a simple method for estimating a birth rate and a level of mortality for an intercensal period. The birth rate is estimated from the intercept of a line fitted to data and the level of mortality from the slope of that line. The formula that is developed is based upon a recent generalization of stable population relations. An estimate of childhood mortality level is an optional but significant piece of additional input. An important by-product of the procedure is an estimate of the true age distribution.

This paper presents a simple method for estimating birth rates and, simultaneously, a parameter representing the level of mortality. It is designed for use in countries lacking good vital registration, which comprise a majority of the world's population (cf. United Nations, 1980). The birth rate is inferred from the intercept of a straight line; the level of mortality is inferred from the slope of that line. The estimation procedure integrates Brass's one-parameter logit mortality system with recently developed equations generalizing stable population theory. Two successive age distributions are required to use the system.

A one-paragraph introduction to this method was presented in Preston and Coale (1982, p. 244). Here we derive the basic formula, extend it to the situation where an independent estimate of child mortality is available, apply the procedure to data, and discuss its sensitivity to various forms of error.

BACKGROUND

Generalized Stable Equations

Preston and Coale (1982) have shown that the following relation pertains at a moment in time in any closed population:

$$c(a, t) = b(t)e^{-\int_0^a r(x,t)dx}p(a, t), \quad (1)$$

where

$c(a, t)$ = proportion of the population that is aged a at time t
$b(t)$ = crude birth rate at time t
$r(x, t)$ = annual growth rate of persons aged x at time t
$p(a, t)$ = probability of survival to age a according to the period life table prevailing at time t.

Closely related versions of this equation also pertain within discrete time intervals. The most direct way to develop an analogous discrete time equation is to assume the age-specific growth rate function, $r(x, t)$, to be constant over time within that interval. In this case, $c(a, t)$ in the above formula is the ratio of ath birthdays during the period to the person-years that were lived in the population, $b(t)$ is births divided by total person-years lived in the time interval, and $p(a, t)$ is a birth-weighted average of the $p(a, t)$ functions that prevailed during the period. If $r(x, t)$ is constant over age, equation (1) reduces to a familiar equation characterizing a stable population.

This equation also applies to an open population as long as $r(x, t)$ is expressed as the sum of the age-specific growth rate at age x, time t, and the rate of net emigration (per person-year) at x, time t.

Hereafter, we will drop the t identifier for convenience.

Rewriting equation (1), we have:

$$\frac{1}{p(a)} = \frac{be^{-\int_0^a r(x)dx}}{c(a)}. \qquad (2)$$

Brass's Logit Transformation

Brass (1971) shows that a linear logit transformation of the survival function $p(a)$, or of its complement, $q(a)$, often provides a good representation of changes in age-specific mortality conditions as levels of mortality change. In particular, he assumed that, of each set of $q(a)$ functions within a model life table system,

$$\ln \frac{q(a)}{p(a)} = \propto + B \ln \frac{q_s(a)}{p_s(a)}, \qquad (3)$$

where $q_s(a)$, $p_s(a)$ are the $q(a)$ and $p(a)$ functions in the "standard" schedule adopted; and $\propto$, B are parameters that relate mortality in any other member of that life table system to that in the standard. $\propto$ is the "level" parameter; when its value is changed, it changes the level of mortality in the same direction for all ages. B is the slope parameter; when it rises, it increases $\ln [q(a)/p(a)]$ for all ages above the age where $q_s(a)/p_s(a)$ is unity, and lowers $q_s(a)/p_s(a)$ at all lower ages. Tabulated values of a logit life table system may be found in Carrier and Hobcraft (1971). Some evidence on the success of the logit transformation is presented below.

Equation (3) can be rewritten as:

$$\frac{1 - p(a)}{p(a)} = e^{\propto} \left[\frac{q_s(a)}{p_s(a)} \right]^B, \text{ or}$$

$$\frac{1}{p(a)} = e^{\propto} \left[\frac{q_s(a)}{p_s(a)} \right]^B + 1. \qquad (4)$$

Deriving the Estimation Equations

Equations (2) and (4) can be combined to give

$$\frac{be^{-\int_0^a r(x)dx}}{c(a)} = e^{\propto} \left[\frac{q_s(a)}{p_s(a)} \right]^B + 1, \text{ or}$$

$$\frac{e^{-\int_0^a r(x)dx}}{c(a)} = \frac{1}{b} + \frac{K}{b} \left[\frac{q_s(a)}{p_s(a)} \right]^B. \qquad (5)$$

We have set $e^{\propto} = K$ for simplicity in exposition. Note that equation (5) requires nonlinear estimation procedures unless $B = 1$. But setting $B = 1$ is a reasonable procedural assumption in any event, since it is rare to have census information of sufficient quality to permit accurate estimation of both $\propto$ and B. B is the more vulnerable of the two parameters to systematic age misstatement (as illustrated below) and in any event the major interest is usually attached to knowing the "level" of mortality. Thus, assuming $B = 1$, we have

$$\frac{e^{-\int_0^a r(x)dx}}{c(a)} = \frac{1}{b} + \frac{K}{b} \frac{q_s(a)}{p_s(a)}. \qquad (6)$$

Equation (6) is a simple linear equation. The intercept is the reciprocal of the birth rate and the slope is K (the "level" of mortality) divided by the birth rate. Elements required for estimating values on the left-hand side are readily derived from two censuses. The variable on the right-hand side, $q_s(a)/p_s(a)$, is of course supplied by an assumed life table, which will eventually be modified by the value of K to be determined.

It is very common that a country will have reasonably good information on levels of child mortality by virtue of Brass questions on numbers of children ever born and surviving. This information can be conveniently integrated into the estimation procedure and should normally be used where available. The most useful index of child mortality for present purposes is $p(5)$, the probability that a child will survive to age 5. If the information on children born and surviving is drawn from the second of two censuses separated by 10 years, $p(5)$ will refer to a point approximately midway through the intercensal period (National

Academy of Sciences, 1981). A midcensal estimate should be the goal in any event.

Define ${}_5p_s(a)$ and ${}_5q_s(a)$ as the probabilities of surviving to and dying before age a, respectively, for someone who has survived to age 5 in the standard life table. Thus ${}_5q_s(5) = 0$ and ${}_5p_s(5) = 1.00$. Assume that the one-parameter logit transformation pertains only to mortality above 5:

$$\frac{{}_5q(a)}{{}_5p(a)} = K \frac{{}_5q_s(a)}{{}_5p_s(a)} \qquad \text{for } a \geq 5 .$$

Table 1 demonstrates that a one-parameter logit transformation of post-5 mortality within the West model life table system represents mortality variation quite effectively within a 30-year range of life expectancies. By transforming the ${}_5p(a)$ function in a life table with $e_0 = 50$, the ${}_5p(a)$ functions corresponding to $e_0 = 35$ and to $e_0 = 65$ can be reproduced quite accurately. A 30-year span surely exceeds the range of uncertainty in the great majority of applications. A graphical demonstration is more convincing still, but the actual and predicted series are so close that they can hardly be separately graphed.

Let us designate the observed $p(5)$ as $p^*(5)$. Since

$$p(a) = p^*(5) \cdot {}_5p(a) \text{ for } a \geq 5 ,$$

we may return to equation (2) and rewrite equation (6) after introducing the new parameterization of mortality:

$$\frac{p^*(5)e^{-\int_0^a r(x)dx}}{c(a)} = \frac{1}{b} + \frac{K}{b}\left[\frac{{}_5q_s(a)}{{}_5p_s(a)}\right], \; a \geq 5 . \qquad (7)$$

Table 1.—Probabilities of Surviving from Age 5 to Various Ages in West Female Model Life Tables at Life Expectancies at Birth of 35 and 65

	Life Expectancy at Birth of 35 ${}_5p(a)$		Life Expectancy at Birth of 65 ${}_5p(a)$	
Age (a)	Actual[a]	Predicted[b]	Actual[a]	Predicted[b]
5	1.000	1.000	1.000	1.000
10	.959	.956	.993	.991
15	.927	.923	.988	.984
20	.888	.882	.979	.974
25	.841	.833	.969	.961
30	.791	.782	.956	.947
35	.737	.729	.942	.930
40	.683	.676	.924	.912
45	.628	.623	.903	.891
50	.573	.568	.876	.867
55	.508	.504	.838	.835
60	.435	.433	.787	.791
65	.345	.349	.713	.726
70	.250	.260	.609	.635
75	.154	.170	.469	.503
80	.075	.092	.305	.334

[a] Source: Coale and Demeny (1966), pp. 8 and 18.
[b] Derived through one-parameter transformation of ${}_5p(a)$ in Coale-Demeny "West" female model life table with $e_o = 50$. To derive the prediction for $e_o = 35$, ${}_5q(a) / {}_5p(a)$ in the model life table with $e_o = 50$ was multiplied by $K = 2.12$ and the results transformed into ${}_5p(a)$. For the prediction at $e_o = 65$, the same procedure was used but with a $K = 0.43$.

Equation (7) remains a simple linear equation whose parameters can readily be estimated. But it now incorporates the important additional information on child mortality levels.

IMPLEMENTING THE SYSTEM

We will assume that an estimate of $p^*(5)$ is available. Data on population age distribution from two censuses are assumed to be available in 5-year wide age groups. The procedure can be applied to data in discrete 5-year age intervals or to specific ages. The latter choice is substantially the simpler and does not appear to sacrifice accuracy. The steps are the following.

(1) Estimate ${}_5r_x$, the age-specific growth rate for the population aged x to $x + 5$, as

$$ {}_5r_x = \frac{\ln {}_5N_x\,(t + h) - \ln {}_5N_x\,(t)}{h} $$

where ${}_5N_x\,(t)$ is the number of persons reported at ages x to $x + 5$ at time t (the first census) and h is the length of the intercensal period in exact number of years. The set of ${}_5r_x$'s should be computed for $x = 0, 5, 10, \ldots$ up to an arbitrary terminal age. Age-specific net emigration rates, if available, should be added to the ${}_5r_x$ series (Preston and Coale, 1982).

(2) Estimate $c(a)$, the ratio of ath birthdays to total person years lived during the period. The most satisfactory procedure for doing so, in view of the assumptions underlying the formulas, is to estimate person-years lived between ages a and $a + 5$ during period t to $t + h$ as

$$ {}_5\hat{N}_a\,(t \text{ to } t + h) = \frac{{}_5N_a\,(t + h) - {}_5N_a(t)}{{}_5r_a \cdot h} $$

for each age and derive $c(a)$ as

$$ c(a) = \frac{{}_5\hat{N}_a\,(t \text{ to } t + h) + {}_5\hat{N}_{a-5}\,(t \text{ to } t + h)}{10 \sum_{x=0}^{\infty} {}_5\hat{N}_x\,(t \text{ to } t + h)} \quad (8) $$

(3) The values of the left-hand side of (7) can now be calculated for all ages $a = 5, 10, 15, \ldots$.

For example,

Age	$\dfrac{p^*(5)e^{-\int_0^a r(x)dx}}{c(a)}$
5	$\dfrac{p^*(5)e^{-5\,{}_5r_0}}{c(5)}$
10	$\dfrac{p^*(5)e^{-5({}_5r_0+{}_5r_5)}}{c(10)}$
15	$\dfrac{p^*(5)e^{-5({}_5r_0+{}_5r_5+{}_5r_{10})}}{c(15)}$
⋮	
etc.	

(4) It is now only necessary to adopt a standard mortality schedule for ages above 5. This should, of course, be as similar as possible to the presumed level and shape of adult mortality in the population under study. If no other information is available, a level might be chosen that corresponds in a model life table system to the particular value of $p^*(5)$. Note that no logit calculations are necessary; the only required values are

$$ \frac{{}_5q(a)}{{}_5p(a)} = \frac{l_5 - l_a}{l_a} $$

in the life table chosen as the standard.

(5) Plot the dependent variable

$$ y = \frac{p^*(5)e^{-\int_0^a r(x)dx}}{c(a)} $$

against the independent variable

$$ x = \frac{{}_5q_s(a)}{{}_5p_s(a)} . $$

The plot should be approximately linear. If it is not, an error in data, calculation, or assumption is implied. The *only* assumption employed, however, is that adult mortality (i.e., beyond age 5) can be represented as a one-parameter logit transformation of a standard. The user can, of course, experiment with different standards and different values of B. Other sources of error are discussed below.

(6) Fit a line to the relationship between the variables. There are several ways to fit a line to the points. The two major alternatives are least-squares procedures, which tend to give more weight to outlying observations, and grouped mean procedures, which will normally be preferred because of their reduced sensitivity to outliers. Regardless of the fitting procedure used, the intercept is the reciprocal of the birth rate. The slope divided by the intercept is K, the factor by which the ${}_5q_s(a)/{}_5p_s(a)$ schedule is to be multiplied in order to estimate adult mortality for the population in question.

APPLICATIONS

India

Table 2 shows the values required to apply the procedure to Indian females during the intercensal period from 1961 to 1971. The points (cols. 4 and 5, designated hereafter y and x) are plotted in Figure 1. The points fall very close to a straight line, except that for age 75. A line was fit to the points by a grouped mean procedure, using the mean values of x and y for ages 5–30 as one "observation" and the mean values of x and y for ages 45–70 as the other. The equation of this line is

$$y = 23.735 + 44.992x, \text{ implying that}$$

$$b = \frac{1}{23.735} = .0421 \text{ and}$$

$$K = \frac{44.992}{23.735} = 1.896.$$

When the estimated value of K is applied to ${}_5q_s(a)/{}_5q_s(a)$ and a complete life table recomputed from age 5, the estimated life expectancy at age 5 is 51.0 years.[1] This lies between the estimate of 53.2 years by Dyson (1979), also using the South model life table system on intercensal data, and that of the Registrar-General of India (1977, p. 16) of 50.2 years. As Dyson points out, there is reason to believe that the Registrar-General's estimate is too low. Preston and Bennett (1982, p. 15), using a method that does not require a model life table system, derive an e_5^0 of 53.6.

The estimated birth rate of 42.1/1000 is very close to the range of 40.5–42.0 given by Adlakha and Kirk (1974) for the intercensal period; it is likely that our estimate is higher because it uses a higher estimate of child mortality based upon newly-published results from a 1965–1966 National Sample Survey. Our estimate is substantially higher than estimates based upon the Indian Sample Registration System for years near the end of the decade.

South Korea

Table 3 and Figure 2 show comparable values used to apply the system in South Korea. A different "standard" table is chosen for mortality above age 5, one that is believed to be more reflective of Korean conditions: West female level 19 ($e_0 = 65$).

The equation of the line fitted by grouped mean procedures to points 5–30 and 45–70 is

$$y = 36.947 + 43.831, \text{ implying that}$$

$$b = \frac{1}{36.947} = .02707 \text{ and}$$

$$K = \frac{43.831}{36.947} = 1.186.$$

The implied levels of birth rates and life expectancy are in good agreement with those presented in Coale et al. (1980). The unweighted mean birth rate for 1967–1975 given in this source is 28.6.[2] It is derived mainly through own-children techniques. Some of the rather small discrepancy between the two figures results from the fact that the intercensal procedure described here implicitly as-

Table 2.—Application of Integrated Procedure to Indian Females, 1961–1971

Age at Start of Interval	Age-Specific Growth Rate, 1961–1971	Cumulation of Age-Specific Growth Rates Up to Age a	Annualized Proportion of Intercensal Person-Years Lived at Age a		$_5q_s(a)/_5p_s(a)$ in South Female Model Life Table Level 13 (e_o = 50)[c]
a or x	$_5r_x$	$\sum_{x=0}^{a-5} {_5r_x}$	$c(a)$ [a]	$\frac{p^*(5)}{\mathrm{Col}(3)} \cdot \exp\{-5 \cdot \mathrm{Col}(2)\}$[b]	
	(1)	(2)	(3)	(4)	(5)
0	.0180				
5	.0232	.0180	.03016	23.515	0.
10	.0340	.0412	.02651	23.823	.0227
15	.0255	.0752	.01982	26.883	.0363
20	.0119	.1007	.01684	27.851	.0561
25	.0127	.1126	.01666	26.527	.0819
30	.0186	.1253	.01498	27.685	.1113
35	.0280	.1439	.01263	29.921	.1433
40	.0207	.1719	.01079	30.448	.1795
45	.0226	.1926	.00897	33.029	.2210
50	.0167	.2152	.00759	34.859	.2706
55	.0271	.2319	.00585	41.605	.3414
60	.0221	.2590	.00480	44.280	.4478
65	.0347	.2811	.00380	50.081	.6410
70	.0232	.3158	.00227	70.483	1.0104
75	.0235	.3390	.00148	96.267	1.8471
80	.0192	.3625	.00080	158.349	4.1223
85	.0271				

[a] Calculated by formula 8b.
[b] Estimated $q(2)$ for the mid-censal period is .184 according to unpublished calculations by P. N. Mari Bhat for the U.S. National Academy of Sciences. This was converted into the corresponding $p(5)$ of .776 via the South model life table system.
[c] Source: Coale and Demeny (1966).

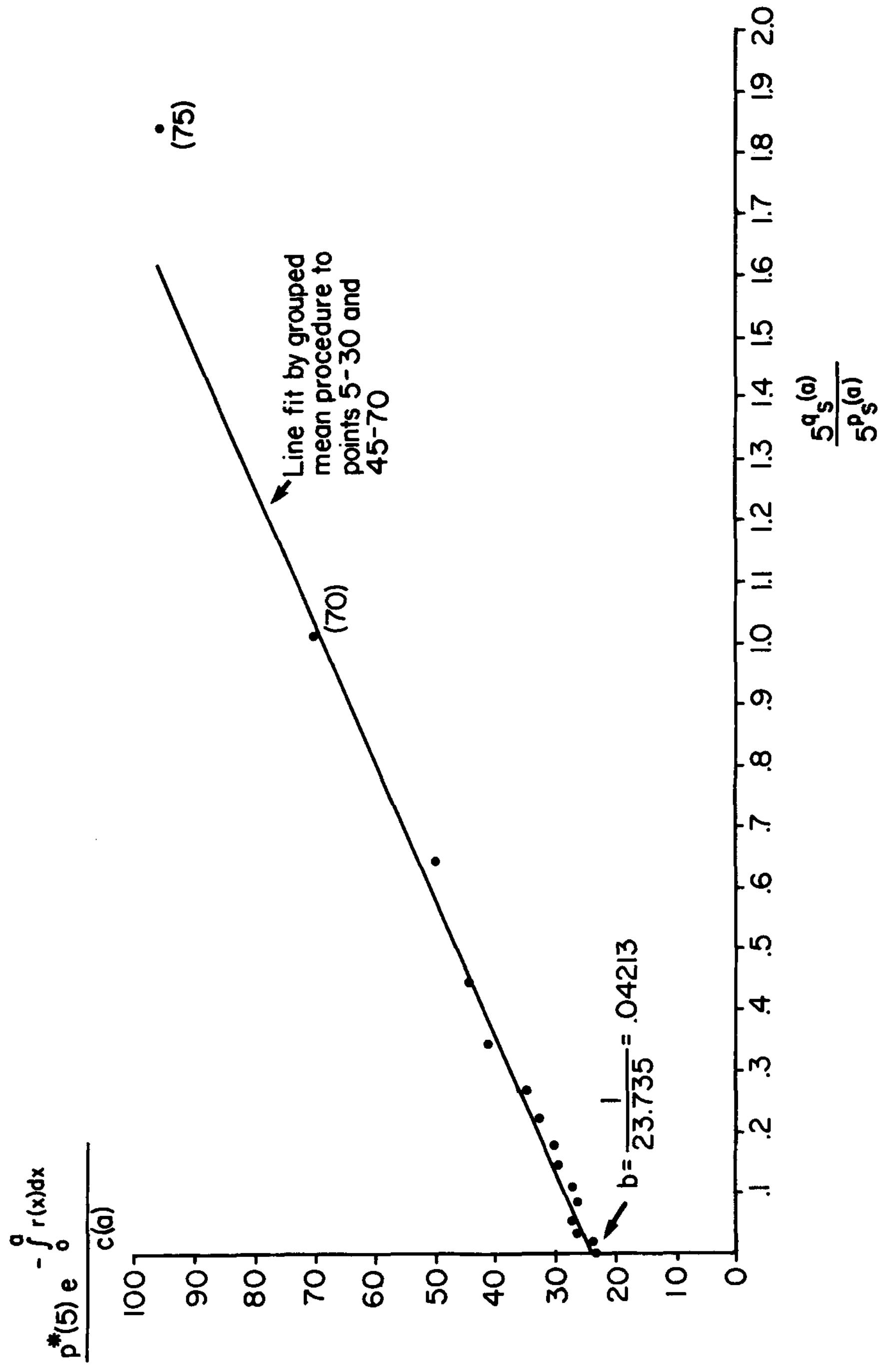

Figure 1.—Application of Integrated Procedure to Indian Females, 1961–1971

Table 3.—Application of Integrated Procedure to South Korean Females, 1966–1975

Age at Start of Interval	Age-Specific Growth Rate, 1966-1975	Cumulation of Age-Specific Growth Rates Up to Age a	Annualized Proportion of Intercensal Person-Years Lived at Age a		$_5q_8(a)/_5p_8(a)$ in West Female Model Life Table Level 19 (e_o = 65)[c]
a or x	$_5r_x$	$\sum_{x=o}^{a-5} {}_5r_x$	$c(a)$ [a]	$\frac{p^*(5)}{\text{Col}(3)} \cdot \exp\{-5 \cdot \text{Col}(2)\}$[b]	
	(1)	(2)	(3)	(4)	(5)
0	-.0066				
5	-.0036	-.0066	.02716	36.265	0.
10	.0255	-.0102	.02619	38.284	.0069
15	.0484	.0153	.02273	38.836	.0123
20	.0358	.0637	.01858	37.298	.0207
25	.0101	.0995	.01567	36.976	.0321
30	.0117	.1096	.01406	39.181	.0458
35	.0306	.1213	.01254	41.434	.0619
40	.0318	.1519	.01101	40.496	.0817
45	.0330	.1837	.00915	41.565	.1069
50	.0281	.2167	.00758	42.542	.1414
55	.0192	.2448	.00633	44.267	.1927
60	.0317	.2640	.00507	50.208	.2709
65	.0229	.2957	.00401	54.175	.4027
70	.0243	.3186	.00294	65.899	.6416
75	.0242	.3429	.00192	89.362	1.1302
80	.0555	.3671	.00110	138.202	2.2769
85	.0693				

[a]Calculated by formula 8b.
[b]$p^*(5)$ estimated at .95292 by weighting life tables for 1966-1970 and 1970-1975 presented in Coale et al. (1980, p. 35).
[c]Source: Coale and Demeny (1966).

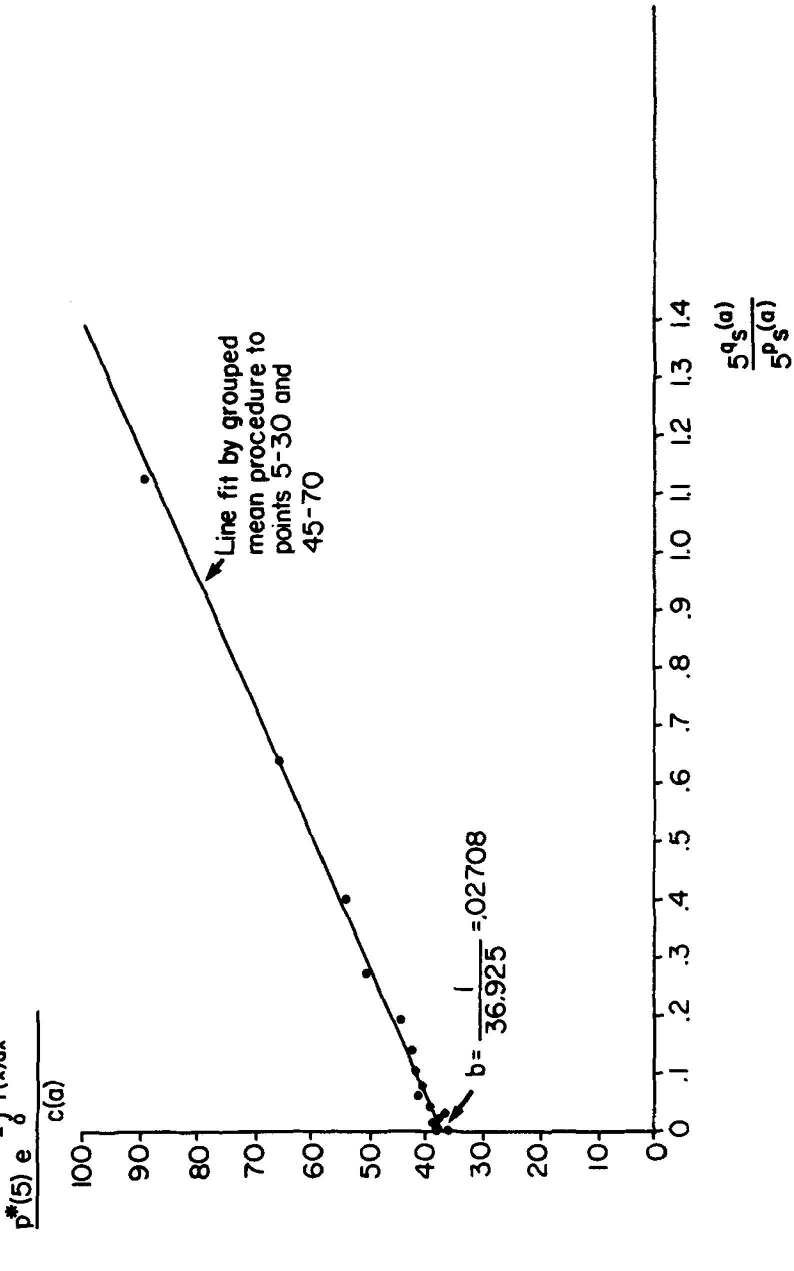

Figure 2.—Application of Integrated Procedure to South Korean Females, 1966–1975

signs greater weight to more recent observations in a growing population such as South Korea; and during the period, Korea's birth rate declined rapidly. A K value of 1.86 translates into an e_5 of 63.07, which is almost identical to the value presented in Coale et al. (1980, p. 35) based on West model life tables for 1966–1970 (63.16) but less than that for 1970–1975 (65.07). Using a nonparametric intercensal estimation procedure, Preston and Bennett (1982) derived an e_5 value of 62.7 for the intercensal period.

SENSITIVITY TO ERROR

There are several advantages to the estimation system proposed here relative to existing procedures. Perhaps the most attractive is that it appears to provide a fairly robust estimate of the birth rate. The apparatus for indirect estimation has been much more extensively developed for mortality than for fertility estimation. We discuss briefly below the sensitivity of results to various forms of error and to the procedures used.

Choice of Points to Fit the Line

Experimentation with differing combinations of points for fitting lines in the two cases shown above (but always including the lowest ages and relatively high ages) results in birth rate estimates that vary by less than 1/1000. The reason is that points for young ages are heavily clustered at low values (i.e., near the intercept that determines b). Consequently, choosing different points for fitting at the high ages still results in a line that goes through the middle of that cluster. Relative to backward projection of the population below age 10, in which birth rate estimates are based exclusively on number of persons below age 10 at the second census, the procedure described here uses much more information on the age distribution. In principle, it uses information on the proportions at all ages to determine b; in practice, points at higher ages (i.e., at higher values of x and y) have successively less influence on the intercept and successively more on the slope.

Changing Census Coverage Completeness

For similar reasons, birth rate estimates (but not estimates of adult mortality) are relatively insensitive to changing completeness of coverage in the two censuses. To illustrate, we have reduced all intercensal growth rates in India by 3/1000, equivalent to a 3 percent deterioration in coverage at all ages in the second census relative to the first. The equation of the new line fit to points 5–30 and 45–70 is

$$y = 24.303 + 62.200x, \text{ implying that}$$

$$b = \frac{1}{24.302} = .0411 \text{ and}$$

$$K = \frac{62.200}{24.302} = 2.559.$$

The birth rate estimate is changed by only 1/1000 as a result of this large simulated deterioration in census coverage. However, the level of adult mortality rises substantially, as more people are absent at the second census and presumed dead. The value of e_5 in this simulated case is 47.2 years, compared to the original estimate of 51.0.

Inaccurate Information on Migration

A failure to incorporate accurate information on international migration in the estimation equation could have several different effects, depending on the age-incidence of migration. If the error in the net emigration series (which is to be added to $r(x)$) is invariant to age, then the effect on estimates is formally equivalent to the effect of differential coverage completeness in the two censuses. As shown above, the effect on birth rate estimates is probably small, but on adult mortality estimates it can be substantial. If emigration is larger than that allowed for, adult mortality estimates will be spuriously inflated, since emigrants appear to be deaths. If emigration is higher than allowed for only at older ages (say, after 25 or 30), then the adult mortality estimate should be the only one affected (again, it will be too high). If, however, emigration is underestimated only at the youngest ages (say, below 15), then the birth rate estimate will be too low, since the entire y series will be spuriously raised by a relatively constant amount. In effect, part of the relatively old age distribution that results from youthful emigration appears instead to result from low fertility. A pattern of error in emigration that is concentrated in the ages 15–30, which is perhaps the most commonly encountered pattern, should have relatively little effect on the birth rate estimate, since the effect on y-values accumulates and becomes substantial only in the middle range of points plotted. But it will, again, raise estimates of adult mortality.

If estimated net emigration rates are too high (i.e., net immigration rates are too low), then results are the converse of those described above.

Age Misstatement and Differential Omission of Children

An important source of error in indirect estimation of developing countries is age misstatement and differential omission from censuses by age. The present method is also vulnerable to these problems. But once again they appear to affect adult mortality estimates more seriously than estimates of the birth rate. One of the most serious problems for most methods of birth rate estimation, especially backward projection but also procedures using the ogive of the age distribution, is differential omission of young children from censuses. To illustrate the effects of such omission on the present procedure, we have arbitrarily reduced the number of children aged 0–4 in both Indian censuses by 10 percent and reestimated the line (again using ages 5–30 and 45–70) as

$$y = 23.671 + 43.701x, \text{ implying that}$$

$$b = \frac{1}{23.671} = .0422 \text{ and}$$

$$K = \frac{43.701}{23.671} = 1.8462.$$

Very surprisingly, the estimated birth rate is changed only from 42.1/1000 to 42.2/1000. And instead of being lowered by the omission of the youngest children, it is raised slightly. What has happened to cause this slight rise is that, when children aged 0–4 are omitted, the proportion at all other ages rises, depressing their y-values. In the cluster of points near the intercept, one rises sharply in value and all others decline. The sum of the first six y-values actually declines, and as the intercept declines the birth rate rises.[3] But on average the cluster remains in much the same place. This insensitivity contrasts with the very considerable sensitivity to omission of children of back-projection methods. If 10 percent of children 0–4 are differentially omitted from the censuses, the back-projected value of b based on children 0–9 would be too low by about 5 percent.

Age misreporting that is essentially random will add scatter to the points without biasing estimates.[4] Transfers of population to adjacent age intervals should likewise have small effects. However, systematic overstatement of age at the higher ages may bias estimates, particularly of adult mortality. Such overstatement will raise y-values in the middle-to-late ages being evacuated, and lower them in the highest ages receiving the spurious inflow of population. This pattern should be detectable in a hill-shaped scatter of points about the line; indeed, it is observed in the India data of Figure 1. An estimation strategy should simply endeavor not to use the points so affected, particularly since the very high values of y and x pertaining to the high ages factor heavily into group averages. Note that the pattern of the scatter produced by age misstatement appears similar to that which would be produced by a value of B (the adult mortality slope parameter) of less than unity. For this reason, it may be misleading to use this estimation procedure to estimate B.

Error in Child Mortality

The sensitivity of b to error in estimated $p^*(5)$ can be stated very simply: the estimated b is inversely proportional to $p^*(5)$. If $p^*(5)$ is too high by 5 percent (a *very* large error relative to typical uncertainty in $p^*(5)$), all y-values will be raised by 5 percent. Both the slope and the intercept also will be raised by 5 percent, and the estimated birth rate will be 5 percent too low. However, the estimate of adult mortality will be unaffected, since K is the ratio of the slope to the intercept and both are distorted by the same factor. Although the estimated b is sensitive to $p^*(5)$, it is fortunate that $p(5)$ is one of the most reliably-observed demographic parameters in developing countries, thanks to the success of the Brass technique.

Error in Estimated Adult Mortality Pattern

The x-values used in plotting the basic relation shown in Figures 1 and 2 are simply ${}_5q_s(a)/{}_5p_s(a)$. Choosing as a standard the wrong level of mortality within a one-parameter logit life table system should have very little effect on either the birth rate estimate or on the estimate of adult mortality. If the one-parameter transform applies exactly, the x-values will all be incorrect by the factor ϵ, the error in assumed mortality level. But this error will be exactly reflected in the estimated K value that is used to correct the initial standard. The final estimated mortality level should be correct, and the intercept should be undisturbed by equiproportionate errors in x-values.

A more serious difficulty arises if the wrong adult mortality pattern is chosen. To illustrate the sensitivity of results to use of very different age-patterns of mortality, Larry Heligman of the United Nations Population Division has repeated the India exercise using the United Nations (1982) new South Asian model life table as a standard. This pattern is quite different from any contained in the Coale-Demeny life table set, with extremely low mortality in the age span 5–50 and unusually high mortality outside of this span. At the same level of $p(5)$ used for India above, the South Asian model has lower mortality than is contained in the South at all ages between 10 and 50, and higher mortality above age 50. The maximum difference between the two ${}_5q_s(a)/{}_5p_s(a)$ functions, .0515, thus occurs at age 50 (.2823–.2308). Using the South Asian pattern as a standard results in an estimated crude birth rate of 41.4 (vs. 42.1 for the South) and an estimated e_5 of 52.7 years (vs. 51.0 for the South). As anticipated, the proportionate difference is greater for e_5^0 than it is for the crude birth rate. The estimated crude birth rate is reduced by using the South Asian pattern because, at a given level of y-value, the x-value is reduced more at lower ages (except that corresponding to age 5) than at higher ages. This distortion has the effect of raising the intercept and lowering birth rate estimates. Experimenting with different B-values—the slope coefficient for adult mortality—also produces changes in the estimated birth rate. With a $B = .841$ in the South Asian pattern, the crude birth rate is 43.1 (compared to 41.4 when $B = 1.00$) and e_5^0 is 50.6, compared to 52.7. Here, the proportionate changes in birth rates and e_5^0 are nearly the same.

ESTIMATING THE TRUE AGE DISTRIBUTION

It is useful to note that, by accepting the recorded intercensal growth rates by age, we are implicitly adopting the "hypothesis of similar errors" previously used by Coale and by Demeny and Shorter (1968). That is, we are assuming that equal proportionate errors have occurred in a particular age group at both censuses, so that recorded growth rates are correct. It is possible to estimate what those errors are by reconstructing the "true" age distribution and comparing it to the recorded one. An estimate of the true age distribution is likely to be one of the most important by-products of this procedure, since there is no other method for estimating it short of making stable assumptions or introducing a dual record system. To estimate the true age distribution, ${}_5\hat{c}_a$, we may accept the estimated b, the recorded age-specific growth rates, and the estimated level of mortality:

$${}_5\hat{c}_a = b\ e^{-\int_0^{a+2.5} r(x)dx}\ \frac{{}_5L_a}{l_0}\ .$$

Such estimates of ${}_5\hat{c}_a$ are not forced to sum to unity, and in several applications they have differed from it by 1 or 2 percent. For final estimates of ${}_5\hat{c}_a$, all such preliminary estimates should be divided by the sum of preliminary estimates. Although such division seems to constitute a revision of b, it is more properly interpreted as a fudge-factor required to fit continuous relations into discrete boxes.

SUMMARY

The estimation system proposed here builds upon several others: Brass's adult mortality estimation via logit transformations of a standard schedule; Brass-type child mortality estimation; and recent generalizations of stable population procedures. It is new only in the degree to which it integrates these components into a simplified, one-step procedure for estimating fertility and mortality conditions. The computational advantage relative to separate application of individual procedures is likely to be greatest in situations where censuses are not separated by exactly 5 or 10 years. In such situations, conventional calculations of adult mortality conditions can be quite laborious, since survival rates are normally tabulated in 5 and 10 years increments, and the graduations or interpolations required can introduce error. The visual display of data points on which estimation is based ought to aid in the identification of error in data. The simplicity of procedures also facilitates analysis of the sensitivity of results to data errors of various kinds. Results of several sensitivity analyses suggest that the procedure provides relatively robust estimates of intercensal birth rates. It shares other intercensal methods' sensitivity of adult mortality estimates to differential census coverage completeness and to age misstatement. An important by-product of the method is an estimate of the true age distribution.

While the discussion and results presented here focus on the situation where an estimation of childhood mortality is available, the procedure can also be employed when it is not. In this case, the "standard" mortality schedule must begin at age zero instead of age 5 and equation (6) must be used. There are several reasons to use an estimate of childhood mortality if it is available. First, it is typically one of the most reliably estimated demographic variables for developing countries. Second, in the absence of child mortality information, the value of B adopted for estimation purposes will apply to relations between childhood and adult mortality, and not simply to relations among younger and older adults. Relations between child and adult mortality are highly variable (Coale and Demeny, 1966) and hence the procedure is liable to greater error (slightly mitigated by the possibility to experiment with values of B). In addition, data errors (such as differential completeness of censuses) that previously would have affected only levels of adult mortality will now affect levels of childhood mortality, to which birth rate estimates are sensitive.

NOTES

[1] Life expectancy for age 5 is derived by first recalculating ${}_5p(a)$ for a = 10, 15, . . . , 80 as ${}_5p(a) = 1/[1.896\, q_s(a)/p_s(a) + 1]$. Using the assump-

tion that deaths are evenly distributed within each age interval, person years lived between 5 and 80 are $5[p(10) + p(15) + \ldots p(75)] + 2.5\,[p(5) + p(80)]$. To this sum is added $p(80)\,\hat{e}_{80}$, where $\hat{e}_{80}$ is taken from a model life table with the same value of $p(80)/p(5)$ (in this case, South level 8.9). The resulting sum is life expectancy at age 5.

[2] Coale et al. (1980, p. 2). These years are chosen because both censuses were held in October.

[3] The decline results from the fact that, under the recommended procedure, ${}_5\hat{N}_0$ figures into only the calculation of $y(5)$, whereas ${}_5\hat{N}_x$ at all other ages figures into $y(a)$ at two ages, x and $x + 5$.

[4] In particular, least-squares estimates are unbiased by errors in the y-values as long as they are normally distributed with mean of zero and constant variance.

ACKNOWLEDGMENTS

The advice of Larry Heligman was instrumental in improving this paper. Ansley Coale, Griffith Feeney, P. N. Mari Bhat, Shanta Danaraj, Duanpon Theerawanviwat, and the referees also made useful comments. This research was supported by a grant from NICHD, HD 1P30 HD10379.

REFERENCES

Adlakha, Arjun, and Dudley Kirk. 1974. Vital Rates in India 1961–71 Estimated from 1971 Census Data. Population Studies 28:381–400.

Brass, William. 1971. On the Scale of Mortality. Pp. 69–110 in William Brass (ed.), Biological Aspects of Demography. London: Taylor and Francis.

Carrier, N. H., and John Hobcraft. 1971. Demographic Estimation for Developing Societies. London: Population Investigation Committee, London School of Economics.

Coale, A. J., and Paul Demeny. 1966. Regional Model Life Tables and Stable Populations. Princeton, N.J.: Princeton University Press.

Coale, A. J., Lee-Jay Cho, and Noreen Goldman. 1980. Estimation of Recent Trends in Fertility and Mortality in the Republic of Korea. Committee on Population and Demography Report No. 1. Washington, D.C.: National Academy of Sciences.

Demeny, P., and F. C. Shorter. 1968. Estimating Turkish Mortality, Fertility, and Age Structure. Istanbul: Istanbul University Press.

Dyson, Timothy. 1979. A Working Paper on Fertility and Mortality Estimates for the States of India. Paper presented to the Panel on India, National Academy of Sciences Committee on Population and Demography. Washington, D.C.

India, Registrar General. 1977. Census of India, 1971. Series 1. Paper 1 of 1977. Life Tables. New Delhi.

National Academy of Sciences. 1981. Manual of Indirect Estimation Techniques. Unpublished manuscript, Washington, D.C.

Preston, Samuel H., and Neil Bennett. 1982. A Census-Based Method for Estimating Adult Mortality. Population Studies Center Research Report No. 81-16. Ann Arbor: University of Michigan. (Population Studies, in press.)

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

United Nations. 1980. Population and Vital Statistics Report. April edition. ST/ESA/STAT/SER.A/132. New York: U.N. Department of International Economic and Social Affairs.

United Nations. 1982. Model Life Tables for Developing Countries. Population Study No. 77. New York: U.N. Department of International Economic and Social Affairs.