# Estimating Infant Mortality Trends from Child Survivorship Data †

GRIFFITH FEENEY*

In this paper a new procedure is developed for estimating infant mortality rates from figures derived from answers to population census questions on the total number of children a woman has borne during her lifetime and on the number of these children who are living at the time of the census. The statistics required are the number of women in a series of quinary age groups, beginning with ages 15–19, the total numbers of children born to women in these age groups, and the numbers of these children born who have survived to the time of the census. Application of the procedure yields estimates of the infant mortality rate for a series of points preceding the census. Both the value of the rate and the number of years preceding the census are derived from the statistics. Figures for ages 15–74 yield a series of twelve estimates ranging from slightly less than 30 years to about one year before the census. Each age group provides one estimate. Older age groups yield estimates further removed from the census, with an average of approximately $2\frac{1}{4}$ years between estimates.

The ideas underlying the new procedure are a natural extension of concepts introduced by W. Brass and subsequently developed by Brass, Coale, Sullivan, Trussell, and others. Relevant details and references are given in the following section. The new procedure differs from those previously proposed in two respects. It results in a dated estimate of the infant mortality rate from each age group of women, and it does not assume knowledge of the rate of change of infant mortality during the years before the census. It thus opens the possibility of estimating from the figures both the level of infant mortality at the time of the census and the rate of change during the years preceding the census.

## BACKGROUND

In the estimation procedure developed in the following section some relatively complex technical methods developed over the past 20 years by Brass are used. The exposition in the literature[1] is exceedingly concise, however, and the broad outlines of the approach are obscured by the emphasis on the Brass 'multipliers,' which represent an ingenious procedure for obtaining approximate results with relatively little computation. In this section we describe and explain Brass's technique which most readers will find essential for reading the following section.

### *Mortality Estimation from Child Survivorship Data*

The proportion of deceased children among all children born to women in a given age group at

* Research Associate, East-West Population Institute, The East-West Center, Honolulu, Hawaii.

† This work has been supported by East-West Center and by Grant Number 1 ROL HD-09927-1 from the National Institute of Child Health and Human Development. My thanks and gratitude are due to the Department of Statistics of Malaysia, to the staff of the Population Institute and research assistant Stephen E. Wilson for their able assistance, and to William Brass, Lee-Jay Cho, Ansley J. Coale, Kenneth Hill, Vasantha Kandiah, James A. Palmore, Jr., Robert D. Retherford and T. James Trussell, for reading drafts of this material and supplying comments.

[1] W. Brass and A.J. Coale, 'Methods of Analysis and Estimation', in *The Demography of Tropical Africa* (Princeton, New Jersey, 1968), pp. 104–109. W. Brass, 'Methods for Estimating Fertility and Mortality from limited and Defective Data,' (Chapel Hill, North Carolina, 1975), pp. 50–59.

a given census may be expresses as $\sum_t q(t)c_i(t)$, where $i$ is an index identifying the age group, $c_i(t)$ denotes the proportion of all children born to women in this age group who were born during the $t$-th year preceding the census, and $q(t)$ denotes the proportion of these children who die prior to the census. Assume for the moment that the values $c_i(t)$ are known, that mortality was constant during the years prior to the census for the whole period during which the births in question occurred, that there is no differential mortality by age of mother at birth or between children of women living at the time of the census and children of women deceased at the time of the census, and that the age pattern of mortality conforms to a known one-parameter model life table family. By the second and third assumptions, the proportion $q(t)$ is equal to one minus the number of person years lived between exact age $t-1$ and exact age $t$ in the life table with radix one describing the mortality experienced by the population. Using the standard life table notation $L_x$ for person-years lived between exact ages $x$ and $x+1$, the proportion of deceased children may thus be written $\sum_{t \geqslant 0} (1 - L_t)c_i(t)$.

We may now ask what model life table in the given family will yield, in combination with the known values of $c_i(t)$, the observed proportion of deceased children. If the model family is given in tabular form, as in the Coale-Demeny models,[2] hypothetical proportions of deceased children may be calculated by combining the $c_i(t)$ values with model life table $L_x$ values for various tables. To estimate mortality one simply looks for the model table which yields the observed proportion of deceased children, interpolating between model tables if necessary to obtain values of $L_x$ which yield the desired proportion to a sufficient level of precision. Estimates of any desired life table statistics may then be obtained by a corresponding interpolation between the two model tables.

If the model family is defined by a mathematical formula, as in Brass's models,[3] the model $L_x$ values may be expressed as functions of the model parameter, and we write $L_x(\omega)$ for the $L_x$ value in the model life table family defined by the parameter $\omega$. The estimation of mortality may then be regarded as the solution of the equation

$$Q_i = \sum_{t \geqslant 0} (1 - L_t(\omega))c_i(t) \tag{2.1}$$

where $Q_i$ denotes the proportion of deceased children for women in the $i$-th age group calculated from the census. This conception of the process subsumes the approach using tabular models, for these correspond to defining the functions $L_x(\omega)$ by interpolation between tabulated values. A numerical solution of equation (2.1) may be obtained by standard iterative procedures.

### *The Brass Relational Model Life Table Families*

The relational model life table families developed by Brass will usually be preferable to other available models, both because they are computationally simpler and because they allow the possibility of tailoring the family to particular statistics. There is a substantial theoretical basis to the Brass models, but in the present context we may simply note the necessary formulae.[4]

[2] A. J. Coale and P. Demeny, *Regional Model Life Tables and Stable Populations* (Princeton, New Jersey, 1966).

[3] W. Brass, 'On the Scale of Mortality,' in W. Brass, Ed., *Biological Aspects of Demography* (London, 1971), pp. 72–74. See also N. H. Carrier and J. Hobcraft, *Demographic Estimation for Developing Societies* (London, 1975), Appendix I; and W. Brass, *op. cit.* in footnote 1, pp. 85–96.

[4] W. Brass, *op. cit.*, footnote 3. See also K. Hill and T. J. Trussell, 'Recent Developments in Indirect Mortality Estimation,' *Population Studies*, 31, 2 (July 1977), pp. 313–334. Also known as the 'logit system;' the term 'relational' is used here to call attention to the parallel with the relational Gompertz model of fertility. See W. Brass, 'Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales,' *Journal of the Royal Statistical Society* A (1974), 137, Part 4, pp. 551–553.

Let $\phi$ and $\phi^{-1}$ denote the functions defined by

$$\phi(x) = \tfrac{1}{2} \log_e \frac{(1-x)}{x}, \quad 0<x<1$$

$$\phi^{-1}(y) = [1+e^{2y}]^{-1}, \quad -\infty<y+\infty$$

The parentheses will be omitted where confusion over the argument is unlikely to arise; in particular we write $\phi l_x$ in place of $\phi(l_x)$.

Given any schedule of life table $l_x$ values, referred to in this context as the 'standard' schedule and denoted $l_x^s$, a two-parameter model life table family is defined by $l_x(A,B) = \phi^{-1}(A + B\phi l_x^s)$, $A$ and $B$ denoting the two parameters. Observe that this is equivalent to $\phi l_x(A,B) = A + B\phi l_x^s$, which states that the transformed $l_x$ values of tables in the family are linearly related. Table 1 shows standard $l_x$ values for Brass's 'general' standard.

Table 1. *Brass's General Standard $l_x$ Values for Single Years of Age to Age 80*

| Age ($x$) | $l_x$ Values for Indicated Ages | | | | |
|---|---|---|---|---|---|
| | $x$ | $x+1$ | $x+2$ | $x+3$ | $x+4$ |
| 0 | 10,000 | 8,499 | 8,070 | 7,876 | 7,762 |
| 5 | 7,691 | 7,634 | 7,590 | 7,554 | 7,526 |
| 10 | 7,502 | 7,475 | 7,448 | 7,422 | 7,394 |
| 15 | 7,363 | 7,323 | 7,280 | 7,233 | 7,183 |
| 20 | 7,130 | 7,073 | 7,013 | 6,951 | 6,889 |
| 25 | 6,826 | 6,766 | 6,705 | 6,645 | 6,585 |
| 30 | 6,525 | 6,465 | 6,406 | 6,345 | 6,285 |
| 35 | 6,223 | 6,160 | 6,097 | 6,032 | 5,966 |
| 40 | 5,898 | 5,829 | 5,759 | 5,687 | 5,612 |
| 45 | 5,535 | 5,455 | 5,373 | 5,287 | 5,198 |
| 50 | 5,106 | 5,009 | 4,909 | 4,805 | 4,697 |
| 55 | 4,585 | 4,470 | 4,351 | 4,227 | 4,099 |
| 60 | 3,965 | 3,823 | 3,676 | 3,524 | 3,369 |
| 65 | 3,210 | 3,049 | 2,886 | 2,719 | 2,551 |
| 70 | 2,380 | 2,202 | 2,023 | 1,846 | 1,671 |
| 75 | 1,500 | 1,335 | 1,177 | 1,027 | 888 |
| 80 | 760 | | | | |

Sources: W. Brass, 'On the Scale of Mortality' by W. Brass, *loc. cit.* in footnote, p. 77, Table 4.
Note: Single-year values over age ten by cubic interpolation on given five-year values. Single-year values between five and 10 by cubic interpolation on given values for ages four, five, ten, and 15.

A one-parameter family may be defined by setting $B = 1$. We shall re-parameterize this family by the infant mortality rate $q_0$ by observing that the value of $A$ for which $1 - q_0 = l_1(A)$ is $\phi(1-q_0) - \phi l_1^s$, whence the re-parameterized family may be written.

$$l_x(q_0) = \phi^{-1}[\phi(1-q_0) + \phi l_x^s - \phi l_1^s] \quad (2.2)$$

*Estimating the Time Distribution of Children Born*

The values $c_i(t)$, $t = 0, 1, \ldots$ in (2.1) may be termed the 'time distribution of children born' to women in the $i$-th age group. The answers to the census questions on children born and children surviving do not provide information on the children's date of birth, nor, indeed, would any answers short of a complete birth history do so. The distribution in time of children born could be tabulated directly if birth histories were available. In this case, however, it is likely that information on age at death for deceased children would also be available, so that mortality rates could be calculated directly and the value of the indirect estimation procedure becomes questionable. Indirect procedures may prove useful even where direct calculations are possible, because of differences in infant mortality among children of mothers of different ages, or response errors on date of birth and date of or age at death, but insufficient experience has been gained as yet to make any general conclusion on this point possible.

Brass has developed a procedure for estimating the time distribution of children born from mean number of children born to women in successive age groups based on the fertility polynomial.[5]

$$f(a) = \begin{cases} k(a-s)(s+33-a)^2 & \text{for } s < a < 33 \\ 0 & \text{otherwise} \end{cases}$$

where $k$ and $s$ denote parameters related to the total fertility rate ($F_t$) and the mean age at childbearing (M) by $k = F_t \div 98826.75$ and $M = s + 13.2$. Brass first derives an expression for the mean number of children born to women in the age group $x$ to $x + 5$, by assuming that the women are uniformly distributed by age within the group,[6] which may be written

$$\frac{k}{n}\left\{\left[\frac{-(33-d_1)^5}{20} + \frac{11(33-d_1)^4}{4} + 98826.75d_1\right] + \right.$$

$$\left. -\left[\frac{-(33-d_2)^5}{20} + \frac{11(33-d_2)^4}{4} + 98826.75d_2\right] + 98826.75d_3\right\} \qquad (2.3.1)$$

where

$$d_1 = \max\{0, \min(33, a+n-s)\} \qquad (2.3.2)$$

$$d_2 = \max\{0, \min(33, a-s)\} \qquad (2.3.3)$$

$$d_3 = \max\{0, (a+n-s) - \max(33, a-s)\} \qquad (2.3.4)$$

Denoting the quantity (2.3.1) by[7] $P(a, n)$, the number of children born $t$ to $t-1$ years before the census to women aged $a$ to $a+n-1$ at the time of the census equals $P(a-t,n)-P(a-t-1,n)$. Division by $P(a, n)$ gives these births as a proportion of all children born to the women by census date, so that

$$c(t) = \frac{P(a-t,n)-P(a-t-1,n)}{P(a,n)}, \quad t = 0, 1, \ldots \qquad (2.4)$$

The constant $k$ in (2.3.1) cancels out in (2.4). The value of $s$ in (2.3.2–4) may be estimated by equating

$$\frac{P(a-5,n;s)}{P(a,n;s)} \qquad (2.5)$$

to the corresponding observed ratio and solving the resulting equation for $s$, where we write $P(a, n; s)$ in place of $P(a, n)$ to indicate explicity the dependence of (2.3.1) on the value of the parameter $s$.

[5] See W. Brass and A. J. Coale, *op. cit.*, footnote 1; and W. Brass, 1975, *op. cit.* in footnote 1.

[6] For a derivation see L. J. Cho and G. Feeney, '*Fertility Estimation by the Own-Children Method, A Methodological Elaboration,*" (Chapel Hill, North Carolina, 1978), pp. 10–12, esp. formulae (35a-b).

[7] It has become more or less standard to index the age groups 15–19, 20–24 ... by 1, 2, ... and to use $P_1$, $P_2$, ... to denote the mean number of children born to women in these age groups. The more explicit notation used here is necessary for the formulae giving the form of the time distribution of children. The term 'mean parity' is often used to mean 'mean number of children born,' and ratios of this form are referred to as 'mean parity ratios'.

In practice, solutions may be read directly from Table 2. Formulae (2.3.2–4) show that the mean parity ratio (2.5) depends only on the quantity $s - a$, which we term the 'displacement' of the age identifying the ratio from $s$. In Table 2 the mean parity ratio for various values of this displacement is tabulated. To obtain the value of $s$ corresponding to any given mean parity ratio (say 0.393) proceed down columns from left to right in the table to the first value greater than the given value (0.396 in this case; note that the values in the table are multiplied by 1000 to eliminate the decimal place) and read off the corresponding value of the displacement $a - s$ (8.6 years in this case). Note that no interpolation is required. The table is constructed so that the value of the displacement $a - s$ corresponding to any interval in the body of the table corresponds to the upper limit of the interval. (The value tabulated is in fact the value of the ratio for the indicated value of the displacement plus 0.05). To estimate the value of $s$ one simply subtracts the displacement $a - s$ from the age $a$ identifying the observed mean parity ratio. Thus, if 0.393 is the mean parity ratio for the age groups 20–24/25–29, $a = 25$ and $s = 25–8.6 = 16.4$ years.

Table 2. *Mean Parity Ratios for Brass's Fertility Polynomial (× 1000)*

| Tenths of a Year | Displacement ($a - s$) in years | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 9 | 64 | 113 | 170 | 233 | 296 | 356 | 412 | 463 | 511 | 554 | 595 | 632 | 667 | 699 |
| 0 | 68 | 118 | 176 | 239 | 302 | 362 | 417 | 468 | 515 | 559 | 599 | 636 | 670 | 702 |
| 1 | 73 | 124 | 182 | 246 | 308 | 368 | 423 | 473 | 520 | 563 | 602 | 639 | 673 | 705 |
| 2 | 77 | 129 | 189 | 252 | 315 | 373 | 428 | 478 | 524 | 567 | 606 | 643 | 677 | 708 |
| 3 | 82 | 135 | 195 | 259 | 321 | 379 | 433 | 483 | 529 | 571 | 610 | 646 | 680 | 712 |
| 4 | 87 | 140 | 201 | 265 | 327 | 385 | 438 | 488 | 533 | 575 | 614 | 650 | 683 | 715 |
| 5 | 92 | 146 | 208 | 271 | 333 | 390 | 443 | 492 | 537 | 579 | 618 | 653 | 687 | 718 |
| 6 | 97 | 152 | 214 | 278 | 339 | 396 | 448 | 497 | 542 | 583 | 621 | 657 | 690 | 721 |
| 7 | 102 | 158 | 220 | 284 | 344 | 401 | 453 | 502 | 546 | 587 | 625 | 660 | 693 | 724 |
| 8 | 107 | 164 | 227 | 290 | 350 | 407 | 458 | 506 | 550 | 591 | 629 | 663 | 696 | 727 |
| 9 | 113 | 170 | 233 | 296 | 356 | 412 | 463 | 511 | 554 | 595 | 632 | 667 | 699 | 729 |

Note: See text for explanation of calculation.

Preston and Palloni have suggested an alternative approach to estimation from child survivorship data in which they use the age distribution of surviving children instead of the time distribution of children born.[8] The approach involves reverse-surviving the surviving children instead of forward-surviving the children born. The age distribution of surviving children may be estimated by matching children to mothers on census household records as is done in the 'own-children' method of fertility estimation. If all surviving children could be properly matched, this approach would eliminate errors resulting from the estimation of the time distribution of children born, and this is its evident advantage. One disadvantage is that it requires special data processing and tabulation operations in the census. Since the greater part of the work consists in matching children to mothers on household records, and since this is being done increasingly in connection with 'own-children' fertility estimation, this disadvantage may be expected to lessen as time goes on and may, in some instances, disappear altogether. A second disadvantage is that the calculated age distribution of surviving children is subject to errors resulting from age misreporting of children and non-matching and mis-matching of children to mothers. Whether these errors are less serious than those introduced in estimating the time distribution of children born is not known, but errors due to faulty estimation of the time distribution of children born are reasonably small despite the frequently unrealistic assumption of constant fertility and conformity to Brass's model fertility schedules.

[8] S. A. Preston and A. Palloni, *'Fine Tuning Brass-Type Mortality Estimates with Data on Ages of Surviving Children,'* Population Bulletin No. 10-1977. New York: United Nations Department of Economic and Social Affairs 1978. See pages 73–74 for a detailed discussion of the advantages and disadvantages of the Preston-Palloni approach. This discussion includes all the disadvantages noted above, except the restriction to younger age groups of women and notes one advantage not noted above, the possibility of application to more general groups of women, as for example groups defined by marital status or parity.

A third disadvantage, which seems to me the most serious of all, is that the procedure applies only to women who are sufficiently young for most of their surviving children to be living with them and thus to be matchable. In practice this would appear to restrict the procedure to women aged under 30. Matching for older women can, of course, be attempted but as the age of the women increases the rate of matching will drop and the calculated age distribution of surviving children will become less trustworthy. Evidence given below indicates (1) that statistics relating to women under age 20 are rendered completely useless by differential infant mortality by age of mother and (2) that numbers of children born and children surviving reported by women aged 30–49 can be as reliable as those reported by women aged 20–29. The restriction to women under age 30 may, therefore, amount to throwing away two-thirds of the available information.

*Brass's Multipliers*

Brass devised a method for obtaining approximate solutions to (2.1) which radically reduces the volume of computation required to produce estimates.[9] Consider the ratio

$$\frac{q(x;\omega)}{\sum_t [1 - L_{t-1}(\omega)]\, c_i(t)} \qquad (2.3)$$

where $q(x;\ \omega)$ denotes the value of $q(x) \equiv 1 - l_x$ in the model life table specified by the parameter value $\omega$. From the assumptions concerning mortality it follows that, for the value of $\omega$ representing the mortality experienced by the population, the numerator of this ratio gives the value of $q(x)$ for the population and the denominator gives the proportion of deceased children among all children born to women in the $i$-th age group. Multiplication of the proportion of deceased children by this ratio will, therefore, give the value of $q(x)$. These observations do not yet obviously advance the cause of mortality estimation, for the ratio depends on $\omega$, which will not at this stage be known. Brass found, however, that when $x$ is suitably chosen in relation to the age group, the ratio is nearly constant with respect to $\omega$, so that $\omega$ may be fixed at an arbitrary value and still yield a value approximately valid for all levels of mortality. The approximate values of $x$ for the age groups 15–19, 20–24, 25–29, 30–34, 35–39, . . . are 1, 2, 3, 5, 10, . . . .

Since the values of $c_i(t)$ estimated by the procedure of the last section depend only on the age group and the value of $s$, the multipliers may be tabulated for each age group and for a series of values of $s$. The simplified estimation procedure in which this table of 'multipliers' is used consists of three logical steps: (1) estimate the value of $s$ for the population; (2) interpolate among the tabulated multipliers to obtain a value corresponding to this value of $s$; (3) multiply the proportion of deceased children for this age group by the interpolated ratio value. The second two steps are repeated for each age group. In practice, the first two steps may be combined, for there is a one-one correspondence between the mean parity ratio and $s$. Instead of first estimating $s$ from the mean parity ratio and then using this $s$ to determine the value of the multiplier, the mean parity ratio values may be incorporated into the multiplier table and one can go directly from the mean parity ratio value to the multiplier value.[10]

*Recent Work*

Brass's multiplier depends on the values $c_i(t)$, $t = 1, 2, \ldots$, and according to the estimation procedure detailed above, these values are determined by the parameter $s$. Since this parameter is estimated from the mean parity ratio for women in two successive age groups, the multiplier is a

[9] W. Brass and A. J. Coale, *op. cit.*, footnote 1, pp. 105–114.

[10] Tables of multipliers are given in W. Brass and A. J. Coale, *op. cit.*, in footnote 1 p. 108 and W. Brass, *loc. cit.* in footnote 3, p. 55. The latter source has the advantage of tabulating $P_2/P_3$ as well as $P_1/P_2$ values.

function of the mean parity ratio. Sullivan investigated this functional relationship by calculating exact values of both the multiplier and the mean parity ratio for all possible combinations of numerous observed fertility schedules and model life tables, generating a total of several thousand values for both quantities. He then regressed the multiplier values on the mean parity ratios for various sub-sets of these observations.[11]

He also developed similar regression results for use with proportions of deceased children among all children born to women classified by duration of marriage. The resulting regression constants provide an alternative means for obtaining multipliers which convert proportions of deceased children to women in a given age group to life table $q(x)$ values.

Trussell subsequently refined this procedure by including further independent variables in the regression and substituting model fertility schedules derived from the Coale-Trussell model for Sullivan's observed fertility schedules.[12] The inclusion of additional variables reduces the standard error of the regressions and the use of the Coale-Trussell model tables introduces a wider range of variation into the fertility patterns on which the regression equations are based.

There have been several attempts to weaken the assumption that mortality has been constant during the years before the census, and this work indicated several ways in which child survivorship estimates may be validly interpreted when mortality is changing.[13] In each case, however, the approximate rate of mortality decline must be specified; hence the trend of mortality is assumed rather than estimated from the data.

The recent work of Preston and Palloni has already been noted above (see footnote 8).

## ESTIMATION PROCEDURE

If mortality is constant, its level may be specified by giving the value of some statistic, such as the infant mortality rate or the expectation of life at birth, without reference to time. If mortality is changing, both a statistic representing the level of mortality and the time at which this level was obtained are required to specify an estimate. This is obvious to the point of triteness, but indirect estimation procedures so often assume constant mortality that the point requires emphasis: a mortality estimate requires both a mortality statistic and a time reference.

### *Estimation Equations for Linear Mortality Decline*

Let $L_x(t)$ denote person-years lived between exact age $x$ and exact age $x + 1$ in the life table with radix one representing mortality risks during the $t$-th year before the census. The proportion of persons born during the first uear preceding the census who survive to the census is simply $L_0(1)$. The proportion of persons born during the second year before the census who survive to the end of this year is $L_0(2)$, and the proportion of these survivors who survive one further year and so are alive at the time of the census is $L_1(1)/L_0(1)$, so the proportion of persons born during the second year preceding the census who survive to the census is the product $L_0(2)L_1(1)/L_0(1)$. In general, the proportion of persons born during the $t$-th year prior to the census who survive to

[11] J. M. Sullivan, 'Models for the Estimation of the Probability of Dying Between Birth and Exact Ages of Childhood,' *Population Studies*, **26**, (1)(March 1972), pp. 82–83.

[12] T. J. Trussell, 'A Re-Estimation of the Multiplying Factors for Determining Childhood Survival,' *Population Studies*, **29**, (1)(March 1975), pp. 97–107. The Coale-Trussell model is discussed in A. J. Coale and T. J. Trussell, 'Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations,' *Population Index*, **40**, (2)(April 1974), pp. 195–258.

[13] J. Sullivan, 'Mortality Estimates Derived from Retrospective Mortality Data During Periods of Fluctuating Mortality, *Demografi Indonesia*, 2, (1) (December 1974) pp. 116–113; W. Brass, *op. cit.*, in footnote 1, pp. 56–59; E. P. Kraly and D. A. Norris, 1976, 'An Evaluation of Brass Mortality Estimates under Conditions of Declining Mortality,' *Demography*, **15**, (4)(November 1978), pp. 549–557.

the census is obtained by surviving this cohort forward one year at a time using the $L_x$ survivorship ratios from the appropriate life table. This gives

$$q(t) = 1 - L_0(t) \prod_{i=1}^{t-1} \frac{L_i(t-i)}{L_{i-1}(t-i)} \tag{3.1}$$

where $q(t)$ denotes the proportion of persons born during the $t$-th year before the census who do not survive to the time of the census.

Assume now that the life table representing mortality during each year preceding the census conforms to a one-parameter model life table family and that the infant mortality rate has been declining linearly at the rate of $r$ infant deaths per thousand births per year during the years preceding the census. Let $\omega$ denote the infant mortality rate at the time of the census. Suppose further that the life table for each year is defined by the infant mortality rate at the mid-point of this year. Arbitrarily given values for $r$ and $\omega$ define a linear trend of mortality before the census and hence, in particular, an infant mortality rate at the mid-point of each year preceding the census. By the model life table family assumption, these infant mortality rates define a complete life table and, in particular, values of $L_x$ for single years of age for this year. Finally, these $L_x$ values may be substituted in formula (3.1) to yield values for the proportions $q(t)$, $t = 1, 2, \ldots$ . We have thus shown how to calculate the values of $q(t)$ which would result from any linear trend in the infant mortality rate. The dependence of $q(t)$ on $r$ and $\omega$ may be symbolized by writing $q(t; r, \omega)$ in place of $q(t)$. Substitution in (2.1) gives the system

$$Q_i = \sum_{t=1}^{n(i)} q(t; r, \omega) c_i(t), \quad i = 1, \ldots, N, \tag{3.2}$$

of $N$ equations in the unknowns $r$ and $\omega$. The upper limit of summation $n(i)$ is to be defined so that $c_i(t) = 0$ for $t > n(i)$ and may be taken as the least integer greater than the upper limit of the age group less the lower limit of the reproductive span. In the case of five-year age groups beginning with ages 15–19 for $i = 1$ and the lower limit of the reproductive span equal to ten, $n(i) = 5(i + 1)$.

The equations (3.2) are in several respects analogous to those based on the assumption of constant mortality. In both cases a single equation for each age group of women is formed by equating the observed proportion of deceased children among children born to women in this age group to the expression of this proportion in terms of the values $q(t)$ and $c_i(t)$, an equation justified by formal identity. In both cases, the values of $c_i(t)$ must be estimated before attempting any solution, and in both cases the values of $q(t)$ are replaced by a parametric expression derived from a one-parameter model life table family. Demographically, the sole difference between the two cases is the assumption of linearly declining as opposed to constant mortality. Mathematically, the sole difference is in the number of unknown parameters in the equations, one in the case of constant mortality, two in the case of linearly declining mortality.

Since data will normally be available for substantially more than two age groups, the estimation equations will usually be overdetermined. There will not, therefore, generally exist values of $\omega$ and $r$ which satisfy every equation. The mathematical structure of the situation suggests three options at this point. The first is to select some, perhaps all, sub-systems of two equations and attempt to solve each sub-system simultaneously for values of $\omega$ and $r$. Since there are $N(N-1)/2$ ways of selecting two from $N$ equations, this approach can yield a total of $N(N-1)/2$ pairs of solutions for $\omega$ and $r$ – 66 pairs, for example, for the case of age groups from 15–19 to 70–74.

It is undoubtedly a nuisance to have 66 potentially different answers for a single problem, but this redundancy may provide valuable information. If the data were perfectly free from error and if the assumptions underlying the estimation equations were strictly valid, each pair of equations would upon solution yield precisely the same values for $\omega$ and $r$. As a matter of pure logic, then, if different pairs of equations yield different values for $\omega$ and $r$, either or both of these propositions must be false, and if this is the case it is in the interest of the analyst to be so warned. It may, in addition, be possible to interpret the dispersion of estimates obtained and make inferences concerning the nature of data errors or the invalidity of the assumptions.

The second approach to a solution would be to choose values of $\omega$ and $r$ which minimize some measure of discrepancy between the 'expected' proportions of deceased children, the expressions on the right in (3.2), and the observed proportions $Q_i$. This approach obviates the need to choose among many different estimates, while at the same time providing some warning of data errors or invalid assumptions in the magnitude of the measure of discrepancy attached to the minimizing values for $\omega$ and $r$. It does introduce the problem of choosing a measure of discrepancy, however. The proper choice may depend on the particular problem at hand, and it will generally be necessary to use iterative numerical methods to find the minimizing values. The one situation where this approach may clearly be preferable is where the numbers of deceased children are very small and hence significantly affected by random variation. In this situation one might expect estimates of $\omega$ and $r$ obtained by minimization of some discrepancy measure to have lower variance than alternative procedures, a conjecture which might be tested by resorting to simulation.

The third approach to the solution of the estimation equations is to calculate a 'solution set' for the equation for each age group. Restricting attention for the moment to a single age group, we give $r$ some assigned value, yielding an equation which may be solved for the single remaining unknown $\omega$. We then repeat this process for a series of values of $r$ covering the plausible range of empirical possibilities. This process yields a series of combinations of values for $r$ and $\omega$, each satisfying the equation. The totality of such combinations is infinite, for there is one for every possible value for $r$, but this totality may be represented by interpolation between the calculated values. It is heuristically useful in this connection to imagine plotting the calculated combination of values for $r$ and $\omega$ on coordinate axes. The solution set as a whole may then be visualized as a line in the plane passing through the plotted points. The entire process may then be repeated with the equation for each age group.

There is a potential technical difficulty in this approach. The solution set might conceivably be very 'bumpy,' so that intermediate points would not be well approximated by interpolation. As it happens, however, this difficulty does not arise, for a simple demographic argument shows that $\omega$ must decrease as $r$ increases. Recall that each point in the solution set represents a linear mortality trend consistent with the observed proportion of deceased children. Increasing $r$ means a more rapid decline in mortality, and if a more rapid decline in mortality is to yield the same proportion of deceased children, then the mortality level at the time of the census must be lower. If this were not the case, the solution set would include two mortality trends one of which represented a higher level of mortality at every moment of time prior to the census than the other, and this is impossible since two such trends would necessarily give different porportions of deceased children. This shows that the solution set is 'smooth' and that approximation by interpolation between a series of calculated points in valid.

### *Mortality Estimates as Intersections of Consistent Linear Trends*

In a technical sense the solution set of the estimation equation for any age group exhausts the information on mortality trends contained in the proportion of deceased children for this age group. It may seem peculiar, even paradoxical, to say that the solution set contains information on mortality trends when it specifies an infinite number of possible trends and provides no guide as to which of these trends will actually have operated. The calculated solution set does rule out some (in fact, an infinite number of) mortality trends as inconsistent with the observed data, however, and it does, therefore, provide information on mortality. Indeed, as just stated, it provides all possible information derivable from the data of a single age group.

Conceding this point, we may very well remain sceptical of the practical value of information which comes in so unwieldly a form. What use can it be to be supplied with an infinite number of answers to the question: What has been the trend of mortality? It turns out that the consistent linear trends of mortality specified by the solution set have, to a very close approximation, a common point of intersection a certain number of years prior to the census. Consider, for example, the 1970 Census of Malaysia. Women aged 25–29 reported 745,983 children born and 703,487 children surviving, yielding a proportion deceased of 0.0570. On calculating the solution set of the equation for this age group we find, that, if the infant mortality rate had been constant prior to the census, it must have been $\omega = 38.8$ infant deaths per thousand births; if it had been declining by $r = 0.001$ per year, then the level at the time of the census must have been $\omega = 34.5$ and so forth, calculated values of $\omega$ for $r = 2, \ldots,$ being as indicated below

| $r$ | $\omega$ |
|---|---|
| 2 | 30.3 |
| 3 | 26.2 |
| 4 | 22.1 |
| 5 | 18.0 |

Each of these combinations of values for $r$ and $\omega$ specifies a linear trend of infant mortality rates for which the level of mortality at the time of the census is $\omega$ and for which the rate of decline during the years before the census is $r$, where $r = 0, 1, \ldots, 5$. When plotted, these trends have a common point of intersection a certain number of years before the census. They, therefore, collectively determine the value of the infant mortality rate at that point in time. The estimated rate is 38.8, the value corresponding to $r = 0$. The value of the years-prior-to-the-census figure may be read off approximately from the graph and is 4.1 years. The census was taken at 1970.7, where time is expressed in decimal form to tenths of a year, hence this estimate corresponds to 1966.6, or *circa* August 8, 1966.

To calculate the coordinates of the intersection of any two trends we put $\omega + rt = \omega' + r't$ and solve for $t$. The time at which the intersection occurs is thus found to be $t = (\omega - \omega') \div (r' - r)$ years before the census. The infant mortality rate at this time may be calculated either as $\omega + rt$ or $\omega' + r't$ (both as a computational check if working by hand). The six calculated trends in the above example would yield a total of $6(6 - 1)/2 = 15$ intersections, each specifying the infant mortality rate a certain number of years prior to the census. Final values for the infant mortality rate and the number of years prior to the census may be obtained by averaging these two sets.

I have calculated a series of infant mortality estimates for 14 data sets with the highest age group varying between 45–49 and 70–74. Six consistent linear trends were calculated for each age group and the estimated infant mortality rate was calculated as the average of the ordinates of the set of all intersections of these six consistent trends. The number of years prior to the census to which this estimate applies is calculated as the average of the abscissae of these intersections. The intersections, though tightly clustered, do not coincide exactly, and both statistics are to some extent imprecise on this account. The level of imprecision of the estimated infant mortality rate may be measured by one hundred times half the range of the values divided by the average value, and similarly for the estimated number of years prior to the census. These figures may be regarded as approximate relative errors, expressed as percentages of the estimated infant mortality rates and years-prior-to-census values. The value of this relative error statistic varies between 0.00 and 4.27 per cent over all age groups in the fourteen data sets from which estimates have been made. The values typically increase with age within each set, however, and for all age groups below 50 years the error is one per cent or less.[14]

[14] For previous discussions of this estimation procedure see G. Feeney, 'Estimating Infant Mortality Rates from Child Survivorship Data by Age of Mother,' *Asian and Pacific Census Newsletter*, 3 (2)(November 1976), pp. 15–16 and 'Estimation of Demographic Parameters from Census and Vital Registration Data,' in International Union for the Scientific Study of Population, *International Population Conference: Mexico 1977* (Liège, 1977) Volume III, pp. 349–370.

*Tabular Solutions to the Estimation Equations*

In the equation $Q_i = \Sigma q(t; r, \omega) c_i(t)$ the values of $c_i(t)$, when estimated as described above, depend only on the age group (represented by the index $i$) and the value of the parameter $s$ in Brass's polynomial fertility model. Values for $Q_i$ and $s$ therefore determine the solution set of $(r, \omega)$ values and, consequently, an estimated infant mortality rate a certain number of years prior to the census. We may, therefore, produce a tabulation of both variables corresponding to specified ranges of $Q_i$ and $s$ values for each age group.

In Table 3 values of the infant mortality rate and the years prior to the census corresponding to selected values of $Q$ and $s$ for each age group are shown, based on Brass's general standard model life table family. This table may be used to obtain solutions of the estimation equations (3.2) by first estimating the value of $s$ and then performing double interpolations to obtain the infant mortality rate and years-prior-to-the-census value.

## ERRORS IN THE ESTIMATES

Mortality estimates produced by the procedure described in the preceding section are subject to two kinds of errors, those due to errors in the child survivorship data, and those due to the invalidity of the various assumptions made. Both types of errors have been discussed in detail elsewhere,[15] hence we need only consider matters peculiar to the new procedure or those which experience suggests require further emphasis.

Under the latter heading it should be pointed out that although the estimation procedure assumes constant fertility, the estimates are so robust against departures from this assumption that this will be a negligible factor, unless errors due to other causes are extremely small. The tendency to avoid child survivorship estimates because of the 'constant fertility' assumption (and because fertility is known or felt to have been changing, perhaps substantially) is naive. The insensitivity to erroneous values of the parameter $s$ is evident in Table 3.[16]

Another source of error which has been noted but requires more emphasis is differential infant mortality by age of mother (it is immaterial here whether or not this differential reflects any causal relation between infant mortality and age of mother at birth). Infant mortality rates by age of mother at birth are generally available only from special surveys or matching studies, for the age of mothers of the decedent is not normally recorded on a death certificate. Some information is available for a considerable diversity of populations, however, and consistently suggests that infant mortality rates are higher for children of mothers at the extremes of the reproductive age span.

The effect of differential infant mortality by age of mother on child survivorship estimates may be gauged by calculating infant fertility rates for children born to mothers below specified ages, as in Table 4, and observing how the rate changes as the age limit of the women approaches the end of the reproductive span. The data restrict Table 4 to giving rates for children of women aged 40 or younger, rather than an age higher than the end of the reproductive span. However, relatively few children are born to women over 40 and the column on the extreme right of the table may be considered in this context as giving the infant mortality rate for children of mothers of all ages. Infant mortality rates for children born to women aged 20 or younger at birth substantially exceed the overall rates in nine out of ten cases and the differences are strikingly high. It seems clear that information from the 15–19 age group should be considered useless for mortality estimation.

[15] W. Brass and A.J. Coale, *op. cit.* footnote 1, pp. 111–119.

[16] And also from the table of multipliers on p. 108 of W. Brass and A.J. Coale, *op. cit.*, footnote 1. The tendency may derive from an exaggerated notion of the level of precision which may be expected from indirect estimation procedures. One should not expect a smaller relative error than five to ten per cent, though, of course, greater precision may be obtained in particular cases. Note, incidentally, that there is always a 'correct' value of $s$ even though the actual age pattern of mortality may not conform to Brass's fertility polynomial.

Table 3. *Infant Mortality Rates and Years-Prior-to-Census Values Estimated from Proportions of Surviving Children for Women in Quinquennial Age Groups and Mean age at Childbearing: Based on Brass's General Standard Model Life Table Family*

| | Infant mortality rate | | | | | | | Years prior to census | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Age Group 15–19 | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 5 | 43.7 | 46.3 | 49.4 | 53.4 | 58.7 | 65.8 | 5 | 1.73 | 1.47 | 1.23 | 1.01 | 0.80 | 0.62 |
| 10 | 88.6 | 93.9 | 100.0 | 107.7 | 118.2 | 132.0 | 10 | 1.72 | 1.46 | 1.22 | 1.00 | 0.79 | 0.61 |
| 15 | 134.9 | 142.7 | 151.6 | 163.0 | 178.3 | 198.4 | 15 | 1.70 | 1.44 | 1.21 | 0.99 | 0.78 | 0.61 |
| 20 | 182.6 | 192.8 | 204.4 | 219.2 | 239.0 | 265.2 | 20 | 1.69 | 1.43 | 1.20 | 0.98 | 0.78 | 0.60 |
| | | | | | Age Group 20–24 | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 5 | 35.6 | 36.7 | 38.0 | 39.6 | 41.4 | 43.7 | 5 | 3.34 | 2.98 | 2.64 | 2.32 | 2.01 | 1.73 |
| 10 | 72.6 | 74.8 | 77.4 | 80.5 | 84.2 | 88.6 | 10 | 3.33 | 2.97 | 2.63 | 2.50 | 2.00 | 1.72 |
| 15 | 111.1 | 114.4 | 118.3 | 122.9 | 128.4 | 134.9 | 15 | 3.32 | 2.96 | 2.62 | 2.29 | 1.89 | 1.70 |
| 20 | 151.2 | 155.6 | 160.7 | 166.7 | 173.9 | 182.6 | 20 | 3.30 | 2.94 | 2.60 | 2.28 | 1.97 | 1.69 |
| | | | | | Age Group 25–29 | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 5 | 31.7 | 32.4 | 33.0 | 33.8 | 34.6 | 35.6 | 5 | 5.35 | 4.92 | 4.50 | 4.10 | 3.71 | 3.34 |
| 10 | 64.9 | 66.1 | 67.5 | 69.0 | 70.7 | 72.6 | 10 | 5.74 | 4.91 | 4.49 | 4.09 | 3.70 | 3.33 |
| 15 | 99.6 | 101.4 | 103.5 | 105.8 | 108.3 | 111.1 | 15 | 5.33 | 4.90 | 4.48 | 4.08 | 3.69 | 3.32 |
| 20 | 135.9 | 138.4 | 141.1 | 144.1 | 147.4 | 151.2 | 20 | 5.32 | 4.89 | 4.47 | 4.07 | 3.68 | 3.30 |
| | | | | | Age Group 30–34 | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 5 | 29.1 | 29.6 | 30.1 | 30.6 | 31.1 | 31.7 | 5 | 7.75 | 7.24 | 6.75 | 6.26 | 5.80 | 5.35 |
| 10 | 59.5 | 60.5 | 61.6 | 62.6 | 63.7 | 64.9 | 10 | 7.73 | 7.22 | 6.73 | 6.25 | 5.79 | 5.34 |
| 15 | 91.5 | 93.0 | 94.6 | 96.2 | 97.8 | 99.6 | 15 | 7.73 | 7.22 | 6.72 | 6.24 | 5.78 | 5.33 |
| 20 | 125.2 | 127.2 | 129.2 | 131.3 | 133.5 | 135.9 | 20 | 7.73 | 7.22 | 6.92 | 6.24 | 5.77 | 5.32 |
| | | | | | Age Group 35–39 | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 10 | 54.5 | 55.5 | 56.5 | 57.5 | 58.5 | 59.5 | 10 | 10.53 | 9.92 | 9.35 | 8.80 | 8.26 | 7.73 |
| 15 | 84.0 | 85.5 | 87.0 | 88.6 | 90.0 | 91.5 | 15 | 10.53 | 9.94 | 9.37 | 8.81 | 8.26 | 7.73 |
| 20 | 115.1 | 117.2 | 119.2 | 121.2 | 123.2 | 125.2 | 20 | 10.55 | 9.96 | 9.38 | 8.81 | 8.27 | 7.74 |
| 25 | 148.1 | 150.6 | 153.2 | 155.6 | 158.1 | 160.6 | 25 | 10.58 | 9.97 | 9.39 | 8.82 | 8.27 | 7.73 |

Table 3 continued

Infant mortality rate | Years prior to census

Age Group 40–44

| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 49.2 | 50.3 | 51.4 | 52.5 | 53.5 | 54.5 | 10 | 13.64 | 13.01 | 12.38 | 11.73 | 11.13 | 10.53 |
| 15 | 75.9 | 77.5 | 79.2 | 80.8 | 82.4 | 84.0 | 15 | 13.65 | 13.01 | 12.37 | 11.75 | 11.13 | 10.53 |
| 20 | 104.4 | 106.6 | 108.8 | 110.9 | 113.1 | 115.2 | 20 | 13.72 | 13.06 | 12.42 | 11.78 | 11.16 | 10.55 |
| 25 | 134.7 | 137.4 | 140.2 | 142.9 | 145.5 | 148.1 | 25 | 13.79 | 13.12 | 12.47 | 11.82 | 11.19 | 10.58 |

Age Group 45–49

| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 43.8 | 44.8 | 45.9 | 47.0 | 47.8 | 49.2 | 10 | 16.81 | 16.20 | 15.58 | 14.95 | 14.12 | 13.64 |
| 15 | 67.8 | 69.3 | 70.9 | 72.6 | 74.3 | 75.9 | 15 | 16.90 | 16.24 | 15.61 | 14.97 | 14.32 | 13.65 |
| 20 | 93.3 | 95.5 | 97.7 | 99.9 | 102.1 | 104.4 | 20 | 16.95 | 16.32 | 15.68 | 15.03 | 14.37 | 13.72 |
| 25 | 120.9 | 123.6 | 126.4 | 129.1 | 131.9 | 134.7 | 25 | 17.10 | 16.45 | 15.80 | 15.13 | 14.46 | 13.79 |

Age Group 50–54

| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 38.7 | 39.6 | 40.6 | 41.7 | 42.7 | 43.8 | 10 | 19.69 | 19.07 | 18.53 | 17.98 | 17.40 | 16.81 |
| 15 | 60.0 | 61.5 | 63.0 | 64.6 | 66.2 | 67.8 | 15 | 19.72 | 19.19 | 18.65 | 18.08 | 17.50 | 16.90 |
| 20 | 83.0 | 85.0 | 87.1 | 89.2 | 91.2 | 93.3 | 20 | 19.86 | 19.33 | 18.77 | 18.20 | 17.55 | 16.95 |
| 25 | 107.7 | 110.3 | 112.9 | 115.6 | 118.2 | 120.9 | 25 | 20.03 | 19.48 | 18.92 | 18.33 | 17.72 | 17.10 |

Age Group 55–59

| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 53.0 | 54.4 | 55.8 | 57.2 | 58.6 | 60.0 | 15 | 22.19 | 21.75 | 21.29 | 20.82 | 20.33 | 19.72 |
| 20 | 72.2 | 75.3 | 77.2 | 79.1 | 87.1 | 83.0 | 20 | 21.51 | 21.93 | 21.46 | 20.98 | 20.48 | 19.87 |
| 25 | 94.4 | 97.9 | 100.4 | 102.8 | 105.4 | 107.7 | 25 | 22.05 | 22.13 | 21.65 | 21.16 | 20.65 | 20.03 |
| 30 | 118.5 | 122.5 | 125.5 | 128.5 | 131.6 | 134.7 | 30 | 22.43 | 22.35 | 21.86 | 21.35 | 20.83 | 20.28 |

Age Group 60–64

| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 46.5 | 47.8 | 49.1 | 49.9 | 50.8 | 53.0 | 15 | 24.34 | 25.97 | 23.59 | 22.43 | 21.41 | 22.19 |
| 20 | 64.6 | 66.3 | 68.1 | 69.9 | 71.2 | 72.2 | 20 | 24.54 | 24.17 | 23.77 | 23.36 | 22.53 | 21.50 |
| 25 | 84.2 | 86.5 | 88.8 | 91.0 | 93.4 | 94.4 | 25 | 24.78 | 24.41 | 24.01 | 23.60 | 23.16 | 22.05 |
| 30 | 105.7 | 108.5 | 111.3 | 114.1 | 116.9 | 118.5 | 30 | 25.05 | 24.67 | 24.27 | 23.85 | 23.40 | 22.44 |

Table 3 continued

| Infant mortality rate | | | | | | | Years prior to census | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Age Group 65–69 | | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 15 | 40.4 | 41.6 | 42.8 | 44.0 | 45.3 | 46.5 | 15 | 25.87 | 25.62 | 25.33 | 25.03 | 24.69 | 24.34 |
| 20 | 56.3 | 57.8 | 59.5 | 61.2 | 62.9 | 64.6 | 20 | 26.35 | 25.81 | 25.53 | 25.22 | 24.89 | 24.54 |
| 25 | 73.6 | 75.8 | 77.9 | 80.1 | 82.0 | 84.2 | 25 | 26.59 | 26.32 | 26.02 | 25.69 | 25.13 | 24.78 |
| 30 | 92.6 | 95.3 | 98.0 | 100.6 | 103.3 | 105.7 | 30 | 26.88 | 26.61 | 26.31 | 25.98 | 25.62 | 25.05 |
| | | | | Age Group 70–74 | | | | | | | | | |
| Q/S | 13 | 14 | 15 | 16 | 17 | 18 | Q/S | 13 | 14 | 15 | 16 | 17 | 18 |
| 15 | 34.4 | 35.6 | 36.8 | 38.0 | 39.2 | 40.4 | 15 | 26.71 | 26.61 | 26.48 | 26.31 | 26.11 | 25.87 |
| 20 | 48.1 | 49.7 | 51.4 | 53.0 | 54.6 | 56.3 | 20 | 27.30 | 27.18 | 27.02 | 26.83 | 26.60 | 26.35 |
| 25 | 62.9 | 65.0 | 67.2 | 69.3 | 71.5 | 73.6 | 25 | 27.49 | 27.38 | 27.23 | 27.05 | 26.84 | 26.59 |
| 30 | 79.8 | 82.4 | 85.0 | 87.3 | 89.9 | 92.6 | 30 | 28.21 | 28.07 | 27.89 | 27.34 | 27.13 | 26.88 |

As regards mortality fluctuations, it should be pointed out that child survivorship data involve a natural filtering effect which filters out high-frequency mortality fluctuations. To illustrate, consider a population which experienced infant mortality rates of 50 and 100 in alternate years preceding the census. Since the children born to women in each age group are born over a period of many years, these children would have experienced both the high and the low mortality risks and the proportion who die before the census will be approximately the same as if the level

Table 4. *Infant Mortality Rates for Births to Women Below Ages 20, 25, 30, 35, and 40: Selected Populations*

| Population | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|
| Argentina[a] | 1,335 | 979 | 845 | 817 | 801 |
| Brazil[a] | 1,041 | 752 | 657 | 648 | 651 |
| Canada[a] | 212 | 190 | 179 | 180 | 182 |
| Chile[a] | 793 | 601 | 546 | 545 | 552 |
| El Salvador[a] | 1,166 | 961 | 889 | 876 | 884 |
| Mexico[a] | 863 | 665 | 582 | 579 | 607 |
| California[a] | 262 | 190 | 177 | 175 | 175 |
| Total U.S.A.[b] | 331 | 268 | 252 | 249 | 250 |
| U.S.A. White[b] | 282 | 233 | 221 | 219 | 220 |
| U.S.A. Non-white[b] | 500 | 440 | 420 | 414 | 413 |
| Israel[c] | 232 | 217 | 224 | 235 | 243 |
| Bangladesh[d] | 1,384 | 1,213 | 1,189 | 1,186 | 1,198 |

*Notes*

[a] Ruth R. Puffer, and C. V. Serrano, *Patterns of Mortality in Childhood,* Table 147, p.245 (Washington, D.C.: 1973) Pan American Health Organization, Scientific Publication No. 262.

[b] U.S. National Center for Health Statistics, *'A Study of Infant Mortality from Linked Records by Age of Mother, Total-Birth Order, and Other Variables: United States, 1960 Live Birth Cohort,'* Vital and Health Statistics, Data from the National Vital Statistics System, Series 20, No. 14. (Washington, D.C.: 1973) Table C, p.11.

[c] H. Pertiz, and U. O. Schmelz, Eds. *Late Fetal Deaths and Infant Mortality: 1948–1972,* Special Series No. 453 (Jerusalem: Central Bureau of Statistics 1974), Table 18, p. 164.

[d] H. Stoeckel and A. K. M. Alauddin Chowdhury, 'Neo-natal and Post-Neo-Natal Mortality in a Rural Area of Bangladesh,' *Population Studies,* 26, 1972, Table 3, p. 117.

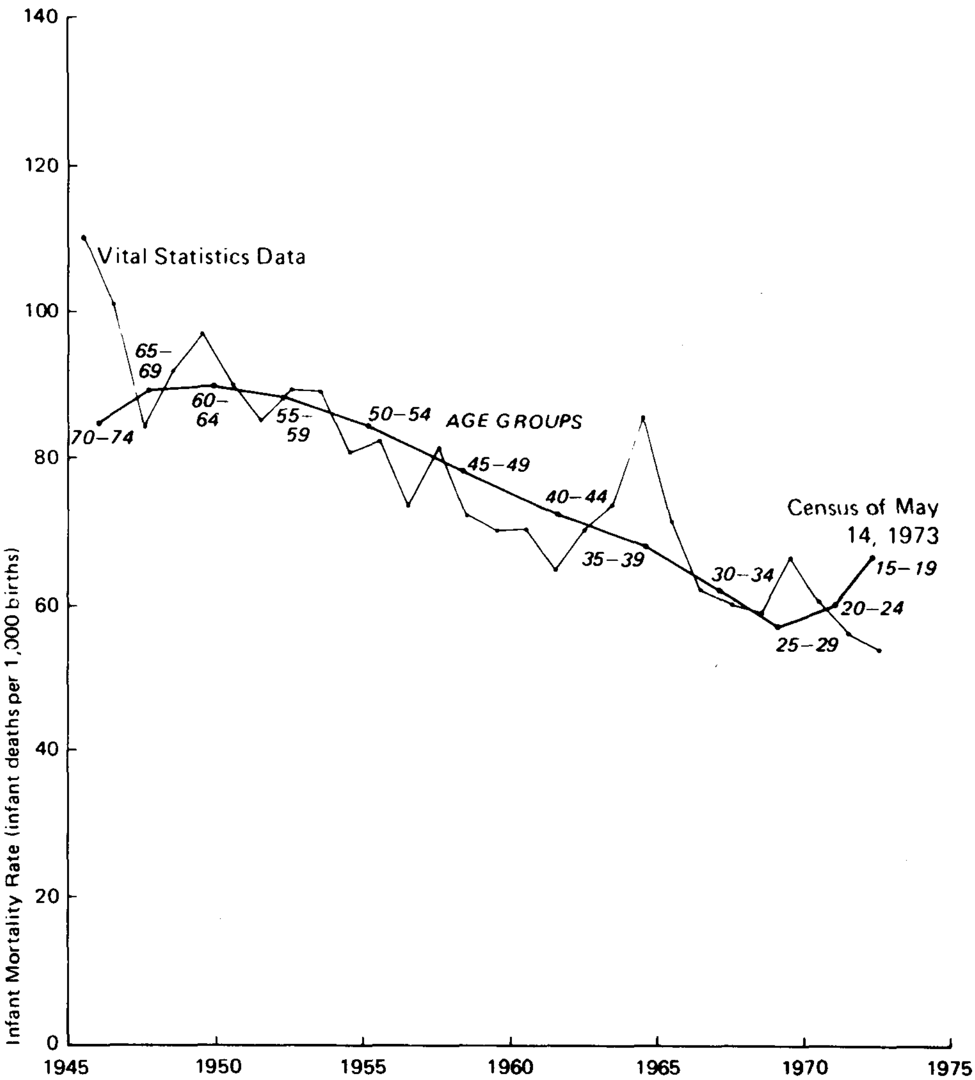
of mortality had been constant at 75 infant deaths per thousand births. The estimated infant mortality rates at specific points in time will therefore be either 25 points too high or 25 points too low. In many applications one will be more concerned with the long-term trend of mortality than with short-term fluctuations, and in such cases this 'estimation error' may actually represent a convenient and automatic smoothing of the time series.

The model life table family assumption is complicated by the consideration of mortality change, for the age pattern of mortality may change over time in such a way that the appropriate model family changes also.[17] This observation suggests the need for mortality models which combine age and time variations.

## APPLICATION TO COSTA RICA AND PENINSULAR MALAYSIA

In Figures 1 and 2 we show child survivorship estimates for Costa Rica and Peninsular Malaysia compared with vital registration figures. The latter have been taken from the United Nations *Demographic Yearbooks.* Both the Costa Rican and the Malaysian statistics show an upturn just prior to the census, reflecting estimates from the 15–19 and 20–24 age groups higher than the estimate for the 25–29 age group. It is quite clear that this apparent rise in mortality is spurious

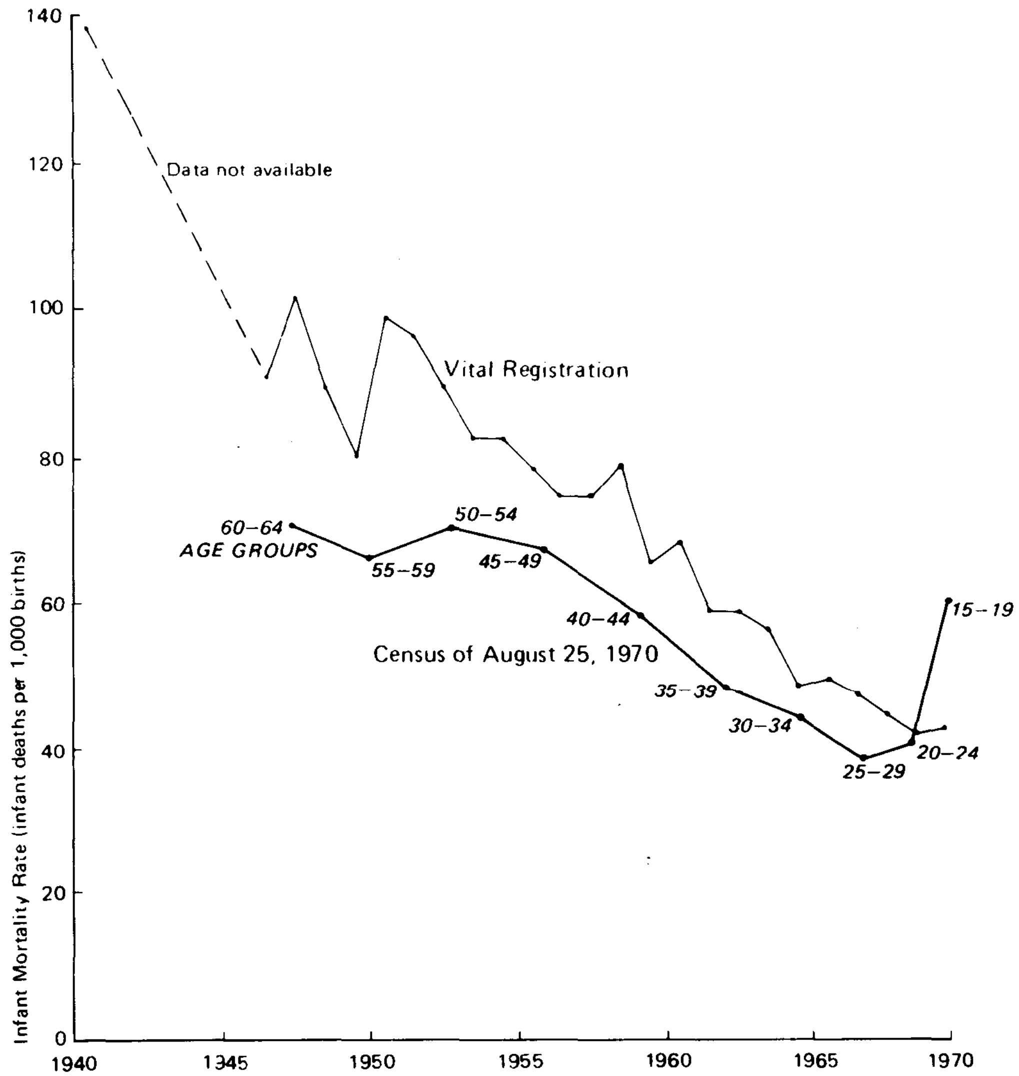
[17] See J. M. Sullivan, 'The Influence of Cause-Specific Mortality Conditions on the Age Pattern of Mortality with Special Reference to Taiwan,' *Population Studies,* 27, (1)(March 1973), p. 141.




Note: Vital registration data tabulated by year of registration prior to 1963

Sources: Vital registration data, Table 6; Census estimation, Table 5

Figure 1. Infant Mortality Rates Estimated from Child Survivorship Data Compared with Rates Calculated from Vital Registration Data: Costa Rica 1945–1972




Sources: Vital registration data, Table 8; Census estimation, Table 5

Figure 2. Infant Mortality Rates Estimated from Child Survivorship Data Compared with Rates Calculated from Vital Registration Data: Peninsular Malaysia 1947–1969

and that the estimates based on the 15–19 age groups are biased upward by the relatively high mortality of children of young mothers. The difference between the estimates from the 20–24 and 25–29 age groups might be explained in two ways, as caused by response error depressing the value of the 25–29 age group estimate or by differential infant mortality by age of mother biasing the value of the 20–24 age group estimate upwards. The comparison with vital statistics rates for Costa Rica suggests that the level of the child survivorship estimates is correct for the age groups over 25, and this suggests that the latter explanation holds true. The Malaysian estimate for the 25–29 age group is substantially lower than the vital statistics rate. Examination of the figures for sub-populations shows that the low value for the age groups over 25 is largely due to the Chinese population for which response error is evidently very high.

Perhaps the most surprising observation to be made in Figure 1 is the absence of any indication of significant differences in response error between ages 20 and 50. The estimates for Costa Rica look, with the exception of the turns in the tails, very much like a series obtained from the vital statistics rates by smoothing. The Malaysian estimates derived from the 25–50 age groups are low, but the estimate from the 45–49 age group is not noticeably inferior to that from the 25–29 age group. Indeed, the relative error is higher for the former group.

This observation suggests that the widespread practice of disregarding information for women over age 30 or 35 may be a mistake. It is plausible that women's memories begin to fade as they pass into extreme old age, but it is not particularly plausible that women forget how many children they have had while they are still of reproductive age, particularly in social situations where the bearing and raising of children is the primary role of women and the principal basis for their social status. It would be foolish to suppose that women necessarily report correctly, but a stance of blind distrust of information given by older women is as inappropriate as a stance of blind trust. It is by no means clear, that response errors are caused by memory lapse, which seems to have given rise to the notion that they should increase with age.

## CONCLUSION

Questions on child survivorship have been included in dozens of national population censuses over the past 20 years and much of this information (though by no means all) has been analyzed by indirect methods. Despite this extensive experience with the problems both of data collection and analysis the notion is still occasionally met that such questions are not practicable in large-scale data collection. The experience of the last two decades refutes this notion decisively and the applications to Costa Rica and Malaysia given here confirm the general conclusion. The Costa Rican example shows that indirect methods can give results quite as good as direct methods, and the example of Malaysia shows that indirect methods may give useful results even when they are substantially less than perfect.

Both the discussion of differential infant mortality by age of mother and the applications to Costa Rica and Malaysia indicate that child survivorship for 15–19 year old women is likely to be very seriously biased, a point which is generally appreciated in the field, but which seems not to have received much emphasis in print. More surprisingly, the Costa Rican application shows no evidence of deterioration in the data until well past age 50 and the Malaysian application suggests that data at ages 45–49 are quite as good as at ages 20–24. This refutes the widely held notion that response error rises substantially with age below age 50 and suggests that the common practice of ignoring information from women over age 30 or 35 may amount to throwing away half the available data. There is remarkably little published evidence to support the view that responses by women in the second half of the reproductive span are so afflicted with response error that they should be ignored for mortality estimation purposes as a matter of routine. The obvious conclusion is that one should make estimation calculations for all available data (within reasonable limits) and evaluate the results before deciding which age groups to reject. This has the advantage of accumulating evidence of response errors in the data as well as not failing to make the most of the available data.

Table 5. *Child Survivorship Data and Estimated Infant Mortality Rates: Costa Rica Census of May 14, 1973*

| Age Group | Women | Children Born | Children Surviving | Infant Mortality Rate | Time |
|---|---|---|---|---|---|
| 15–19 | 111,317 | 17,901 | 16,772 | 67.0 | 72.3 |
| 20–24 | 84,765 | 93,097 | 86,056 | 60.1 | 71.0 |
| 25–29 | 63,064 | 159,466 | 145,957 | 57.9 | 69.2 |
| 30–34 | 50,400 | 207,823 | 187,075 | 62.4 | 67.1 |
| 35–39 | 46,498 | 255,968 | 225,822 | 68.3 | 64.5 |
| 40–44 | 39,577 | 253,195 | 218,671 | 72.7 | 61.6 |
| 45–49 | 31,689 | 211,484 | 177,366 | 78.4 | 58.4 |
| 50–54 | 27,213 | 177,485 | 143,607 | 84.4 | 55.2 |
| 55–59 | 20,101 | 128,987 | 100,551 | 88.3 | 52.3 |
| 60–64 | 18,887 | 114,249 | 85,767 | 90.5 | 49.7 |
| 65–69 | 11,899 | 74,150 | 53,761 | 89.7 | 47.6 |
| 70–74 | 9,687 | 58,832 | 41,502 | 84.9 | 46.1 |

Source: Dirección General de Estadistica y Censos. *Republica de Costa Rica Censos Nacionales de 1973: Población,* Volume 1, Women and Children born, Table 25, p. 150. Children Surviving, Table 26, p.155.

Table 6. *Infant Mortality Rates from Vital Registration Data: Costa Rica 1930–1973*

| Year ($y$) | Rate for Indicated Year (per 10.000) | | | | |
|---|---|---|---|---|---|
| | $y$ | $y + 1$ | $y + 2$ | $y + 3$ | $y + 4$ |
| 1930 | 1547 | 1787 | 1491 | 1637 | 1356 |
| 1935 | 1570 | 1529 | 1417 | 1217 | 1401 |
| 1940 | 1344 | 1259 | 1613 | 1239 | 1275 |
| 1945 | 1123 | 1107 | 1081 | 933 | 997 |
| 1950 | 902 | 859 | 895 | 897 | 810 |
| 1955 | 828 | 739 | 817 | 723 | 704 |
| 1960 | 708 | 653 | 707 | 741 | 861 |
| 1965 | 718 | 628 | 603 | 597 | 671 |
| 1970 | 615 | 565 | 544 | 448 | |

Sources: United Nations *Demographic Yearbooks,* as follows:
1930–1939, *1951 Yearbook,* Table 19, Pages 228–235;
1940–1950, *1953 Yearbook,* Table 11, pp. 216–225;
1951–1964, *1966 Yearbook,* Table 14, pp. 280–295;
1965–1974, *1974 Yearbook,* Table 20, pp. 342–363.

*Notes:*

Costa Rica is coded 'U,' representing 'said to be unreliable (incomplete),' for 1930–1950 and 'C,' representing 'said to be relatively complete,' for 1951–1973. The 1966 *Yearbook* indicates that the Costa Rican figures are classified by year of registration, not by year of occurrence, before 1963.

Table 7. *Child Survivorship Data and Infant Mortality Rate Estimates: Peninsular Malaysia, Census of 24/25 September 1970*

| Age Group | Total Women | Children Born | Children Surviving | Infant Mortality Rate | Time |
|---|---|---|---|---|---|
| 15–19 | 491,615 | 56,689 | 53,525 | 60.9 | 69.7 |
| 20–24 | 377,426 | 384,650 | 364,974 | 40.9 | 68.4 |
| 25–29 | 275,710 | 745,983 | 703,487 | 38.8 | 66.6 |
| 30–34 | 268,250 | 1,141,633 | 1,060,271 | 44.2 | 64.5 |
| 35–39 | 214,824 | 1,149,231 | 1,051,810 | 48.6 | 62.0 |
| 40–44 | 186,528 | 1,079,425 | 960,107 | 58.5 | 59.1 |
| 45–49 | 157,268 | 884,325 | 960,157 | 67.9 | 55.8 |
| 50–54 | 135,543 | 708,065 | 593,434 | 70.7 | 52.7 |
| 55–59 | 105,870 | 520,452 | 431,550 | 66.5 | 49.9 |
| 60–64 | 93,116 | 422,035 | 336,601 | 71.3 | 47.4 |

Source: Unpublished tabulations, 1970 Census, courtesy of the Department of Statistics, Kuala Lumpur, Malaysia.

Table 8. *Infant Mortality Rates from Vital Registration Data: Peninsular Malaysia 1930–1940 and 1946–1972*

| Year ($y$) | Rate for Indicated Year (per 10,000) | | | | |
|---|---|---|---|---|---|
| | $y$ | $y + 1$ | $y + 2$ | $y + 3$ | $y + 4$ |
| 1930 | 1783 | 1583 | 1497 | 1575 | 1739 |
| 1935 | 1589 | 1598 | 1487 | 1494 | 1311 |
| 1940 | 1385 | | | | |
| 1945 | | 917 | 1022 | 896 | 806 |
| 1950 | 1016 | 973 | 900 | 834 | 831 |
| 1955 | 784 | 752 | 755 | 796 | 660 |
| 1960 | 684 | 597 | 594 | 568 | 484 |
| 1965 | 500 | 479 | 451 | 422 | 432 |
| 1970 | 408 | 385 | 379 | | |

Sources: United Nations *Demographic Yearbooks,* as follows:
1930–1939, *1951 Yearbook,* Table 19, pp. 328–335;
1940–1950, *1953 Yearbook,* Table 11, pp. 216–225;
1951–1964, *1966 Yearbook,* Table 4, pp. 280–295;
1965–1972, *1974 Yearbook,* Table 20, pp. 342–363.

*Notes:*

The 1951 *Yearbook* indicates '1930–1933: former British Malaya, excluding the unfederated Malay states. 1934–1940: former British Malaya.' The 1953 *Yearbook* indicates 'Prior to 1940, territory of former British Malaya, i.e., excluding Singapore.' The 1966 *Yearbook* indicates that the 1962 figure is 'provisional.' Peninsular Malaysia is coded 'C,' representing 'data stated to be relatively complete,' throughout the period.