International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

An Analysis of Indirect Mortality Estimation

Arthur, W.B. and Stoto, M.A.

IIASA Working Paper

WP-81-143

September 1981

AN ANALYSIS OF INDIRECT
MORTALITY ESTIMATION

W. Brian Arthur
Michael A. Stoto

September 1981
WP-81-143

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

THE AUTHORS

W. Brian Arthur has been at IIASA since March 1977. He was previously an Associate at the Center for Policy Studies of the Population Council, New York.

Michael A. Stoto is Assistant Professor at the J. F. Kennedy School of Government, Harvard University.

ABSTRACT

This paper investigates the robustness of the Brass child-survivorship indirect mortality estimation technique. It develops an analytical method for studying the error or bias caused in indirect mortality estimates by poor data, badly chosen model functions, and specific demographic assumptions that are often violated in practice. The resulting analytical expressions give insight into the rationale of indirect methods, the conditions under which they are robust, and the magnitude of errors that occur when specific assumptions are violated.

AN ANALYSIS OF INDIRECT MORTALITY ESTIMATION

W. Brian Arthur and
Michael A. Stoto

1. INTRODUCTION

Since the seminal work of Brass and Coale (1968), demographers have become highly skilled in the estimation of demographic parameters from indirect data. In many developing countries, the classical demographic data sources -- a registration of vital events or periodic censuses -- are far from adequate. Brass, Coale and their co-workers have developed a set of powerful and less demanding techniques based on simple survey or census questions to replace the classical methods.

All of the new methods capitalize on the substantial regularity of the age pattern of demographic events across regions and time. These methods use the minimum amount of information required to match a standard schedule to a specific situation. By careful choice, the indirect methods rely on easy-to-obtain data which are the least subject to known sources of bias.

Simulation studies, internal consistency checks, and comparisons with independent results have shown the new methods to be accurate and reasonably robust. Yet it is natural to examine the sensitivity of such methods to the many assumptions on which they are built. Just how robust are the methods? To which assumptions are they most sensitive? What would their error be in certain cases? How might they be corrected?

Questions like these have been examined before, largely through the medium of numerical studies and regression analyses. Some results are well known and the methods clearly understood. This paper adds to this literature by providing and employing an analytic technique for studying the sensitivities of the estimates to the assumptions that underlie them. This new technique provides algebraic expressions which are both more general and easier to interpret than computer results.

One purpose of this paper is to develop an analytic method for the derivation of errors in indirect data estimation. A second is to use the method to derive general and specific results. The resulting expressions give us insight into the rationale of the indirect methods, the conditions under which they are appropriate, and the possibilities for correcting the effects of inappropriate assumptions.

We illustrate the use of the analytic method in the simplest and most widely used indirect data technique -- the Bass childhood survivorship method. As we will indicate more concretely later, we expect the approach to be fruitful for other techniques as well.

Our plan is as follows. In the next section we briefly set out the notation and assumptions of the Brass childhood mortality method. Section 3 develops a general theory of errors for this estimator. We follow this, in Section 4, with four specific analyses of practical interest. A concluding section sums up the specific and general results for Brass's childhood mortality estimator.

2. THE BRASS CHILD-SURVIVORSHIP TECHNIQUE

The Brass child-survivorship technique (Brass and Coale (1968); Brass (1975)) is designed to estimate q(M), the probability of dying before age M.[1] Ideally, to estimate q(M), we would like to identify by census a large group of children at birth, follow them M years, and see how many do not survive. But in countries where census data are unreliable, this direct method is impossible: death and birth records may undercount certain social groups and be badly incomplete.

The Brass technique circumvents census-record problems by identifying the group it follows *indirectly*, as the children ever born to a representative collection of mothers who are directly questioned. For mothers of the same age x, the ratio of their children who have died to all children ever born to them, D_x, is a mortality statistic both easy to obtain and relatively reliable. The only trouble that enters is that the children "indirectly surveyed" do not conveniently all have the same age M -- they are spread over a range of ages. D_x, the proportion dead, is thus a composite of child-mortality levels. Brass's technique must provide a map from the statistic D_x to the sought-for mortality level q(M).

The technique does this in an ingenious way. In the absence of knowledge of the true ratio $q(M)/D_x$, it simulates this ratio, by calculating it in a "model" or artificially constructed population, chosen under particular assumptions to be similar to the population surveyed. With the simulated "translation ratio" k

[1] For related methods of mortality estimation, see Feeney (1980) and Preston and Palloni (1978).

at hand, the demographer need only multiply his measured D_x by k, to estimate the unknown q(M). In its simplest form, the Brass estimate for q(M) may thus be written as[1]

$$\hat{q}(M) = k.D_x \quad . \tag{2.1}$$

To examine this procedure in more detail, we need to distinguish between three different populations: the *actual population* which is the target population whose vital rates we want to estimate; the *survey population* -- children of the mothers selected for interview; and the artificial or *model population*, chosen in the simulation of the translation ratio k.

TABLE 1

q(a)	= probability of dying between birth and age *a* in the *actual population*	c(a)	= density or relative frequency of children at "age" *a* (whether alive or dead) of mothers aged x in the *actual population*
$q^s(a)$	= probability of dying before age *a* for children in the *survey population*	$c^s(a)$	= density of children at "age" *a* (whether alive or dead) in the *survey population* at the time of survey (where mothers selected have age x)
$q^*(a)$	= probability of dying before age *a* in the *model population*	$c^*(a)$	= density of children at age *a* (whether alive or dead) of mothers aged x in the *model population*

We summarize in Table 1, for each of these populations, the functions that play a key role in the technique. "Age" denotes

[1]In practice, x, the age-group of mothers questioned, is chosen so that their children are clustered around M, the estimation age. Each M therefore has a "corresponding" mothers' age-group, x, and for each of these age-groups a translation ratio k must be calculated. Often k is keyed to ancillary information. Brass (1975) provides a table indexed by the parity ratio, P_1/P_2, the number of children ever born to women aged 15 to 19 divided by the similar number for women aged 20 to 24. Sullivan (1972) provides an equation relating k to P_1/P_2. Trussell (1975) improves this equation by including P_2/P_3.

throughout this paper, years since birth whether children are living or deceased. An asterisk denotes model or guessed functions; and an S-superscript survey population functions. "True" demographic functions, the ones for the target population in question, have no superscript. A"^" will denote an estimate.

With the help of Table 1, we may write the proportion of deceased children measured by the survey as

$$D_x = \int c^S(a)\, q^S(a)\, da \tag{2.2}$$

where integration here and throughout the analysis is understood to be taken over the appropriate age range of children. We may also write the translation ratio k, the ratio of the model probability of death by age M to the model proportions deceased, as

$$k = q^*(M) / \int c^*(a)\, q^*(a)\, da \tag{2.3}$$

The Brass estimate of q(M) is the survey proportion deceased times the translation ratio. Written in terms of the survey and the model functions, it becomes

$$\hat{q}(M) = \frac{q^*(M)}{\int c^*(a)\, q^*(a)\, da} \cdot \int c^S(a)\, q^S(a)\, da \tag{2.4}$$

Note immediately a key virtue of this estimator. If there are no errors -- if the survey population perfectly represents the actual population, so that $q^S = q$ and $c^S = c$, and if the model functions have been chosen perfectly so that $q^* = q$ and $c^* = c$ -- then the estimate is exact: $\hat{q}(M) = q(M)$. Furthermore, if the choice of $q^* = \alpha q$, the α cancels in (2.4) and the estimate is still exact. Thus, the demographer need only guess the shape, not the level, of the true mortality curve.

It is clear that in general, the usefulness of (2.4) as an estimator for mortality at age M depends crucially on whether the survey can be executed with accuracy and on whether the model functions can be chosen judiciously. If women surveyed are representative of their age group in the actual population, if women's ages and children's numbers and deaths are correctly reported, and if there are no sampling errors, then the survey functions c^s and q^s correctly represent the true population functions c and q, and D_x measures the true proportion of children deceased, to all women in the population aged x. If vital rates have not changed in the years preceding the survey, if the actual mortality function is close to some member of a selected model family of mortality functions, if the true age density of children whose mothers aged x can be simulated by a model density function calculated from a standard family of model fertility functions, then c^* and q^* can be accurately chosen to simulate the true population functions c and q.[1] If all such conditions underlying the technique are fulfilled, $\hat{q}(M)$ will be an accurate estimate. If, on the other hand, women interviewed are a biased sample of the actual population at large, or if the true mortality experience in no way resembles that of the model mortality family, the estimate $\hat{q}(M)$ will be in error.

In the analysis that follows, we aim to sharpen our knowledge of the robustness of the child-survivorship technique to errors in the collection of the survey statistic D_x, to imperfect choice of model schedules, and to certain specific demographic assumptions underlying the technique that are likely to be violated in practice. We adapt methods of demographic sensitivity analysis (Arthur (1981)) to this purpose.

---

[1] It is usual not to choose c^* directly from a model family but to calculate it from an assumed model fertility schedule m^* as

$$c^*(a) = \frac{m^*(x-a)}{\int_0^x m^*(y)\,dy} \quad .$$

Thus the model age density of children aged *a* of mothers aged x is simply the proportional fertility rate *a* years ago, when mothers were aged x-a.

3. ERROR ANALYSIS

We may write (2.4), the Brass estimator of q(M), more conveniently in terms of the survey statistic D_x and the model functions q^* and c^* as

$$\hat{q}(M) = \frac{q^*(M)}{\int c^*(a)q^*(a)da} \cdot D_x \quad . \tag{3.1}$$

This will serve as our standard form of the estimate.

We have already established that if the survey statistic is correct, and the model schedules are chosen perfectly, the estimator will be correct. This fact provides the starting point for our analysis. Observe that errors can arise from only three sources: the statistic D_x may be in error; the model schedule c^*, which must be guessed, may be in error; or the model schedule q^*, which also must be guessed, may be in error. Our strategy will be to analyze errors from each source separately, using the correct estimate as a bench mark. In each case we view the source of error as a differential or small perturbation from the true observation or true vital schedule, and assume the other inputs to be correct. We then use differential calculus to derive analytical expressions for the differential--the first-order approximation to the actual change caused in the estimate $\hat{q}(M)$. The differential measures the error in $\hat{q}(M)$ due to errors in D_x or to incorrect selection of q^* and c^*. Stated another way, we view the estimate $\hat{q}(M)$ as a number that depends on three inputs, the datum D_x and the guessed functions q^* and c^*. We seek general analytical expressions for the differential in $\hat{q}(M)$ assuming each one of these inputs in turn is in error. (Exactly *how* a particular error, in c^* say, arises is not considered in this section; it is taken up in section 4.)

Since the total differential in the estimate is the sum of the differentials from each source of error, we may treat each source of error separately.

3.1. *Errors in the Survey Statistic, D_x*

Sampling errors, or systematic bias such as caused by the omission of children who have died, in general mean that the population surveyed misrepresents the actual population. Both c^s and q^s, the age density of children in the survey population and their mortality experience, may differ from c and q, the "true" density of children of mothers aged x, and "true" mortality experience in the population as a whole. This will in turn cause D_x to deviate from the "true" proportion dead in the actual population. (To say exactly how D_x deviates, would require additional assumptions about the nature of the omissions or the sampling process.) We seek an expression that links the general error or deviation δD_x in D_x with the error caused in the estimate.

We start by assuming all parts of the estimate are correct, so that

$$\hat{q}(M) = \frac{q(M)}{\int c(a)\,q(a)\,da} \cdot D_x \quad . \tag{3.2}$$

The differential $\delta\hat{q}(M)$ caused by the deviation δD_x is simply

$$\delta\hat{q}(M) = \frac{q(M)}{\int c(a)\,q(a)\,da} \cdot \delta D_x \quad .$$

(In this case the differential $\delta\hat{q}(M)$ exactly equals the error $\hat{q}(M)-q(M)$.) In proportional form, we can write

$$\text{Err } \hat{q}(M) = \frac{\delta\hat{q}(M)}{q(M)} = \frac{\delta D_x}{\int c(a)\,q(a)\,da} = \frac{\delta D_x}{D_x} \quad . \tag{3.3}$$

We have, in this case, the simple general result that the proportional error in the estimate equals the proportional error in D_x.

3.2 *Error in Choice of c^*.*

Now assume that only c^*, the model age density function, is in error, and that it deviates from the true function c by the function $\delta c = c^* - c$. Using standard operations from differential calculus we can calculate the associated differential in $q(M)$. At the starting reference point, where all parts of the estimate are correct, we can write (3.1) in quotient form as

$$\hat{q}(M) = q(M) = U/V \tag{3.4}$$

where $U = q(M)D_x$ and $V = \int c(a)q(a)da$. We may view the substitution of the guessed density c^* in V for the true density c as causing a perturbation δc in the function c; this changes V (exactly) by the differential

$$\delta V = \int \delta c(a)q(a)da \quad .$$

It causes no change in U, so that $\delta U = 0$. From the quotient rule in calculus, we can write the differential $\delta\hat{q}(M)$ as[1]

$$\delta\hat{q}(M) = \frac{V\delta U - U\delta V}{V^2} = -\frac{U}{V} \cdot \frac{\delta V}{V} \quad . \tag{3.5}$$

Therefore, dividing through by $q(M) = U/V$, the relative change or relative differential in $\hat{q}(M)$ due to the error in choosing c^*

[1]The reader will recall from elementary calculus that the differential is the linear part of the change in $f(x)$ caused by the change δx in x. In our case the change δc is itself a function, and $\delta\hat{q}$ is therefore technically called a functional (or Fréchet) differential. For details see Arthur (1981).

instead of c is

$$\text{Err } \hat{q}(M) = \frac{\delta\hat{q}(M)}{q(M)} = \frac{-\delta V}{V}$$

or

$$\text{Err } \hat{q}(M) = -\frac{\int \delta c(a)\, q(a)\, da}{\int c(a)\, q(a)\, da} \quad . \qquad (3.6)$$

We shall use this general result in our subsequent analyses.

3.3 *Error in Choice of* q^*

Now assume that c^* and D_x are correct, but that q^*, the model mortality function, deviates from the true mortality function q by the function δq. In this case the differential in the estimate, as before, can be computed from (3.4). Here

$$\delta U = \delta q(M) D_x$$

$$\delta V = \int c(a)\, \delta q(a)\, da,$$

so that

$$\delta\hat{q}(M) = \frac{\delta U}{V} - \frac{U}{V}\frac{\delta V}{V} \quad .$$

$$= \frac{\delta q(M) D_x}{\int c(a)\, q(a)\, da} - q(M) \frac{\int c(a)\, \delta q(a)\, da}{\int c(a)\, q(a)\, da}$$

Therefore, the proportional error is

$$\text{Err } \hat{q}(M) = \frac{\delta q(M)}{q(M)} - \frac{\int c(a)\,\delta q(a)\,da}{\int c(a)\,q(a)\,da} \quad . \tag{3.7}$$

Again, we shall use this general result in subsequent analyses.

Note that (3.7) confirms our earlier remark that the user need only guess the shape of the mortality curve, not the level. If the guessed mortality schedule is off by a multiplicative constant, so that $q^* = \alpha q$, then $\delta q = (\alpha - 1)q$, and the relative error is zero. In this special case the error cancels itself. This is one key advantage of the Brass technique. The user need not worry about precise choice of the correct level of mortality function in the model family. Providing all functions in the family have the same more or less "correct" shape, no appreciable error will be introduced.

3.4 *Practical Implications*

The above analytical results provide some guidance for the practicalities of using the Brass technique. Little can be said about protection against errors in the datum D_x beyond the simple observation that "representativeness" in the survey population is crucial.

Choice of the model schedule c^* (or equivalently, of the model fertility function m^* on which c^* is based) merits some comment. We see from (3.6) that the effect of an error in the choice of c^* -- in the simulation of the actual population's age density of children to mothers aged x -- is, in general, neither self-cancelling nor avoidable. There is no recourse beyond fitting c^* as correctly as possible. This is reflected in the usual practical procedure of basing the selection of c^* (or of m^*) on ancillary information that improves greatly its accuracy: the parity ratios P_1/P_2 and P_2/P_3 are often used to this purpose.

Choice of the model mortality schedule q^* is in a somewhat better position. We have already seen that what matters for the model mortality schedule is that it have the right shape. Guessing the "shape" of the unknown life-table may not be easy; but here an extra measure of protection can be afforded by a wise choice of the estimation age M. We see from the error expression (3.7) that, for some $M = \gamma$, the error would be minimised or zero. Unfortunately, however, the "unbiased estimation age" γ varies with the specific character of the error function δq. As a *very rough* guideline, we can say that if M is set not far from A, the average age of children of mothers aged x, the technique will be reasonably robust against errors in choice of q^*.[1] The reason is that D_x estimates the probability of death, very approximately, at the average survey age A. If the technique is forced to map this observation into a q(M) at age M far from A, it is forced to extrapolate along a guessed mortality function that may have the wrong shape. Error will result.

On this last point, we note in passing that the indirect mortality technique is poorly suited to the estimation of infant mortality. To estimate q at M = 1, we should, by the above advice, include only very young children (with average age about one year) in the survey, which means we should interview only very young mothers, aged 15 - 20 say. But responses of women in this age group are unreliable. Furthermore, the denominators of both (3.6) and (3.7) are small for young women, so the estimates would be especially sensitive to any errors in the model fertility and mortality schedules. The alternative, to interview older women, would raise the average age of children surveyed far above one. Estimation of q(1) would then be an "extrapolation" using a particularly poorly known part of the guessed mortality function--the infant years. In general, indirect mortality estimation performs best for ages five upwards.

---

[1] To first order, $\int c(a)q(a)da = q(A) + q'(A)\int(a-A)c(a)da + \int 0^2 da \simeq q(A)$ where A is $\int ac(a)da$, the average age of children, alive or deceased, of mothers aged x. In turn, from (3.7),

$$\text{Err } \hat{q}(M) \simeq \delta q(M)/q(M) - \delta q(A)/q(A) \quad ,$$

which is zero when M is set at A.

4. SPECIFIC ANALYSES

In this section we present four specific analyses based on the general theory of the previous section. Our goal in these analyses is to understand better the structure of the estimation technique and to explore its robustness in the face of various assumptions that are often violated or only partially fulfilled in practice.

The first and second analyses look at the effect on the estimate of fertility and mortality rates that are not stationary over time. The third example studies the effect of specific errors in the shape of the model mortality schedule. The final example uses sensitivity results to explore the tradeoffs between census and survey data.

Two particular age densities and two average values appear often in these analyses. As noted above, c(a) is the "age" distribution of all children (whether living or dead) of mothers in the population aged x. We denote the expected value of a over the distribution of c as A; it is the average age of all such children, had they survived. A second distribution,

$$c_d(a) = \frac{c(a)\,q(a)}{\int c(a)\,q(a)\,da} \qquad (4.1)$$

is the "age" distribution of deceased children of mothers aged x in the population. The expected value of a with respect to this distribution is A_d; it is the mean present "age" of the non-surviving children. Note that the mean age of non-survivors, A_d, will be greater than the mean age of all children, A, since chances of non-survival increase with age.

4.1. *Changing Fertility Rates*

A key assumption of the standard version of the childhood survivorship technique is that the fertility and mortality schedules of the target population have not changed in the recent past. But this assumption is frequently not valid, especially in developing countries where we most commonly apply the technique.

It is easy to see qualitatively how falling fertility rates would bias the estimate. If we assume mistakenly that present low fertility rates obtained in the past as well, and calculate c^*, the simulated age density of children of mothers aged x, using a model fertility schedule that underestimates past fertility, we will under-calculate the frequency of children at higher ages (when fertility was high) and over-calculate it at younger ages. Since q increases with age, as in Figure 1, the guessed model proportion dead $\int c^*(a)q(a)da$ (the denominator of the estimate) will be smaller than it should be and $\hat{q}(M)$ will over-estimate.

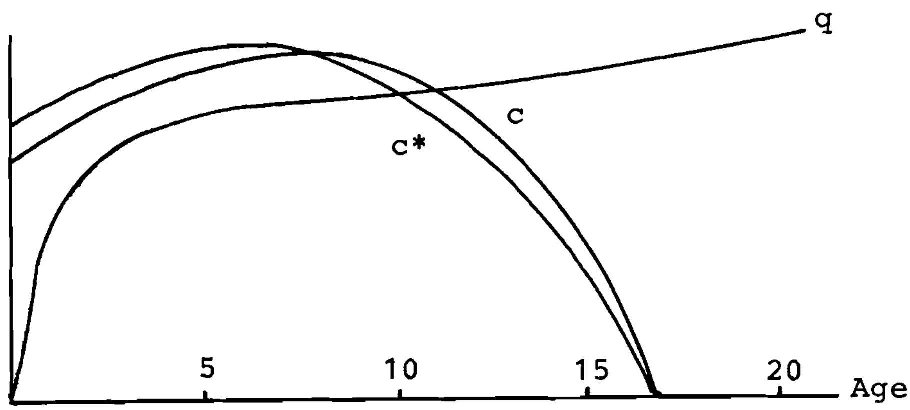

Figure 1.

To make a more precise analysis of this type of error, we must assume some specific dynamics for fertility change. Let $m(y,\tau)$ be the fertility rate for y year-old women in the population τ years before the survey. And suppose the fall in fertility is linear over time, so that

$$m(y,0) = (1 - \beta\tau)\ m(y,\tau) \quad . \tag{4.2}$$

(Since $m(y,0)$ must be positive, we assume $\beta\tau$ is less than one.) Suppose also the surveyed population is properly representative of the actual population, and that there are no measurement errors in D_x. And finally, suppose we have exact knowledge of the present true fertility rates in the actual population: we err only by assuming in our calculation of the model c^* that these rates have applied in the past. Under this mistaken assumption we calculate c^* as

$$c^*(a) = \frac{m(x-a,0)}{\int_0^x m(x-a,0)\,da} \quad . \tag{4.3}$$

Now, the true fertility schedule *a* years ago equals m(y,a), so that the *actual* age-density of children (of mothers aged x) in the population is

$$c(a) = \frac{m(x-a,a)}{\int_0^x m(x-a,a)\,da} \quad . \tag{4.4}$$

Using (4.2) we substitute $(1-\beta a)m(x-a,a)$ for $m(x-a,0)$ in (4.3) and obtain

$$c^*(a) = \frac{(1-\beta a)m(x-a,a)}{\int_0^x (1-\beta a)m(x-a,a)\,da}$$

$$= \frac{1-\beta a}{1-\beta A}\, c(a)$$

where A is the average age of children of women aged x, in the actual population. The differential δc is then

$$\delta c(a) = c^*(a) - c(a)$$

$$= \left(\frac{1-\beta a}{1-\beta A} - 1\right) c(a)$$

$$= \frac{\beta(A-a)}{1-\beta A} \quad c(a) \quad .$$

From the error expression (3.6)

$$\text{Err } \hat{q}(M) = -\frac{\beta}{1-\beta A} \quad \frac{\int (A-a)c(a)q(a)\,da}{\int c(a)q(a)\,da} \quad .$$

Noting that the expression $\int ac(a)q(a)da/\int c(a)q(a)da$ is A_d, the average age of deceased children of mothers aged x in the population, this becomes

$$\text{Err } \hat{q}(M) = \frac{\beta(A_d - A)}{1-\beta A} \quad . \tag{4.5}$$

This is the result we seek. Since A_d always exceeds A, the erroneous assumption of fertility constant at present levels does indeed cause q(M) to overestimate. The overestimation, moreover, is more than proportional to the rate of fertility decline. At younger ages, the error is usually not too serious. In a typical case[1] for 22.5 year old women, A is 2.22 years and A_d-A is 0.83 years, thus the relative error is 1.7% with $\beta = 0.02$ and 4.6% with $\beta = 0.05$. But, as we would expect, the error is more serious for older women, whose children were born when fertility differed considerably from present rates. With the same fertility and mortality schedules as before, for 42.5 year old women A is 14.0 years and A_d-A is 0.79 years, yielding relative errors of 2.2% with $\beta = 0.02$ and 13.4% with $\beta = 0.05$.

4.2. *Changing Mortality Rates*

The bias introduced by mortality rates that fall over the period before the sample has been investigated, using numerical methods, by Kraly and Norris (1978), Sullivan and Udofia (1979), and Palloni (1979,1980). Here we seek analytical expressions. Changing mortality is more difficult to analyze than changing fertility because the mortality rate we seek, q(M), itself depends on time. We must first specify the time at which we measure q(M) then analyze the error.

Let $q(a,\tau)$ be the probability that a child born τ years before the survey date dies before age a. Our target estimate is q(M,M) the probability that a child born M years ago lives to today. As in the previous example, we assume a simple model for

[1]Assumptions: (1) q(a) from Brass's European life table derivatives evaluated numerically; (2) present fertility from Brass's (1975) fertility polynomial, s = 14.5; (3) fertility declining linearly with time at rate $\beta = 0.02$ or 0.05.

the falling rates: $q(a,\tau)$ is a multiple of $q(a,0)$ and the level falls linearly with time so that

$$q(a,0) = (1-\tau\beta)q(a,\tau) \quad . \tag{4.6}$$

We further assume that the survey population is representative and correctly measured and that we guess the shape of the current mortality $q(a,0)$ correctly. Error enters because we believe mistakenly that this mortality schedule has obtained in the past, so that we select $q^*(a) = q(a,0)$.

The situation is illustrated in Figure 2, where mortality schedules of past cohorts are shown as proportionally higher than the present curve, $q(a,0)$. Each age group of children surveyed

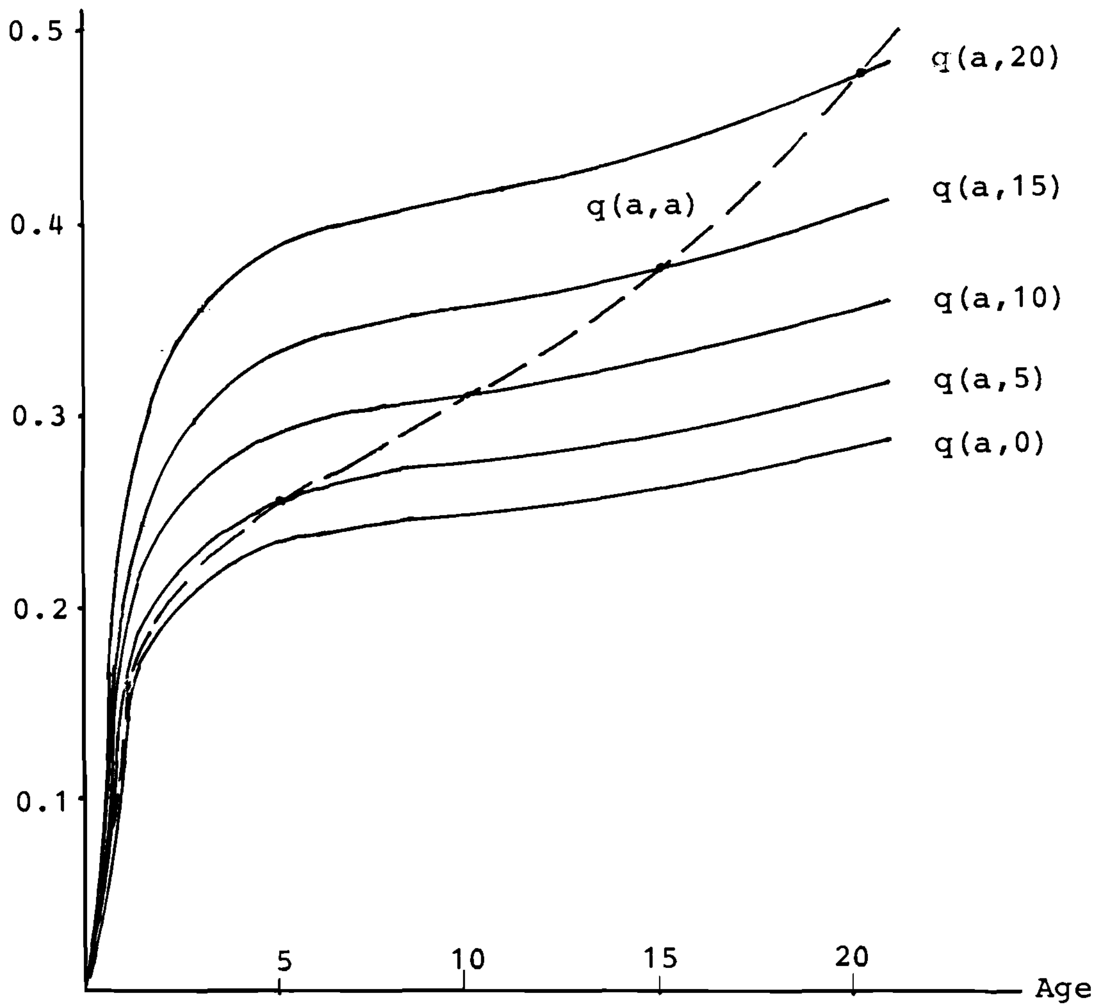

Figure 2.

will have an associated schedule, with higher mortality schedules "belonging" to children born further in the past -- children who are older. Children aged one at the time of survey have $q(1,1)$; children aged five have $q(5,5)$; and so on. Thus the true mortality schedule of children in the actual population is the composite schedule $q(a,a)$. To avoid excessive notation, we shall write this simply as $q(a)$.

Now, from (4.6) we can write $q^*(a)$, the chosen mortality schedule, as $(1-\beta a)q(a)$, so that the relative error in q^* due to believing present mortality rates have held in the past is

$$\delta q(a) = (1-\beta a)q(a) - q(a) = -\beta a\, q(a) \quad .$$

Substituting this into (3.7) yields the relative error expression:

$$\text{Err } \hat{q}(M,M) = -\beta(M - A_d) \quad . \tag{4.7}$$

The sign can be positive or negative, reflecting the fact that the mistaken q^* appears in both numerator and denominator of the estimator, and the numerator depends on the choice of M. It is not uncommon for A-A_d to be three or four years, yielding large relative errors. For instance, with Brass' European standard life table and fertility polynomial with $S = 14.5$, for women aged 32.5 the value of A_d is 7.8 years, and M is usually taken to be 5 years. With $\beta = 0.02$, this leads to a relative error of 5.6%. With $\beta = 0.05$, the relative error is 14.1%.

We would expect this error in mortality estimation to be larger still if we were to make the further mistake of believing that $\hat{q}(M,M)$ -- the mortality estimate of children born *M years ago* -- were an estimate for the mortality, $q(M,0)$, of children born *today*. We have

$$q(M,0) = q(M,M) \quad (1-\beta M) \quad .$$

We may write (4.7) as

$$\frac{\hat{q}(M,M) - q(M,M)}{q(M,M)} = \beta(A_d - M) \quad .$$

Combining these yields

$$\frac{\hat{q}(M,M) - q(M,0)}{q(M,0)} = \frac{\beta A_d}{1-\beta M} \tag{4.8}$$

As expected, q overestimates the mortality of those born today, and by an amount somewhat greater than the rate of fall of mortality times the average time elapsed since the death of the nonsurviving children in the population. To return to the example just given, a β of 0.02 now leads to a relative error of 17.4%, and a β of 0.05 now yields an error of 52.2%.

4.3. *Errors in the Assumed Mortality Pattern*

Here we analyze a case where the assumed or model pattern of mortality q^* differs from the true pattern in a specific way. Brass (1975) has found that a simple two-parameter equation adequately represents most life-tables. In particular, for any two life tables ℓ_1 and ℓ_2 observed in practice, one can find parameters α and β that relate them according to

$$\text{logit } (1-\ell_1(a)) = \alpha + \beta \text{ logit } (1-\ell_2(a)). \tag{4.9}$$

Let the true life table for the population be ℓ, with the mortality function q given by $1 - \ell$.

Suppose now we guess a mortality function q^*. By (4.9), we can represent it as

$$\text{logit } (q^*(a)) = \alpha + \beta \text{ logit } (q(a)). \tag{4.10}$$

Values of zero for α and one for β imply that the guessed function is correct. We can therefore represent errors in the choice of life table as departures of α from zero and β from one.

In the range under consideration q(a) is generally small, so that

$$\text{logit } (q(a)) \cong \frac{1}{2} \ln(q(a)) \quad ; \tag{4.11}$$

thus from this approximation and (4.10)

$$q^*(a) = e^{2\alpha} . q^{\beta}(a) \quad . \tag{4.12}$$

First, we see that non-zero values of α correspond to errors in the level of the mortality function, which we showed in Section 3 to have no effect on the relative error in the estimate. The choice of α therefore makes no direct difference. Second, values of β different from one correspond to an error in the assumed pattern of mortality. We may write

$$\delta q = e^{2\alpha} . q^{\beta} - q = e^{2\alpha}(q^{\beta-1}-1)q \quad .$$

The relative error, from (3.7), thus becomes

$$\text{Err } \hat{q}(M) = e^{2\alpha}(q^{\beta-1}(M)-1) - \frac{\int e^{2\alpha}(q^{\beta-1}(a)-1)c(a)q(a)da}{\int c(a)q(a)da} \tag{4.13}$$

$$= e^{2\alpha}\{q^{\beta-1}(M) - \int q^{\beta-1}(a)c_d(a)da\}$$

$$\text{Err } \hat{q}(M) = e^{2\alpha}\{q^{\beta-1}(M) - \text{Av}(q^{\beta-1})\} \tag{4.14}$$

where Av denotes an average taken with respect to the density c_d.

Now, whereas q increases with age, $q^{\beta-1}$ has the useful property that it remains relatively constant for β close to one in the age range over which deceased children are spread. Thus in general $q^{\beta-1}(M)$ differs little from $\text{Av}(q^{\beta-1})$ as the example in Table 2 shows.

TABLE 2[a]

β	0.8	0.9	1.0	1.1	1.2
$q^{\beta-1}(M) - \text{Av}(q^{\beta-1})$	0.0042	0.0018	0.0	-0.0015	-0.0021

[a]Note: Assumes M = 5 and c and q as given in the example in 4.1.

These errors are in all cases less than half a percent. In general we can conclude that providing the model mortality schedule is chosen from the correct Brass logit family, the mortality estimate will be robust to choice within the family (choice of α and β). It is this property that lends the logit model family its power in indirect estimation of mortality.

4.4 *Census Versus Survey Statistics*

When the Brass procedure is applied to complete census data, there is no sampling error in the observed D_x (although there may, of course, be bias errors corresponding to the exclusion of certain mothers or deceased children). Errors arise because the model schedules c^* and q^* are incorrectly guessed. With survey data, on the other hand, the D_x are observed with random variation and are therefore subject to sampling error, but we have an advantage that we can include specific questions that help in guessing c^*. A technique due to Preston and Palloni (1978), for instance, allows us to estimate c^* with some accuracy from additional survey data. In census versus survey statistics, there is therefore often a tradeoff between the accuracy of the model schedules c^* and q^* and that of the statistic D_x. The theory developed earlier and assumptions about the variance of D_x in a random sample allow us to compare the size of error involved.

We illustrate the census versus survey tradeoff, using a rather simple, stylized example. For census data, we assume that D_x is correct, but that in absence of good information on c, the model age density c^* has been calculated under a "typical", not large, error in the choice of model fertility m^*, corresponding to being off by one or two years in the Brass (1975) polynomial family. (The model fertility schedule has $s = 14.5$ or 13.5, not $s = 15.5$ as we assume for the actual population.) These assumptions correspond to parity ratios (P_2/P_3) of 0.49 or 0.54 rather than 0.44 and so are not very large errors.

For survey data, we assume that c^* is correctly selected, but that D_x is subject to sampling error. Since D_x is a proportion and is approximately equal to q(A), we can take the standard deviation of a sample of N births, as the "typical" error in D_x:

$$\delta D_x \simeq \sqrt{\frac{q(A)\,(1-q(A))}{N}} \quad . \tag{4.15}$$

Note that both estimates are equally sensitive to errors in q^*, so we ignore these in the illustration.

Table 3 lists the absolute values of the relative error in the estimate in this illustrative case for the census estimate (c^* in error) and the survey estimate at different sample sizes N.

TABLE 3.

	M		2	3	5	10	15
Census %	Err $\hat{q}(M)$	S = 14.5	6.4	2.8	1.9	1.8	1.9
		S = 13.5	10.5	5.0	3.5	3.3	3.5
Survey %	Err $\hat{q}(M)$	N = 500	9.5	8.7	8.2	7.9	7.5
		N = 2000	4.7	4.3	4.1	4.0	3.8
		N = 5000	3.0	2.8	2.6	2.5	2.4

We do not wish to conclude from this example that a survey is better or worse than a census. The user of the technique should be aware, however, that where surveys carry with them specific information not reliably gleaned from census date, the value of the additional information can often compensate for the main disadvantage in survey data, namely sampling error. This is especially true in the case of young women being surveyed, where the mortality estimates are most sensitive to fertility assumptions.

5. CONCLUSIONS

In this paper we have attempted to study the robustness of the Brass Childhood Survivorship mortality estimate to the assumptions that underlie it. To do so, we introduce a method that gives algebraic expressions for the error or bias caused by poor data, badly chosen model functions, and specific demographic assumptions that are often violated in practice. As a most general conclusion, the technique is relatively robust with regard to poor choice of the mortality schedule q^*, as long as the estimation age M is chosen not far from the average age, A, of the target children -- children of mothers aged x. Errors caused by poor choice of c^* are more difficult to protect against; additional information that improves the accuracy of c^* is the only safe recourse. And good estimates of the infant mortality rate are difficult to obtain under the Brass method. If we are willing to model exactly how certain specific demographic assumptions are violated, we can derive algebraic expressions for the bias in the estimate.

The error theory we have developed rests on an application of differential calculus. As such, our general error expressions (3.3), (3.6) and (3.7) are differentials -- first-order approximations to the true error. We would expect these approximations to be reasonably close, however, for the reason that the estimate, as in (3.1), is fairly linear in both c^* and q^*. For the specific analyses in Section 4, we made further assumptions and approximations. Although not exact, our results here should be regarded as indicative of the type of bias introduced, its magnitude, and the factors on which it depends.

Although we have not calculated numerical values for all the error expressions, they are well suited to computation. In specific situations such calculation could help provide error bounds or sensitivity analyses.

Throughout we have been concerned with robustness of the technique and the structure of biases introduced. We have had little to say about the calculation of correction factors based on additional information.

The child-survivorship technique considered in this paper is but one of a growing number of indirect estimation techniques. Hill and Trussel (1977) describe similar techniques based on data on surviving parents, spouses and siblings. Preston and Palloni (1978) introduce a method that replaces the model fertility schedule with the age distribution of surviving children. Similar analyses of these various techniques could be performed. They would provide useful information about the techniques themselves and the conditions under which one might be considered better than another.

REFERENCES

Arthur, W.B. 1981. The Analysis of Causal Linkages in Demographic Theory. Research Report. Forthcoming. International Institute for Applied Systems Analysis.

Brass, W. 1975. Methods for Estimating Fertility and Mortality from Limited and Defective Data. Chapel Hill: University of North Carolina.

Brass, W. and Coale, A.J. 1968. "Methods of Analysis and Estimation", in W. Brass (Ed.), Demography of Tropical Africa. Princeton: Princeton University Press.

Feeney, G. 1980. Estimation of Mortality Trends from Child Survivorship Data. Population Studies, 34, 109-128.

Hill, K. and Trussell, T.J. 1977. Further Developments in Indirect Mortality Estimation. Population Studies, 31, 313-334.

Kraly, E.P. and Norris, D.A. 1978. An Evaluation of Brass Mortality Estimates under Conditions of Declining Mortality. Demography 15, 549-557.

Palloni, A. 1979. A New Technique to Estimate Infant Mortality with an Application for El Salvador and Colombia. Demography 16, 455-473.

Palloni, A. 1980. Estimating Infant and Childhood Mortality Under Conditions of Changing Mortality. Population Studies, 34, 129-142.

Preston, S.H. and Palloni, A. 1978. Fine-Tuning Brass-Type Mortality Estimates with Data on Ages of Surviving Children. Population Bulletin 10. New York: United Nations Department of Economic and Social Affairs.

Sullivan, J.M. 1975. Models for the Estimation of the Probability of Dying between Birth and Exact Ages of Early Childhood. Population Studies, 29, 97-107.

Sullivan, J.M. and Udofia, G.A. 1979. In the Interpretation of Survivorship Statistics: The Case of Non-Stationary Mortality. Population Studies, 33, 365-374.

Trussell, T.J. 1975. A Re-estimation of the Multiplying Factors of the Brass Technique for Determining Childhood Survivorship Rates. Population Studies, 29, 97-107.