

AN ANALYSIS OF INDIRECT MORTALITY ESTIMATION

W.B. Arthur

Food Research Institute, Stanford University, Stanford, California 94305, USA

M.A. Stoto

*John F. Kennedy School of Government, Harvard University,
Cambridge, Massachusetts 02138, USA*

RR-84-21

November 1984

Reprinted from *Population Studies*, volume 37 (1983)

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Reprinted with permission from *Population Studies*, 37:301–314.
Copyright © 1983 Population Investigation Committee, London School of Economics.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the copyright holder.

Printed by Novographic, Vienna, Austria

PREFACE

For some years, IIASA has placed considerable emphasis on both population studies and methods of analysis and estimation. These two strands of research are combined in this paper, which develops an analytical method for the derivation of errors in estimations based on indirect data, and illustrates this method by applying it to Brass's child survivorship estimation technique.

The method described here makes it possible to study the sensitivities of estimates to the underlying assumptions, and provides algebraic expressions that are both more general and easier to interpret than computer results. These expressions give some insight into the rationale of the indirect methods, the conditions under which they are appropriate, and the possibilities of correcting the effects of inappropriate assumptions.

An Analysis of Indirect Mortality Estimation*

W. B. ARTHUR† AND M. A. STOTO‡

I. INTRODUCTION

Since the seminal work of Brass and Coale,¹ demographers have become highly skilled in the estimation of demographic parameters from indirect data. In many developing countries, the classical sources of demographic data – 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 the new methods are based on the substantial regularity of the age pattern of demographic events across regions and time. They use the minimum amount of information required to match a standard schedule to a specific situation. By careful choice, the indirect methods rely on data which are easy to obtain and are least subject to known sources of bias.

Simulation studies, checks on internal consistency, and comparisons with independent results have shown the new methods to be accurate and reasonably robust. Yet it is natural to examine their sensitivity to the many assumptions on which they are based. 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 exploratory paper is to develop an analytical method for the derivation of errors in estimations in which indirect data are used. A second is to use the method to derive both general and specific results. The resulting expressions give an 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 analytical method in the simplest and most widely used technique – Brass's childhood survivorship method. As we shall 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 Brass's

* We wish to thank Griffith Feeney, who suggested that demographic sensitivity analysis methods developed in W. B. Arthur, 'The Analysis of Causal Linkages in Demographic Theory', IIASA Research Report RR-81-27 (1981), be applied to incomplete-data estimation techniques. Helpful comments were provided by Douglas Ewbank, Kenneth Hill, and Joseph Potter. The research was carried out at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

† Visiting Professor, Food Research Institute, Stanford University, Stanford, California, 94305, USA.

‡ Assistant Professor of Public Policy, J. F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts, 02138, USA.

¹ W. Brass and A. J. Coale, 'Methods of analysis and estimation'. In W. Brass (ed.), *Demography of Tropical Africa* (Princeton University Press, Princeton, 1968), pp. 88–150.

childhood mortality method. In Section 3 we develop a general theory of errors for this estimator. We follow this, in Section 4, with four specific analyses of practical interest. In a concluding section we sum up the specific and general results for Brass's childhood mortality estimator.

2. BRASS'S CHILD-SURVIVORSHIP TECHNIQUE

Brass's child-survivorship technique² is designed to estimate $q(M)$, the probability of dying before birthday M .³ Ideally, to estimate $q(M)$, we should like to identify a large group of children at birth, follow them for M years, and see how many do not survive. But in countries where census data are unreliable, this direct method cannot be used: certain social groups may be undercounted in death and birth records which may be badly incomplete.

Brass's technique circumvents census-record problems by identifying the group indirectly, as children ever born to a representative collection of mothers who are directly questioned. The ratio of children born to mothers of the same age x who have died to all children ever born to them, D_x , is a mortality statistic which is both easy to obtain and relatively reliable. The only trouble is that the children 'indirectly surveyed' are not all conveniently of 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 required mortality level $q(M)$.

Brass accomplished this in an ingenious way. In the absence of knowledge of the true ratio $q(M)/D_x$, he 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 at hand, the observed value of D_x need only be multiplied by k to estimate the unknown $q(M)$. In its simplest form, Brass's estimate for $q(M)$ may thus be written as⁴

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

To examine this procedure in greater 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 .

We summarize in Table 1, for each of these populations, the functions that play a key role in the technique. 'Age' denotes, throughout this paper, years since birth whether children are living or deceased. An asterisk denotes model or guessed functions; and an s -subscript survey population functions. 'True' demographic functions, those for the

² W. Brass and A. J. Coale, *loc. cit.* in footnote 1, and W. Brass, *Methods for Estimating Fertility and Mortality from Limited and Defective Data* (Chapel Hill, 1975).

³ For related methods of mortality estimation see G. Feeney, 'Estimation of mortality trends from child survivorship data', *Population Studies*, **31** (1977), pp. 109–128.

⁴ 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 'corresponds' to a 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, *op. cit.* in footnote 2, 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. J. M. Sullivan, 'Models for the estimation of the probability of dying between birth and exact ages of early childhood', *Population Studies*, **26** (1972), pp. 79–97, provides an equation relating k to P_1/P_2 . T. J. Trussell, 'A re-estimation of the multiplying factors of the Brass technique for determining childhood survivorship rates', *Population Studies*, **29** (1975), pp. 97–107, improves this equation by including P_2/P_3 .

Table 1. *Definition of terms*

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

target population in question, are without a subscript. The circumflex accent denotes an estimate.

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

$$D_x = \int c_s(a) q_s(a) da, \quad (2.2)$$

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

$$k = q^*(M) / \int c^*(a) q^*(a) da. \quad (2.3)$$

Brass's estimate of $q(M)$ is the survey proportion deceased multiplied by the translation ratio. Written in terms of the survey and the model functions, it becomes

$$\hat{q}(M) = \frac{q^*(M)}{\int c^*(a) q^*(a) da} \int c_s(a) q_s(a) da. \quad (2.4)$$

Note 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 $c^* = c$ – the estimate is exact: $\hat{q}(M) = q(M)$. Furthermore, if $q^* = \alpha q$, α cancels in (2.4) and the estimate is still exact. Thus, only the shape, and not the level, of the true mortality curve needs to be guessed.

It is clear that in general the usefulness of (2.4) as an estimator of mortality at age M depends crucially on whether the survey can be carried out with accuracy and on whether the model functions can be chosen judiciously. If the women surveyed are representative of their age group in the actual population, if women's ages and numbers and deaths of children are correctly reported, and if there is no sampling error, 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 with 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 .⁵ If all these conditions hold, $\hat{q}(M)$ will be an accurate estimate. If, on the other hand, women interviewed are a biased sample of the actual population, 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 analysis⁶ to this purpose.

3. ERROR ANALYSIS

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

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

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 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 schedules q^* , which also must be guessed, may be in error. Our strategy will be to analyse errors from each source separately, using the correct estimate as a benchmark. 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 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 in 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 $\hat{q}(M)$ by assuming in turn that each of these inputs is in error. (Exactly *how* a particular error arises is not considered in this section; we consider this problem 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 is 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

⁵ It is usual not to choose c^* directly from a model family but to calculate it from an assumed model fertility schedules m^* as

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

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 $x-a$.

⁶ See W. B. Arthur, 'The Analysis of Causal Linkages in Demographic Theory', RR-81-27 (1981), International Institute for Applied Systems Analysis, Laxenburg, Austria.

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 show 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

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

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

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

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

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

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 methods from the differential calculus we can obtain the associated differential in $\hat{q}(M)$. At the starting reference point, where all parts of the estimate are correct, we can write (3.1) in quotient form as

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

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

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

It causes no change in U , so that $\delta U = 0$. From the quotient rule we can write the differential $\delta \hat{q}(M)$ as⁷

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

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^* instead of c is

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

or

$$-\frac{\int \delta c(a) q(a) da}{\int c(a) q(a) da}. \quad (3.6)$$

We shall use this general result in our subsequent analyses.

⁷ 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 W. B. Arthur, *op. cit.* in footnote 6.

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 an amount 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

$$\begin{aligned} \delta \hat{q}(M) &= \frac{\delta U}{V} - \frac{U \delta V}{V^2} \\ &= \frac{\delta \hat{q}(M) D_x}{\int c(a) q(a) da} - q(M) \frac{\int c(a) \delta q(a) da}{\int c(a) q(a) da}. \end{aligned}$$

Therefore, the proportional error is

$$\frac{\delta \hat{q}(M)}{q(M)} - \frac{\int c(a) \delta q(a) da}{\int c(a) q(a) da}. \quad (3.7)$$

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, and not its level. If the guessed mortality schedule is in error 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 out. This is one key advantage of Brass's technique. The user need not worry about the precise choice of the correct level of mortality function in the model family. Providing all functions in the family are of the same more or less 'correct' shape, no appreciable error will be introduced.

3.4. Practical implications

These results provide some guidance for the practical use of Brass's 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 its accuracy greatly: the parity ratios P_1/P_2 and P_2/P_3 are often used for this purpose.

Choice of the model mortality schedule q^* is somewhat easier. We have already seen that what matters for the model mortality schedule is that it be of the right shape. Guessing the 'shape' of the unknown life table may not be easy; but an extra measure of protection can be afforded by a wise choice of the estimation age M . We see from Equation (3.7) that, for some $M = \gamma$, the error would be minimized 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 taken as not very different from A , the average age of children of mothers aged x , the technique will be reasonably

robust against errors in choice of q^* .⁸ The reason is that D_x estimates the probability of death, very approximately, at the average survey age A . If the technique is used to map this observation into a $q(M)$ at an age M far from A , it is forced to extrapolate along a guessed mortality function that may be of 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 include only very young children (with an average age of about one year) in the survey, which means we should interview only very young mothers, aged 15–20, say. But for women in this age group sample sizes are small and responses are subject to a number of reporting biases. 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 year. Estimation of $q(1)$ would then be an ‘extrapolation’ from a particularly poorly known part of the guessed mortality function – the infant years. In general, indirect mortality estimation performs best for ages from five upwards.

4. SPECIFIC ANALYSES

In this section we present four specific analyses based on the general theory of the previous section. Our object 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.

In the first and second analysis we look at the effect on the estimate of fertility and mortality rates that are not stationary over time. In the third example we study the effect of specific errors in the shape of the model mortality schedules. In the final example sensitivity results are used to explore the trade-offs between census and survey data.

Two particular age densities and two average values often appear in these analyses. As noted above, $c(a)$ is the ‘age’ distribution of all children (whether living or dead) of mothers 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,

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

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. Mathematically, we may write

$$A = \int ac(a) da$$

and

$$A_d = \int ac_d(a) da = \frac{\int ac(a)q(a) da}{\int c(a)q(a) da}.$$

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.

⁸ To first order, $\int c(a)q(a) da = q(A) + q'(A) \int (a-A)c(a) da + \int O^2 da \cong q(A)$ where A is $\int ac(a) da$, the average age of children, alive or deceased, of mothers aged x . In turn, from (3.7),

$$\frac{\delta q(M)}{q(M)} \cong \delta q(M)/q(M) - \delta q(A)/q(A),$$

which is zero when M is set at A .

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 the technique is most commonly applied.

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 shall underestimate the frequency of children at higher ages (when fertility was high) and overestimate it at younger

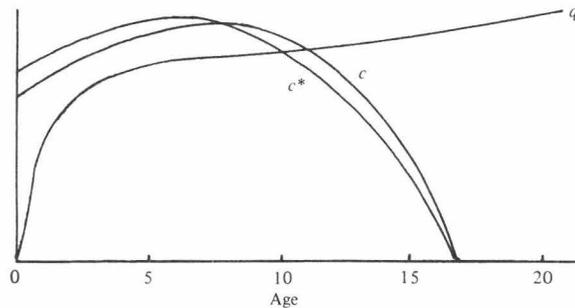


Figure 1. Actual (c) and assumed (c^*) age distributions when fertility is falling.

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 overestimate.

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

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

(Since $m(y, 0)$ must be positive, we assume $\beta\tau < 1$.) 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 know the present true fertility rates in the actual population exactly: 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

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

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

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

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

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

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

$$\begin{aligned} \delta c(a) &= c^*(a) - c(a) \\ &= \left(\frac{1 - \beta a}{1 - \beta A} - 1 \right) c(a) \\ &= \frac{\beta(A - a)}{1 - \beta A} c(a). \end{aligned}$$

From the error expression (3.6),

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

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

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

This is the result we seek. Since $A_d > A$, the erroneous assumption that fertility is constant at present level does indeed cause $\hat{q}(M)$ 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 case⁹ for women aged 22.5, $A = 2.22$ and $A_d - A = 0.83$ years, thus the relative error is 1.73 per cent with $\beta = 0.02$ and 4.64 per cent 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 women aged 42.5 $A = 14.0$ and $A_d - A = 0.79$, and now $\beta = 0.02$ yields a relative error of 2.21 per cent. To check the accuracy of (4.5), which is a first-order approximation, we can calculate the *exact* error in $\hat{q}(M)$ in the three situations we have just considered. The exact relative errors are, respectively, 1.76 per cent, 4.87 per cent and 2.30 per cent. Since the differences between these actual errors and our theoretical ones are small, we conclude that (4.5) is a good approximation.

4.2. Changing mortality rates

The bias introduced by mortality rates that fall during the period before the sample has been investigated by using numerical methods.¹⁰ Here we seek analytical expressions.

⁹ Assumptions: (1) $q(a)$ from Brass's European life table derivatives evaluated numerically; (2) present fertility from Brass's fertility polynomial with $s = 14.5$, from W. Brass, *op. cit.* in footnote 2; (3) fertility declining linearly with time at rate $\beta = 0.02$ or 0.05.

¹⁰ See E. P. Kralej and D. A. Norris, 'An evaluation of Brass mortality estimates under conditions of declining mortality', *Demography*, **15** (1978), pp. 549-557; J. M. Sullivan and G. A. Udofia, 'On the interpretation of survivorship statistics: the case of non-stationary mortality', *Population Studies*, **33** (1979), pp. 365-374; A. Palloni, 'A new technique to estimate infant mortality with an application for El Salvador and Colombia', *Demography*, **16** (1979), pp. 455-473; and A. Palloni, 'Estimating infant and childhood mortality under conditions of changing mortality', *Population Studies*, **34** (1980), pp. 129-142.

Changing mortality is more difficult to analyse than changing fertility because the mortality rate we see, $q(M)$, itself depends on time. We must first specify the time at which we measure $q(M)$, and then analyse the error.

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

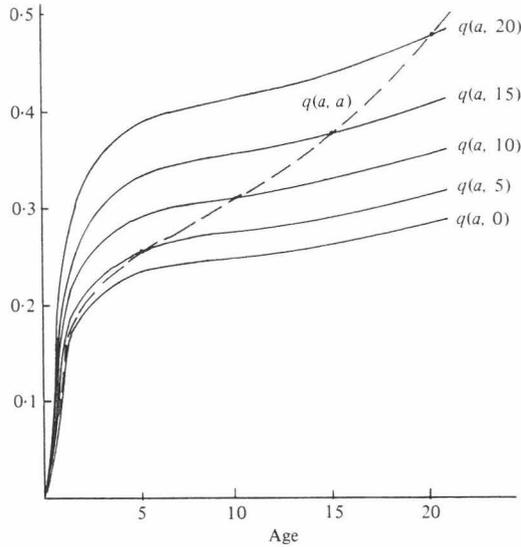


Figure 2. Actual (—) and composite (---) mortality curves when mortality is falling.

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

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

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, $q(a, 0)$. To each age group of children surveyed will correspond a mortality schedule, with higher schedules 'belonging' to children born further in the past – children who are older. That for children aged one at the time of survey will be $q(1, 1)$; that for children aged five $q(5, 5)$; and so on. Thus the mortality schedule of children in the actual population is the composite schedule $q(a, a)$. For the sake of brevity 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 obtained in the past is

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

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

$$\frac{\delta \hat{q}(M, M)}{q(M, M)} = -\beta(M - A_a). \quad (4.7)$$

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

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

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

We may write (4.7) as

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

Combining these yields

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

As expected, \hat{q} overestimates the mortality of those born to-day, and by an amount somewhat greater than the rate of fall of mortality multiplied by the average period of time elapsed since the death of the non-surviving children in the population. To return to the example just given, $\beta = 0.02$ now leads to a relative error of 17.4 per cent, and $\beta = 0.05$ now yields an error of 52.2 per cent.

4.3. Errors in the assumed mortality pattern

Here we analyse a case where the assumed or model pattern of mortality q^* differs from the true pattern in a specific way. Brass¹¹ 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, parameters α and β may be found that relate them according to

$$\text{logit}(1 - l_1(a)) = \alpha + \beta \text{logit}(1 - l_2(a)). \quad (4.9)$$

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

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

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

$\alpha = 0$ and $\beta = 1$ 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 unity.

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

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

thus from this approximation and (4.10)

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

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. Secondly, values of $\beta \neq 1$

¹¹ *Op. cit.* in footnote 2.

correspond to an error in the assumed pattern of mortality. We may write

$$\delta q = e^{2\alpha} q^\beta - q.$$

The relative error, from (3.7), thus becomes

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

$$\begin{aligned} &= e^{2\alpha} \{q^{\beta-1}(M) - \int q^{\beta-1}(a) c_a(a) da\} \\ &= e^{2\alpha} q^{\beta-1}(M) - \bar{q}^{\beta-1}, \end{aligned} \quad (4.14)$$

where $\bar{q}^{\beta-1}$ denotes a weighted average, with weights $c_a(a)$.

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

Table 2. *Relative error due to the incorrect choice of β*

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

Note: assumes $M = 5$ and c and q as given in the example in 4.1.

These errors are in all cases smaller than 0.5 per cent. In general we can conclude that provided 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 Brass's procedure is applied to complete census data, there is no sampling error in the observed D_x (although there may, of course, be errors due to bias 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 there is an advantage in that specific questions which help in guessing c^* can be included in the survey. A technique due to Preston and Palloni,¹² for instance, allows us to estimate c^* with some accuracy from additional survey data. In comparing census with survey statistics, there is, therefore, often a trade-off 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 this trade-off with a rather simple, stylized example. For census data, we assume that D_x is correct, but that in the absence of reliable 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 a difference of one to two years in Brass's

¹² S. H. Preston and A. Palloni, 'Fine-tuning Brass-type mortality estimates with data on ages of surviving children', *Population Bulletin of the United Nations*, no. 10 (1977).

polynomial family.¹³ (In the model fertility schedule $s = 14.5$ 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 :

$$\delta D_x \cong \sqrt{\left[\frac{q(A)(1-q(A))}{N} \right]}. \quad (4.15)$$

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

Table 3. *Relative errors in $\hat{q}(M)$ due to 'typical' errors in census or survey applications (per cent)*

	M	2	3	5	10	15
Census	$S = 14.5$	6.4	2.8	1.9	1.8	1.9
	$S = 13.5$	10.5	5.0	3.5	3.3	3.5
Survey	$N = 500$	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

Note: for census application we assume a shift of one or two years in the assumed fertility schedule. For survey applications we assume a shift of one standard deviation with sample sizes of 500, 2000 and 5000.

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 .

We do not wish to conclude from this example that a survey is better or worse than a census. Users of the technique should be aware, however, that where surveys include specific information not reliably obtained from census data, the value of the additional information can often compensate for the main disadvantage of survey data, namely sampling error. This is especially true when young women are surveyed, and when the mortality estimates are most sensitive to fertility assumptions.

5. CONCLUSIONS

In this paper we have attempted to study the robustness of Brass's childhood survivorship mortality estimate to the assumptions that underlie it. To do so, we introduce a method that yields 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 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 not very different 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 guard 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 from Brass's 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.

¹³ See W. Brass, *op. cit.* in footnote 2.

The 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, because 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. Note that since first-order differentials are always additive, in practical terms, it is sufficient to study each source of error independently.

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 analysis.

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 Trussell¹⁴ have described similar techniques based on data on surviving parents, spouses and siblings. Preston and Palloni¹⁵ have introduced a method that replaces the model fertility schedule with the age distribution of surviving children. Similar analyses of these different 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.

¹⁴ K. Hill and T. J. Trussell, 'Further developments in indirect mortality estimation', *Population Studies*, 31 (1977), pp. 313–334.

¹⁵ *Loc. cit.* in footnote 11.

