Demographic Techniques: Indirect Estimation

Tom A Moultrie, Centre for Actuarial Research (CARe), University of Cape Town, Rondebosch, South Africa

This article is a revision of the previous edition article by K. Hill, volume 5, pp. 3461–3464, © 2001, Elsevier Ltd.

Abstract

This article describes the principles and characteristics of techniques of indirect estimation used to derive estimates of demographic parameters from limited and/or defective data. The two broad classes of estimation techniques are described, giving examples. While the hope was that eventually these methods would be made redundant by improved data collection procedures, the suite of methods is still in widespread use half a century after they were first proposed.

'Indirect estimation' refers to the range of methods used by demographers to estimate rates of fertility, mortality (and, to a much lesser extent, migration) from data that are limited or defective, or both. The origin of the term stems from the pioneering work of the demographer William Brass in Africa in the 1960s. With the first censuses conducted after independence it was apparent, due to the observed deficiencies in the census data and the substantial incompleteness of vital registration systems of births and deaths, as well as the high levels of illiteracy and innumeracy in those populations, that direct estimation of rates of fertility and mortality, based on an analysis of the number of events occurring in a defined period of time and the exposure to risk of the population over that same period, could not produce robust and reliable estimates. Brass' important insight was that answers to simple questions, which might be better collected by enumerators and reported by respondents, could be manipulated by means of demographic models of fertility and mortality in conjunction with established mathematical relationships between demographic quantities to produce better estimates of the parameters of interest, and be robust to common patterns of error in the data.

A second category of indirect methods are those that seek to correct directly derived estimates of demographic parameters for commonly encountered errors in demographic data. Methods in this category use established relationships between demographic quantities to evaluate the consistency of the data, highlight possible sources of error, and adjust the directly observed rates accordingly.

Although these categories are rather different in their development and application, they nonetheless share the important characteristic of being based either on mathematically or logically robust relationships between different aspects of demographic data, or use model patterns of fertility, mortality, or migration to estimate or correct the parameters. To facilitate the process of demographic estimation by users who may not have extensive training in demographic estimation, parsimony of data requirements and simplicity of application by the end user are other important hallmarks of these methods.

While the methods derived by Brass and his collaborators have been developed, extended, and refined by many other demographers over the last 50 years, the approaches now in use are rooted in Brass' work, and still find use across the developing world. The material presented in this article cannot hope to provide even a summary description of all the methods that have been developed. Most of them were first written up comprehensively in *UN Manual X* (UN Population Division, 1983), the first systematic attempt to document the methods of demographic estimation from limited and defective data. Aspects of this work were revised and updated by Sloggett et al. (1994) and again (in the case of methods for estimating adult mortality) by the UN Population Division (2002). More recently, a project to revisit, revise, and update the canon of indirect methods of demographic estimation has been completed (Moultrie et al., 2013).

Indirect Estimation

The first category of methods of indirect demographic estimation seeks to produce more reliable and robust estimates of demographic parameters indirectly; that is, using information only loosely related to the parameters desired, and relying heavily on mathematical manipulation using demographic models to produce estimates of the desired parameters. These methods, all of which require very limited data as inputs, have been described by Ansley Coale – a collaborator with Brass in his work on African demography in the 1960s – as being akin to one of "making bricks with very little straw" (Coale, 2000: 55). The foundational logic of several of these methods is outlined below.

Estimation of Child Mortality

The classic, and still most widely applied, indirect method is what has come to be known as the Brass children surviving–children ever born method of estimating child mortality. The method, first set out in Brass (1964) and elaborated on in the seminal *Demography of Tropical Africa* (Brass et al., 1968), makes use of very simple summary data collected from women: how many children had they ever given birth to and how many of those were still alive, and transforms these inputs into estimates of the probability of a child dying between birth and a defined set of ages. Brass' insight was that the proportion of children dead of those ever born to women is closely associated with the age of the mother and the overall level of child mortality. Accordingly, using mathematical models of fertility and mortality, which exhibit strong regularities across a variety of demographic settings, Brass showed

Table 1 Brass' rubric for determining the ages to which the cumulative probabilities of mortality should apply, based on mother's age

Mother's age	*Best estimate of the age*, x, *to which the cumulative probability of mortality*, q(x), *applies*
15–19	1
20–24	2
25–29	3
30–34	5
35–39	10
40–44	15
45–49	20

how to convert these proportions of dead, controlling for the age of the mother, into cumulative probabilities of child mortality by age x, $q(x)$. Specifically, he showed that these proportions of dead could be mapped with some robustness (considering the poor quality or limited nature of the original data used as inputs) onto cumulative probabilities of death as shown in **Table 1**.

While Brass' approach has been refined and improved upon several times since its first exposition, the single most important contribution to the method was Feeney's (1980) observation that when mortality cannot be assumed to be constant over time, approximate reference dates could be put on the estimates of cumulative probabilities of mortality from different age groups of mother by assuming that changes in mortality are approximately linear over time.

A further refinement was to assume that the age pattern of mortality in a given setting follows a constant shape, with the level of mortality changing over time. This assumption, in conjunction with the Brass relational logit system of life tables, allows those different estimates of child mortality to be further converted into estimates of a single parameter, for example, $q(5)$, the probability of death before age 5, and for those estimates of the same parameter to be located in time using Feeney's approach, thereby allowing the identification of an approximate trend in mortality over time.

An example of the application of the method is presented in **Figure 1**, and shows the estimates of under-5 mortality in Malawi from the 1987 Malawi Census. Estimates of the same parameter from the United Nations Population Division are shown for comparison. A commonly encountered bias in the indirect method is indicated by the last data point. This estimate is derived from the reports of children ever born and children surviving from women aged 15–19 at the time of the census. The children of women who give birth at very young ages are more likely to die than children of women who give birth at older ages, resulting in the suggestion that child mortality in the country may have been increasing in the period just before the census. In most practical applications, this data point is ignored.

As Hill (2001: 3462) has noted, "[t]he Brass method and the developments of it have been very widely applied. It revolutionized knowledge of mortality conditions, particularly in sub-Saharan Africa but also in Latin America. It also gave rise to a concerted effort to find indirect estimation methods for adult mortality and migration. As such, it has proved one of the most influential ideas of demography." More recently, however, the method's applicability has been compromised by the spread of the HIV epidemic in sub-Saharan Africa. Most crucially, the epidemiology of HIV means that one of the central underlying assumptions of the method – namely the independence of the mortality of mothers and their children – is violated. Significant effort has been expended in recent years to further modify Brass's approach to accommodate the effects of HIV, largely without success. Even with the most promising approach developed so far, proposed by Ward and Zaba (2008), difficulties are encountered if the level of HIV prevalence in the population has changed over the preceding 15 or so years.

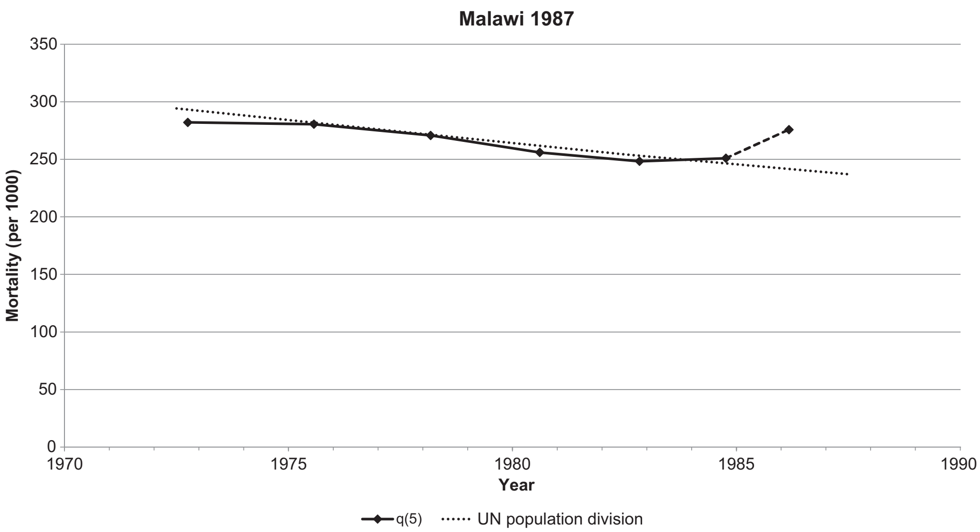

Figure 1 Estimates of under-5 child mortality for Malawi from the 1987 Malawi Census using the Brass children ever born-children surviving technique. Data from the 1987 Malawi Census held by IPUMS (ipums.org). Estimates derived using the Princeton 'North' Model Life table and the spreadsheet available at demographicestimation.iussp.org.

Estimation of Adult Mortality

Following the success of the Brass method for estimating child mortality, similar approaches were developed by Brass and his colleagues to estimate adult mortality using responses on the proportion – by respondents' age – of their mothers and fathers, spouses, and siblings who are reported to have died (respectively, the maternal and paternal orphanhood methods, the widowhood method, and the siblinghood method). The last of these has found particular use in the estimation of maternal mortality, and in the estimation of adult mortality from data collected in surveys where the sample size is typically too small to produce reliable estimates of adult mortality but, by asking about the survival of (typically in developing countries) relatively large numbers of siblings, allows mortality rates to be estimated.

As with the Brass children ever born–children surviving method, the robustness of the widowhood and orphanhood methods has been somewhat compromised by the effects of HIV/AIDS, although a simple adjustment has been proposed by Timæus and Nunn (1997) to compensate for this. In addition, problems with the application of the orphanhood method have been identified in many countries where adult death rates are high and fostering of the children of dead adults is common; in these situations, it is not uncommon for respondents to report on the vital status of their foster, rather than their biological, parents, resulting in an underestimation of mortality.

Estimation of Migration

Attempts have been made (Hill, 1987b; Zaba, 1985, 1987) to devise simple questions that can be asked in censuses that can be used to estimate migration in a similar fashion to the methods outlined above. These methods have not found widespread use, and doubts persist as to their robustness.

Estimation of Fertility

Estimates of trends in fertility can also be derived using indirect methods. The age structure of a population contains information that allow past trends in fertility to be measured. The population enumerated at a point in time of a given age, x to $x+1$, represent the survivors of births that must have occurred x to $x+1$ years previously. If the population can be assumed to be closed to migration, or if migration is negligible, an estimate of the mortality of successive cohorts of births allows one to estimate the number of births in preceding years. Furthermore, since it is possible to back-project the entire population, or that of women of reproductive age, to each preceding year using the same approach (or more simply, by assuming that the population has grown exponentially at a specified rate), it is possible to derive an estimate of either the crude birth rate or the general fertility ratio for approximately 15 years before the enumeration. (The limit of 15 years is placed to avoid the potential for in-or out-migration of young adults to bias the back-projected estimates). Despite, or perhaps because of, its simplicity, the method as outlined above has not been particularly useful, not least because it cannot produce estimates of age-specific fertility rates or of the total fertility rate.

This last limitation was addressed with the development of the 'own-child' method of fertility estimation (Cho et al., 1986). The method, further refined by Avery et al. (2013), relies on the ability to link children enumerated in a census in a household to their mothers present in the same household, and to thereby establish the mother's age at the time the child was born. Complex algorithms have to be employed to perform the matching, and those children whose mother cannot be identified in the same household have to have an appropriate mother's age imputed to them. Where children almost invariably live with their parents, the extent of the imputation required may be quite small; where fostering of children, or cohabitation in extended family groups, is widespread, the effects may be rather greater.

Recently, an alternative approach for estimation of fertility using reverse survival techniques has been proposed by Timæus and Moultrie (2013). This method has the advantage of providing the metrics desired (age-specific and total fertility rates) but avoids the issues associated with having to link children to their mothers by assuming that the shape of the age-specific fertility changes in a systematic and gradual fashion over time.

Consistency Checks

The second category of indirect techniques are those that rely on comparisons of two slightly different measures related to the same demographic parameter, and using the results to derive more robust estimates of that parameter.

Fertility

The Brass P/F ratio method, also developed by William Brass, is one such example. Brass had hypothesized (and evidence suggests that the hypothesis is generally correct) that, with census data on fertility, two specific errors frequently manifest themselves. The first relates to women's reports of lifetime fertility. Brass argued that older women's reports of lifetime fertility would tend to be biased systematically downward, particularly in very high fertility, and less literate or numerate environments, as a result of women neglecting to report dead children as such, or as a consequences of a failure to accurately estimate the exposure to risk. Conversely, he argued, younger women's reports of their lifetime fertility would be more accurate.

The second error hypothesized by Brass was that while women's recent reports – that is, collected on the basis of the number of births in the preceding year, or women's reports of fertility in a defined period of time – are frequently underreported, this underreporting is mostly independent of the age of the respondent. Consequently, while the level of fertility implied by direct estimates from the data may be too low, the shape of the fertility distribution would remain substantially correct.

Brass observed that if fertility had been constant, the cumulated period fertility up to a given age should be equal to the lifetime fertility of women of that age. Further, since the lifetime fertility of younger women is presumed to be reasonably accurate, the ratio of lifetime to cumulated period fertility at younger ages provides a basis for adjusting the observed fertility schedule. The trend in these ratios by age of respondent also provides an important diagnostic of recent trends in fertility and possible errors in the data (Brass, 1975).

While the method has been in use almost without modification since it was first proposed, a significant refinement of the model by Zaba (1981) and Booth (1984) in the form of the relational Gompertz model (described in detail by Moultrie (2013)) allows the assumption of constant fertility to be relaxed. While this model is computationally rather different from the Brass P/F method, it is based on the same underlying logic of comparing lifetime and recent fertility to adjust the observed fertility schedules.

A number of other methods based on the P/F approach, allowing for estimation of fertility from two sets of census data have been derived (UN Population Division, 1983; Zlotnik and Hill, 1981). While these methods are highly sensitive to differential errors between censuses in recording lifetime fertility, an advantage is that they allow the assumption of constant fertility to be relaxed. These have all been recast as variants of the relational Gompertz method in the updated manual of indirect techniques, *Tools for Demographic Estimation* (Moultrie et al., 2013).

Adult Mortality

Indirect methods based on consistency checks have played an increasingly important role in the estimation of adult mortality. While the denominators used to estimate adult mortality rates can be sourced from census data, the numerator (that is, the number of deaths by age) comes from either a vital registration system, or – quite commonly – answers to questions in the census about the number of deaths that occurred in each household over a defined period before the census. Both sources of the numerator are prone to underreport the number of deaths, either because the vital registration system is incomplete or because not all adult deaths occurring in a household in the defined period are reported accurately.

Two similar, but nonetheless distinct, methods have been developed to correct the recorded or reported numbers of deaths for incompleteness. The first is the generalization of the Growth Balance method, first described by Brass (1975). Brass' original formulation assumed that the growth rate in the population had been stable for an extended period of time and that migration was negligible. Under these conditions, the rate at which people enter any open-ended age interval ($a+$) is equal to the sum of the (stable) population growth rate and the death rate of those in that age interval. However, this death rate is usually understated as a result of not all deaths being accurately recorded or reported. Brass' insight here was that if the completeness of death recording or reporting was constant with respect to age, the observed death rate in the open-ended interval $d^{o}(a+)$ could be expressed as the product of the actual death rate in the interval, $d(a+)$ and the proportion of deaths, c, that were correctly recorded or reported.

Hence, mathematically,

$$b(a+) = r + (1/c)d^{\circ}(a+)$$

where, in addition to the terms defined above, $b(a+)$ is the rate at which people enter the open interval, $a+$, which can be derived from the stable population age structure alone, and r is the stable population growth rate. In this formulation, then, there is a linear relationship between b and d, with the intercept being given by r, and the slope by the reciprocal of the completeness of death registration, c. With this estimate of c, the true estimate of the death rate in the open interval can be derived, and hence the reported or recorded deaths can be adjusted for the extent of underreporting.

In a major advance, Hill (1987a) extended Brass' idea and showed how it could be generalized to populations with unstable, or variable, growth rates using data from two censuses. Hill's generalization, termed the General Growth Balance method, further allows for the possibly differential completeness of the two censuses to be taken into account in estimating the level of completeness.

The second approach, developed by Bennett and Horiuchi (1981, 1984) and termed the 'Synthetic Extinct Generations' method, relies on the observation that, in a population closed to migration, the number of people alive in the population aged a and over must – since everyone must die – equal the number of deaths that will occur in the future of people aged a and over. In a stationary population (that is, one that has a constant birth rate equal to a constant death rate, and hence with a growth rate of zero), this number of future deaths will also be the number of annual deaths of those aged a and over. Bennett and Horiuchi further showed that this reasoning and their method could be generalized to nonstable populations where the growth rate in the population is not zero, using data from two censuses.

An example of the application of this method is presented by Dorrington (2013) using data from the 2001 Census and the 2007 Community Survey in South Africa. The results are shown in **Figure 2**. Although the model life table used to smooth the results allows for the effects of AIDS on mortality, the fit is not particularly good, but nonetheless is much better than those produced when using a life table that does not incorporate the effects of AIDS.

There is some debate as to which of these two consistency-based approaches to estimating adult mortality is preferable. Hill et al. (2009) suggest, based on the results of a series of simulations, that combining the two methods, using the General Growth Balance method to estimate the change in completeness of coverage of the data from the two censuses, and then using this as an input in the Synthetic Extinct Generations method produces more robust results than either method applied separately. However, Dorrington and Timæus (2008) observed that this was not the case in populations severely affected by AIDS.

Future Directions

Indirect estimation methods were not intended as a long-term replacement for direct estimation of demographic rates from census and vital registration data; the methods were derived as

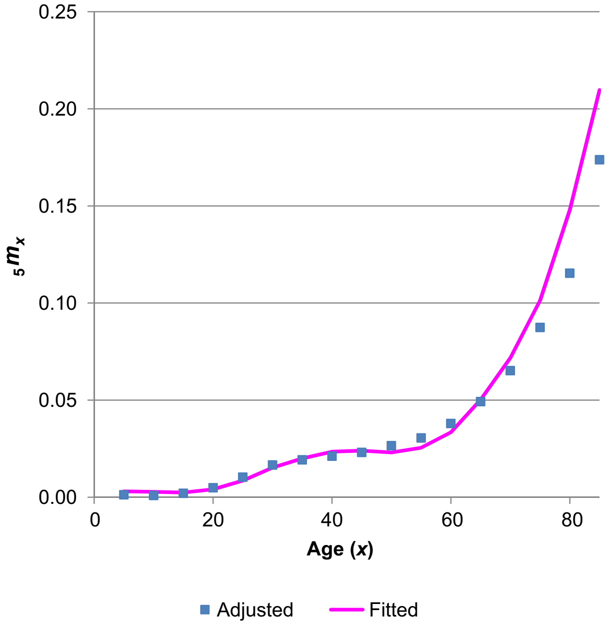

Figure 2 Estimation of male adult mortality in South Africa, applying the synthetic extinct generations method to data from the 2001 South African census and the 2007 Community Survey. Source: Dorrington, R.E., 2013. The synthetic extinct generations method. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris. http://demographicestimation.iussp.org/content/synthetic-extinct-generations-methods. Material on this site is licensed under a CC BY-NC-SA license.

a stopgap way of estimating rates from limited and defective data, with the hope that, in time, improvements in census data collection and improving completeness of vital registration systems would render the methods obsolete. Unfortunately, progress in this regard in developing countries, and especially in sub-Saharan Africa, has been slow. Consequently, the canon of indirect methods still plays an important role in demographic estimation where the quality of data remains poor, or – increasingly – in situations such as in the demographic estimation of parameters for small areas where the data are too sparse to permit direct calculation. With much greater quantities of data now available, even from developing countries, work still continues on validating, refining, and improving the suite of indirect methods using simulation techniques as well as comparison with gold standard data. A recent example is the work of Masquelier (2013) in his investigations of the nature of biases in the estimates of adult mortality produced by the siblinghood methods.

See also: Demographic Measurement: General Issues and Measures of Fertility; Demographic Measurement: Nuptiality, Mortality, Migration, and Growth; Demographic Models; Demographic Techniques: Data Adjustment and Correction; Demographic Techniques: Inverse Projection; Demographic Techniques: Small-area Estimates and Projections; Microsimulation in Demographic Research; Population Dynamics: Theory of Nonstable Populations.

Bibliography

Avery, C., St. Clair, T., Levin, M., Hill, K., 2013. The 'own children' fertility estimation procedure: a reappraisal. Population Studies 67 (2), 171–183. http://dx.doi.org/10.1080/00324728.2013.769616.

Bennett, N.G., Horiuchi, S., 1981. Estimating the completeness of death registration in a closed population. Population Index 47 (2), 207–221. http://dx.doi.org/10.2307/2736447.

Bennett, N.G., Horiuchi, S., 1984. Mortality estimation from registered deaths in less developed countries. Demography 21 (2), 217–233. http://dx.doi.org/10.2307/2061041.

Booth, H., 1984. Transforming Gompertz' function for fertility analysis: the development of a standard for the relational Gompertz function. Population Studies 38 (3), 495–506. http://dx.doi.org/10.2307/2174137.

Brass, W., December 1964. Uses of census or survey data for the estimation of vital rates. In: Paper Prepared for the African Seminar on Vital Statistics, Addis Ababa 14–19 1964. Document No. E/CN.14/CAS.4/V57. United Nations, New York.

Brass, W., 1975. Methods for Estimating Fertility and Mortality from Limited and Defective Data. Carolina Population Centre, Chapel Hill, NC.

Brass, W., Coale, A.J., Demeny, P., Heisel, D.F., et al. (Eds.), 1968. The Demography of Tropical Africa. Princeton University Press, Princeton, NJ.

Cho, L.J., Retherford, R.D., Cho, M.K., 1986. The Own-Child Method of Fertility Estimation. University of Hawaii Press, Honolulu.

Coale, A.J., 2000. Ansley J. Coale: An Autobiography. American Philosophical Society, Darby, PA.

Dorrington, R.E., 2013. The synthetic extinct generations method. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris. http://demographicestimation.iussp.org/content/general-growth-balance.

Dorrington, R.E., Timæus, I.M., 2008. Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors, revisited. In: Paper Presented at Annual Meeting of the Population Association of America. Louisiana, New Orleans, 17–19 April 2008.

Feeney, G., 1980. Estimating infant mortality trends from child survivorship data. Population Studies 34 (1), 109–128. http://dx.doi.org/10.1080/00324728.1980.10412839.

Hill, K., 1987a. Estimating census and death registration completeness. Asian and Pacific Census Forum 1 (3), 8–13, 23–24. http://hdl.handle.net/10125/3602.

Hill, K., 1987b. New approaches to the estimation of migration flows from census and administrative data sources. International Migration Review 21 (4), 1279–1303. http://dx.doi.org/10.2307/2546515.

Hill, K., 2001. Demographic techniques: indirect estimation. In: Smelser, N.J., Baltes, P.B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Pergamon, Oxford, pp. 3461–3464. http://www.sciencedirect.com/science/article/pii/B0080430767020945.

Hill, K., You, D., Choi, Y., 2009. Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data error. Demographic Research 21 (Article 9), 235–254. http://dx.doi.org/10.4054/DemRes.2009.21.9.

Masquelier, B., 2013. Adult mortality from sibling survival data: a reappraisal of selection biases. Demography 50 (1), 207–228. http://dx.doi.org/10.1007/s13524-012-0149-1 (Online: 2 October 2012).

Moultrie, T.A., 2013. The relational Gompertz model. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population,, Paris. http://demographicestimation.iussp.org/content/relational-gompertz-model.

Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), 2013. Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris. Available at: demographicestimation.iussp.org.

Sloggett, A., Brass, W., Eldridge, S.M., Timæus, I.M., Ward, P., Zaba, B. (Eds.), 1994. Estimation of Demographic Parameters from Census Data. Statistical Institute for Asia and the Pacific, Tokyo.

Timæus, I.M., Moultrie, T.A., 2013. Estimation of fertility by reverse survival. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris. http://demographicestimation.iussp.org/content/reverse-survival-methods.

Timæus, I.M., Nunn, A.J., 1997. Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method. Health Transition Review 7 (Suppl. 2), 23–43. http://www.jstor.org/stable/pdfplus/40652324.pdf.

UN Population Division, 1983. Manual X: Indirect Techniques for Demographic Estimation. United Nations, Department of Economic and Social Affairs, New York. ST/ESA/SER.A/81. http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html.

UN Population Division, 2002. Methods for Estimating Adult Mortality. United Nations, Department of Economic and Social Affairs, New York. ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html.

Ward, P., Zaba, B., 2008. The effect of HIV on the estimation of child mortality using the children surviving/children ever born technique. Southern African Journal of Demography 11 (1), 39–73.

Zaba, B., 1981. Use of the Relational Gompertz Model in Analysing Fertility Data Collected in Retrospective Surveys. Centre for Population Studies Research Paper 81–82. Centre for Population Studies, London School of Hygiene & Tropical Medicine, London.

Zaba, B., 1985. Measurement of Emigration Using Indirect Techniques: Manual for the Collection and Analysis of Data on Residence of Relatives. Liège, Belgium. Ordina Editions.

Zaba, B., 1987. The indirect estimation of migration: a critical review. International Migration Review 21 (4), 1395–1445. http://dx.doi.org/10.2307/2546519.

Zlotnik, H., Hill, K., 1981. The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality. Demography 18 (1), 103–122. http://dx.doi.org/10.2307/2061052.

Relevant Website

demographicestimation.iussp.org – The methods described here are all written up, together with spreadsheets that implement the methods, in the recently published *Tools for Demographic Estimation*.