Statistical release
P0302

Mid-year estimates

2002

Embargo: 13:00
Date: 1 July 2002

Published by
Statistics South Africa
Private Bag X44
Pretoria
0001

CONTENTS

A complete set of Stats SA publications is available in the Stats SA Library, and in the following public libraries:

KEY FINDINGS

BY MID-2002 THE POPULATION OF SOUTH AFRICA IS ESTIMATED TO BE 45,45 MILLION (WITHOUT TAKING INTO ACCOUNT ADDITIONAL DEATHS DUE TO HIV/AIDS) AND 45,17 MILLION WHEN SUCH ADDITIONAL DEATHS ARE CONSIDERED. THESE ESTIMATES HAVE BEEN ARRIVED AT USING THE 1996 CENSUS FIGURES AS THE BASE POPULATION AND MAKING CERTAIN ASSUMPTIONS IN THE ESTIMATION OF FERTILITY AND MORTALITY.

Summary of methods used in preparing population estimates for mid-2002

The methods underlying the preparation of the mid-year estimates for 2000 and 2001 have been maintained in the preparation of this year's mid-year estimates. In keeping with the practice of maximising available information, the mid-year estimates that take into account additional deaths due to HIV/AIDS, have been revised in the light of the following additional information:

a) The Department of Health's sero-prevalence survey data for 2001
b) Causes of death data from 1990 to 1996
c) Deaths data (without breakdown by causes of death) from the Population Register from 1996 to 2000.

Summary of population estimates for mid-2002

The report presents results at the national level and at 16 levels of disaggregation, namely: urban and non-urban areas; five population groups (including 'other/unspecified'); and nine provinces. For each of these levels of dissaggregation, the results are shown separately for males, females and total. The results of the mid-year estimates are shown in Table 1.1 through 1.3. By mid-2002 the population of South Africa is estimated to be 45,45 million (without taking into account additional deaths due to HIV/AIDS) and 45,17 million when such additional deaths are considered. Therefore, the implied additional deaths due to HIV/AIDS is 282 303. These estimates have been arrived at using the 1996 census figures as the base population, and making certain assumptions in the estimation of fertility and mortality. The growth rates that underlie these estimates are summarised in Table E. These growth rates are 'inferred' growth rates, which differ from actual growth rates obtained from two consecutive censuses. Details on the methods used and the assumptions made are given in the explanatory and technical notes.

pp
P J Lehohla
Statistician-General: Statistics South Africa

TABLES

1. Mid-year population estimates for 2002, without additional deaths due to HIV/AIDS
1.1 South Africa, urban and non-urban populations

		Population estimates for mid-2002
South Africa	Male	21 870 730
	Female	23 583 480
	Total	45 454 211
Urban	Male	11 612 775
	Female	12 275 503
	Total	23 888 278
Non-urban	Male	10 257 955
	Female	11 307 978
	Total	21 565 933

1. Mid-year population estimates for 2002, without additional deaths due to HIV/AIDS
1.2 South Africa and population groups

		Population estimates for mid-2002
South Africa	Male	21 870 730
	Female	23 583 480
	Total	45 454 211
African	Male	17 019 919
	Female	18 454 273
	Total	35 474 192
Coloured	Male	1 902 285
	Female	2 015 763
	Total	3 918 048
Indian/Asian	Male	548 655
	Female	573 159
	Total	1 121 815
White	Male	2 212 492
	Female	2 342 797
	Total	4 555 289
Other and unspecified	Male	187 379
	Female	197 488
	Total	384 867

2. Mid-year population estimates for 2002, with and without additional deaths due to HIV/AIDS

2.1 South Africa and provinces

		Population estimate taking into account additional deaths due to HIV/AIDS	Population estimate without taking into account additional deaths due to HIV/AIDS	Implied additional deaths due to HIV/AIDS
South Africa	Male	21 748 567	21 870 730	122 164
	Female	23 423 341	23 583 480	160 139
	Total	45 171 908	45 454 211	282 303
Western Cape	Male	2 099 675	2 103 106	3 432
	Female	2 214 284	2 218 737	4 453
	Total	4 313 959	4 321 844	7 885
Eastern Cape	Male	3 321 970	3 333 644	11 674
	Female	3 810 171	3 825 199	15 027
	Total	7 132 141	7 158 843	26 702
Northern Cape	Male	433 360	434 439	1 079
	Female	455 030	456 425	1 395
	Total	888 390	890 864	2 474
Free State	Male	1 406 585	1 415 142	8 557
	Female	1 452 496	1 463 851	11 355
	Total	2 859 081	2 878 993	19 913
KwaZulu-Natal	Male	4 328 707	4 370 197	41 490
	Female	4 883 415	4 938 368	54 953
	Total	9 212 123	9 308 565	96 442
North West	Male	1 796 770	1 808 029	11 258
	Female	1 863 132	1 878 024	14 892
	Total	3 659 902	3 686 053	26 151
Gauteng	Male	4 104 604	4 132 541	27 938
	Female	4 001 586	4 037 844	36 258
	Total	8 106 190	8 170 386	64 196
Mpumalanga	Male	1 535 356	1 546 061	10 705
	Female	1 620 916	1 634 980	14 064
	Total	3 156 272	3 181 041	24 769
Limpopo	Male	2 721 540	2 727 570	6 031
	Female	3 122 311	3 130 052	7 741
	Total	5 843 851	5 857 622	13 772

Explanatory and technical notes

In preparing the mid-year population estimates, a number of data sources and methods are used. As data become available, they are incorporated in the exercise either as independent controls or as input data and the methods used are continuously refined and improved as necessary. Further details on the data and methods used are given below.

1. Base population

For the purpose of computing central death rates, it was necessary to use the mid-1996 population. In order to bring the population to mid-year ('younging'), the total population had to be multiplied by a factor. An estimate of the mid-year population was obtained by using a modified form of the balancing equation. This is the balancing equation with adjustments for undercount on the terms dealing with deaths, births and migration. The ratio of this estimate to the 1996 census population gives an adjustment factor used to 'young' the population.

For the projections, the base population used was the actual 1996 census data adjusted for undercount by the post-enumeration survey. The reported ages of the population were used as they were, without any adjustments made to them. A small proportion of the population had unstated ages. This proportion was distributed among the population of known ages through a standard procedure explained below. Projections were done on the population with known ages and an adjustment factor for pro-rating the population of unknown ages, *k,* was calculated and applied to the projected population. The adjustment factor, *k*, is calculated as follows:

$$k = P_T/(P_T - P_u) \qquad (1)$$

where

P_T is the total population over all ages (including those with unknown ages), and
P_u is the population of unknown ages.

The adjustment factors obtained are given in Table A.

2. Estimation of population aged 0-4

From the census data, reported fertility (births in the past twelve months) per woman and parity (average number of children ever born per woman) were obtained. However, due to the misunderstanding of the reference period, reported fertility often underestimates true fertility. There is therefore the need to adjust the reported fertility estimates. The method opted for is that of Arriaga (1983). In that method, the reported parities are transformed into age-specific fertility rates and are subsequently cumulated. These cumulated rates are compared with another set of cumulated rates obtained from the reported fertility to obtain adjustment factors. Adjustment factors could be based on age groups 20-24, 25-29 or 20-29. Adjustment factors based on women in the age group of 20-29 years have been used, as they usually give estimates that lie between those obtained from the 20-24 age group and the 25-29 age group. The program used for doing this is FERTPF in the United Nations Software Package for Mortality Measurement (MORTPAK). Note that there are other methods for indirectly estimating fertility which would not necessarily give the same results. With direct techniques, it is possible to arrive at a fixed value, but with indirect methods, the values and variances vary slightly within an acceptable range.

The adjusted age-specific fertility rates are used to obtain an estimate of the average annual number of births. This is multiplied by five to obtain average number of births for five years. This number is separated into males and females using an assumed sex ratio at birth (1,03). For each pair of male and female life tables, the respective number of five-year births is multiplied by the five-year survivorship ratio from birth to obtain an estimate of the population aged 0-4.

Table A. Adjustment factors for unknown ages in the 1996 census data

Broad category	Unit	Gender	Adjustment factor for unknown ages, k
Total	South Africa	Male	1,013249
		Female	1,011280
Urban/non-urban	Urban	Male	1,013399
		Female	1,010921
	Non-urban	Male	1,013069
		Female	1,011682
Population group	African	Male	1,012713
		Female	1,011033
	Coloured	Male	1,008511
		Female	1,007423
	Indian/Asian	Male	1,009092
		Female	1,008806
	White	Male	1,016343
		Female	1,013680
	Other and unspecified	Male	1,081407
		Female	1,049809
Province	Western Cape	Male	1,012425
		Female	1,008996
	Eastern Cape	Male	1,009130
		Female	1,007416
	Northern Cape	Male	1,012227
		Female	1,009987
	Free State	Male	1,011974
		Female	1,010441
	KwaZulu-Natal	Male	1,015138
		Female	1,013524
	North West	Male	1,009349
		Female	1,007395
	Gauteng	Male	1,014680
		Female	1,012884
	Mpumalanga	Male	1,019284
		Female	1,016576
	Limpopo	Male	1,013774
		Female	1,012319

Note: The categories used in the first column are non-overlapping since the provincial totals add up to the national total, the population group totals add up to the national total and the total for the urban and non-urban locations also add up to the national total.

3. Incorporation of internal migration into the projections

3.1 Background

When projections for all the regions of a country are required and the appropriate data are available, a multiregional approach should be considered (United Nations, 1992). The method combines the cohort-component perspective of demographic projection with matrices of inter-area migration rates (Rees, 1994). In this approach, regional population projections are generated simultaneously for a system of several interacting regional populations and it guarantees that the total migration flows between regions will sum to zero (or to the assumed level of international migration). The required mathematics for doing so (known variously as multistate demography, multiregional mathematical demography and multidimensional mathematical demography) has been developed through the works of Rogers (1973 and 1975) among others. In the case of migration, the data come in different forms. Research in multiregional demography has seen the emergence of different techniques to handle different data types. First is the distinction between migration data recorded as relocation events (as in the case of migration information derived from a Population Register) and migration data derived by a comparison of place of residence in two points in time (as in the case of retrospective questions asked in censuses and surveys). The former is known as the 'movement approach' or 'movement perspective' and the latter as 'transition approach' or 'transition perspective' (Rees, 1997; Ledent, 1980). Unlike births and deaths, each type of migration data has its own peculiar twist. For migration data that come in the form of 'place of previous residence', the application of multiregional demography has been outlined by Philipov and Rogers (1981). For migration data that comes in the form of 'lifetime migration', the application of multiregional demography has been outlined by Ledent (1981) while for 'last-move' migration data, the topic has been covered by Schmertmann (1999).

The prerequisite for a full-scale multiregional population projection is the multiregional life table model. Developments in multistate demography have seen the vector generalisation of ordinary life table functions to suit the multiregional context. In particular, works on multiregional life tables from within the transition perspective have been investigated by Rogers (1975), Willekens and Rogers (1978), Ledent and Rees (1980), and Ledent (1982) among others. The estimation of the full multiregional life table requires age-specific flows among the various states considered in the analysis. For the internal migration component, all elements of the 'migration cube' (age, origin and destination) need to be provided (Rees and Kupiszewski, 1999). For the mortality component, age-specific mortality rates have to relate not only to the state of residence at age x but also to the state in which the deaths actually occur (Ledent, 1982). Naturally, the data requirements for the application of the full multiregional population projection model are quite demanding and in many cases the data are not available. To overcome these data shortcomings, the estimation of missing data has been addressed by Willekens (1982). Another work has investigated the extent to which the full internal migration matrix can be simplified without seriously affecting the performance of the resulting multiregional model (van Imhoff *et al*., 1997). Lastly, another work has developed model multistate migration schedules for use in situations where the available migration data are inadequate or inaccurate (Rogers and Castro, 1982).

3.2 Application of multiregional demography to relatively 'data-poor' countries

Attempts at applying multiregional demography to relatively 'data-poor' countries have been undertaken by only a few researchers. An early attempt was by Doeve (1984) in which he used Thailand's data. In that study, Doeve made the following remarks:

> The operationalization of something as sophisticated as multistate demography in developing countries is not necessarily a 'dirty job'. (Doeve, 1984)

Some aspects of this operationalisation have been addressed by Bah (1990). In that paper, the following remarks were made about Doeve's work:

> What could … be learnt from Doeve's exercise is the lesson that data which suffers from underenumeration and omissions should not be dismissed as being unfit for the purpose of construction of multiregional life tables. Or to phrase this in other words, we cannot say that because African data is beset by problems of underenumeration [,] omission and misreporting, we should not attempt [to use them] to construct multiregional life tables. (Bah, 1990:15)

In another attempt, the United Nations (1992) applied a modified multiregional projeciton model to Indonesian data. This method has been adopted in this report. Details are given in the next section.

3.3 United Nations' modified multiregional migration model

In attempting to reduce the data needs for doing a full multiregional population projection model, the United Nations (1992) developed a spreadsheet program based on a simplification of the multiregional approach of Rogers (1975, 1985). From the methodological steps outlined in the example, one can mathematically describe the simplification procedure as follows:

In a single region (with no migration considered), one has that,

$$ {}_5P_{x+n,t+n} = {}_5P_{x,t} * \frac{{}_5L_{x+n}}{{}_5L_x} \tag{2} $$

where

${}_5P_x$ is the population aged x to $x+4$ at time t
${}_5P_{x+n}$ is the population aged $x+n$ to $x+n+4$ at time $t+n$
and
${}_5L_{x+n}/{}_5L_x$ is the life table survivorship ratio from age $x+4$ to age $x+n+4$

With the inclusion of migration, the survivors in region j (5 years later), out of the original population in region i, is given by:

$$ {}_5P^{ij}_{x+n,t+n} = {}_5P^{i}_{x,t} * \frac{{}_5L^{i}_{x+n}}{{}_5L^{i}_x} * \left[{}_5M^{ij}_{x+n,t+n} \right] \tag{3} $$

The non-migrant population in region i is given by:

$$ {}_5P^{ii}_{x+n,t+n} = {}_5P^{i}_{x,t} * \frac{{}_5L^{i}_{x+n}}{{}_5L^{i}_x} * \left[1 - \sum_{j \neq i} {}_5M^{ij}_{x+n,t+n} \right] \tag{4} $$

where ${}_5M^{ij}_{x+n,t+n}$ is a conditional probability (also known as conditional migration proportion). It denotes the probability of being in region j at time $t+n$, conditional on survival to time $t+n$. It is loosely called 'migration rate' but is different from the usual age-specific mortality and fertility rates. It expresses the number of migrants as a fraction of the population of region i, aged x to $x+n$ at the beginning of the period.

Mathematically, ${}_5M^{ij}_{x+n,t+n}$ is defined as:

$$ {}_5M^{ij}_{x+n,t+n} = \frac{{}_5O^{ij}_{x+n,t+n}}{\left[{}_5P^{i}_{x+n,t+n} - \left[{}_5O^{\cdot i}_{x+n,t+n} - {}_5O^{i\cdot}_{x+n,t+n} \right] \right]} \tag{5} $$

where

${}_5O^{ij}_{x+n,\ t+n}$ is the age–specific migration stream from province i to province j, over the previous five-year period. This refers to persons whose last moves occurred within the specified interval, originating in i and terminating in j,

${}_5O^{\cdot i}_{x+n\ ,t+n}$ is the sum of all age–specific migration streams into province i, over the previous five-year period, and
${}_5O^{i\cdot}_{x+n\ ,t+n}$ is the sum of all age–specific migration streams out of province i, over the previous five-year period.

This formula is similar to the version suggested in United Nations (1970). However, further clarification on the formula was offered by van Imhoff (2000) and Ledent (2000 and 2001). In this expression, the denominator is composed of persons who resided in i throughout the entire interval plus all persons whose last move (made after time t) originated in i, regardless of where they resided between time t and time of last move.

The formula (3) is similar to the 'Option 2' method proposed by Rogers (1975) that considers observed proportions surviving (Willekens, 2001). The formula also compares closely with the one proposed by Ledent (1982) as given below:

$$s_x^{ij} = \frac{1-(n/2)m_x^{i\delta}}{1+(n/2)\sum_k \overline{s_x^{ik}}\, m_{x+n}^{k\delta}}\,\overline{s_x^{ij}} \qquad (6)$$

where:

s_x^{ij} is the proportion of those aged x to $x+n$ in region i who survive n years later in region j,

$\overline{s_x^{ij}}$ is the conditional proportion (that is, without mortality) of those aged x to $x+n$ in region i who survive n years later in region j and

$m_x^{i\delta}$ is the mortality rate of those aged x to $x+n$ in region i.

First, note that formula (2) (with a slight change in the symbols to suit Ledent's formulation) may be rewritten as:

$$s_x^{ij} = \frac{{}_nP_{x+n}^{ij}}{{}_nP_x^{i}} = \frac{{}_nL_{x+n}}{{}_nL_x}\,{}_nM_x^{ij} = \frac{{}_nL_{x+n}}{{}_nL_x}\,\overline{s_x^{ij}} \qquad (7)$$

(since ${}_nM^{ij}_x$ is the same as $\overline{s_x^{ij}}$).

Now, on assuming a linear integration of the person-years lived, the above becomes :

$$s_x^{ij} = \frac{1-(n/2)m_x^{i\delta}}{1+(n/2)m_{x+n}^{i\delta}}\,\overline{s_x^{ij}} \qquad (8)$$

Then, comparing (8) with (6) indicates that the only difference between the two formulas lies in the denominator which, by the way, is common to all the s_x^{ij} regardless of j.

In (8), the mortality considered at the denominator (that is, after the move to another region) refers to region i. Any individual initially in region i remains subject to the mortality of region i until the end of the interval.

By contrast, in (6), the mortality considered in the denominator is a weighted sum of the mortality rates in the various regions where the weights are the conditional probabilities of being in those regions at the end of the interval. Any individual initially in region i remains subject to the mortality of region i until the end of the interval if he/she stays there but becomes subject to the mortality of any region j if he/she moves there.

In practice, the difference between (8) and (6) has to be small since the difference Δ_i between the two denominators has to be small. This difference can be written as:

$$\begin{aligned}\Delta_i &= [1 + (n/2)\sum_k \overline{s_x^{ik}}\, m_{x+n}^{k\delta}\ ] - [\ 1 + (n/2)\, m_{x+n}^{i\delta}\ ] \\ &= (n/2)\ [\sum_{k\neq i} \overline{s_x^{ik}}\, m_{x+n}^{k\delta} - m_{x+n}^{i\delta}\,(1 - \overline{s_x^{ii}}\ )\ ] \qquad (9)\end{aligned}$$

and finally, since $1 - \overline{s_x^{ii}} = \sum_{k\neq i} \overline{s_x^{ik}}$, as:

$$\Delta_i = (n/2)\ [\sum_{k\neq i} (m_{x+n}^{k\delta} - m_{x+n}^{i\delta}\ )\ \overline{s_x^{ik}}\ ] \qquad (10)$$

In other words, Δ_i is made up of a sum of destination-specific terms in which the term pertaining to region j is proportionate to 1) the intensity of migrating to region j and to 2) the mortality differential between that regions i and j (Ledent, 2001).

Note that in the United Nations document where the modifed multiregional method was outlined, the migration rates were obtained from the Rogers and Castro model. However, in this case, migration rates have been estimated, since the census data provide figures on migration for the five years prior to the census. This UN model is simplified in that survival, migration and births are calculated independently. Whereas in the full multiregional migration model, all the demographic rates are applied simultaneously, in some modified models (as in this case), it is done sequentially (van Imhoff *et al.*, 1993). The model uses fertility and mortality rates for the region at the start of the interval, irrespective of whether a migration occurred. This could probably lead to some distortion, especially if the migration is from a region of lower mortality to a region of higher mortality (van Imhoff, 2000).

The method described above was applied to the nine provinces of South Africa, using data from the 1996 census. The spreadsheet file used for doing so was developed at Stats SA. This is the first stage towards applying a near-full multiregional migration model. As the migration propensities from the 1996 census are low, the modified model is probably sufficient at this stage.

4. Incorporation of deaths due to HIV/AIDS

In estimating additional deaths due to HIV/AIDS, a number of models are available for use. In this case, the AIDSProj spreadsheet was used to estimate the number of new AIDS cases, which was subsequently translated into AIDS deaths. The AIDSProj model was modified to handle South Africa's nine provinces. The main input data needed for the application of the model was base populations, annual growth rates and sero-prevalence data from antenatal clinic surveys. The user has the option of accepting the sero-prevalence data as reported or make some adjustments to them (taking account of male HIV infections, women of all ages and lower fertility among HIV+ women) using the data supplied in the model or otherwise. At the core of the model lies the fitting of a two-parameter gamma model onto the sero-prevalence data.

The equation of the gamma distribution is given by:

$$f(x;\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{(\alpha-1)} e^{-\frac{x}{\beta}} \tag{11}$$

where α and β are parameters of the model. The user is also expected to fit the gamma curve to the sero-prevalence data (by changing the parameters of the model, α and β). The sero-prevalence data (with provincial level breakdowns) used were those reported by the Department of Health, based on the ongoing national antenatal clinic surveys that have been conducted since 1990.

In making the adjustments to the sero-prevalence data to make them applicable to the whole population, two controls were considered:

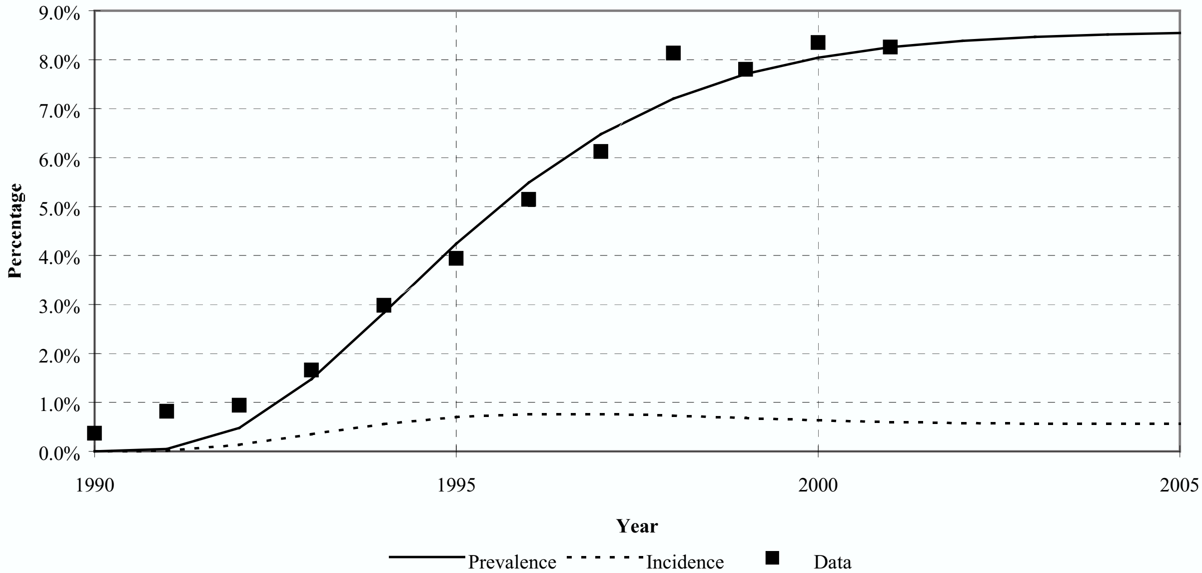
a) The estimated HIV/AIDS deaths for 1996 should not exceed the total deaths due to HIV/AIDS and HIV/AIDS-related deaths (a very broad category) obtained from the 1996 causes of deaths data.

b) The total estimated HIV/AIDS deaths for 2000 should be a proportion of the (upwardly) adjusted deaths obtained from the population register for 2000. As we do not have evidence suggesting that half of all deaths are due to HIV/AIDS, only a proportion less than 50% would be considered acceptable.

The adjustments were subsequently made (taking the controls into account) and the gamma functions were successfully fitted to the available sero-prevalence data (1990 to 2001). Figure 1 shows how the gamma function was successfully fitted on the national data.

Several assumptions are needed to translate new AIDS cases to 'additional deaths due to HIV/AIDS'. First, it is assumed that the period from new AIDS cases to death is a year. Second, it is assumed that all the new AIDS cases in 1995 which led to deaths in 1996 have been included in the life tables constructed for 1996 (adjusted for under-registration of deaths). In short, those AIDS deaths have become part of 'normal mortality' and are not part of 'additional deaths due to HIV/AIDS'. By October 2001, the additional deaths due to HIV/AIDS that would have occurred to the population alive in 1996 are taken as the deaths occurring between January 1997 to October 2001 less the AIDS deaths that would occur over this period as part of 'normal mortality'.

Figure 1: Fitting a gamma function to adjusted national adult HIV prevalence data, South Africa



5. Estimation of under-registration of death registration and the subsequent construction of life tables

5.1 Background and overview

Before selecting methods for estimating the degree of completeness of death registration, it is possible to obtain a reasonable range estimate within which calculated values should lie. In the case of South Africa, the disparity between urban and non-urban areas bears heavily on estimates of completeness of death registration. To get an idea of what this means, we take an example using the 1996 deaths data and figures from the 1996 census. The 1996 census reports 54% of the population living in urban areas and 46% in non-urban areas. The 1996 death data however reports 76% of deaths occurring to residents of urban areas and 24% to residents of non-urban areas. If we calculate the crude death rates, this gives 11,4 per thousand for the urban areas and 4,1 per thousand for the non-urban areas. This does not seem plausible. It could only mean that the deaths in the non-urban areas are considerably under-reported compared to deaths in the urban areas.

Since the population of South Africa is 46% non-urban, the average for South Africa as a whole would be lowered even if the completeness in the urban areas were high. If, for illustrative purpose, we apply the death rate in the urban areas to the non-urban areas, this would yield a figure of 248 313 for deaths in non-urban areas instead of the recorded 77 330. This would mean that deaths in non-urban areas are under-recorded by just over 3,2 times. So for South Africa as whole, the level of completeness of death registration as at 1996 would be 71% at the maximum. It is hard to see how it can be greater than this when births registered in 1996 was only about 50% complete at the maximum (based on comparison of currently registered births with population register figures which are continuously updated for late registrations).

To continue along the same lines of reasoning, if the death rate in the non-urban areas is higher than that in urban areas (a phenomenon that should not be taken for granted) the level of completeness of death registration in South Africa would even be lower. While there are techniques for estimating completeness of death registration, as mentioned earlier, these are only usable when registration of deaths is sufficiently high. When registration of deaths is lower than the suggested threshold figure, the techniques have to be used with great caution. Stats SA has devised a way of going around this. The method, however, would need estimates of completeness for South Africa as a whole and for urban areas. These would have to be obtained by one of the existing methods for estimating the level of completeness of death registration.

Preston (1984) provides a comprehensive review of direct and indirect techniques for estimating the completeness of death registration. Each of the techniques was developed based on one of three assumptions, namely: the assumption of population stability, the assumption of quasi-stability and the assumption that registered deaths represent a constant proportion of true deaths at each age within the age range considered.

While there are many indirect methods proposed for estimating the completeness of death registration, four of them stand out as being most widely applied. These are the Brass method (Brass, 1975), the Preston and Coale method (Preston *et al.*, 1980), the Preston and Hill method (Preston and Hill, 1980) and the Bennet and Horiuchi method (Bennet and Horiuchi, 1981). In his review of techniques available for estimating the completeness of death registration, Preston (1984) also gave more attention to these four techniques. The first two methods were widely applied in the model life table construction project of the United Nations and in that of the Organisation for Economic Co-operation and Development (OECD) (Preston, 1984).

5.2 The method adopted and justification for doing so

There is some confusion on the application of the Preston-Coale and the Bennet-Horiuchi methods with regards to the registered deaths used. The equation on which the method is based requires that all data refer to a ‘population’- so if the growth rates are intercensal, so should be all other data (Preston, 2000). In his 1984 review, Preston had noted the following:

> Note, however that if intercensal growth rates are used, the Bennet and Horiuchi (and the Preston and Coale) procedures should also be used with intercensal deaths not simply with deaths centred on one of the two censuses. (Preston, 1984:69-70)

In the paper by Preston *et al.* (1980) where this method was outlined, the method was applied to Korean data and regarding the deaths, *D(a)*, and the population, ${}_5N_a$, used in the calculation, the following footnote was written:

> D(a) is deaths over the intercensal period 1970-1975; ${}_5\,N_a$ is five times the average of the population in the two censuses. (Preston *et al.*, 1980:197)

However it is possible to use data at one census to estimate what the intercensal data would be. Obviously, some error can creep into this substitution (Preston, 2000).

Similarly, Bennet and Horiuchi define the deaths term ${}_5D_{a-5}$ used in their formula as,

> the number of deaths occurring in the cohort (or the stationary population age group) in the intervening period. (Bennet and Horiuichi, 1981:210)

Indeed, in the BENHR program included in MORTPAK for applying the Bennet-Horiuchi method, the deaths data stated are clearly written as ‘intercensal deaths’. The Preston and Hill method also requires intercensal deaths and two censuses.

In short, in their original formulation, three of the four methods mentioned above require intercensal deaths and two censuses. In the South African context, going the route of using the 1991 census and intercensal deaths is not a very appealing option and is beset by many problems. Some of the problems deal with the changing demarcation of provincial and national boundaries over this period. The 1991 and 1996 censuses were referring to different geographic entities. A more serious problem is the exclusion of death statistics from the former TBVC states. If, for example, the Preston and Coale method is to be used, one would have to ‘experiment’ with different growth rates in a situation where one does not want to use the intercensal growth rates. While this procedure has worked well in at least one case, the caution from one of the authors of the technique is that, in general, it may not be satisfactory (Preston, 1984). The route of using deaths centered on one year (instead of the intercensal deaths) and ‘experimenting’ with different growth rates, it seems, is an exercise in futility. The situation is even more serious in the case of the Bennett and Horiuchi technique, where the ‘experimentation’ has to be done for all the five-year age groups. A golden rule in indirect techniques is that one cannot force the application of a technique just because of its more solid theoretical basis. Indeed, where the data are available, the choice has to be for the technique with the more solid theoretical basis. This was the case in the Island of Mauritius. Census data were available for 1972 and 1983 and registered deaths for the intervening years. Since the intercensal period was 11 years (not a multiple of five), the Preston-Hill method was not found to be suitable. As Mauritius had a full set of uninterrupted intercensal deaths and had the advantage of being a small island (without changes in national boundaries), the Bennet and Horuichi method was appplied and yielded good results (Bah, 1995).

If one has to use data from only one census (1996) and deaths for that year and one does not want to use either intercensal growth rates or intercensal deaths, then one’s options are very limited. One would have to resort to the Brass method or employ direct methods, to estimate the degree of incompleteness of death registration. While the Brass method is more appropriate under conditions of stability, it has been found to work adequately in quasi-stable conditions. As Preston (1984:69) observed,

> When errors in the data are abundant, the analyst may wish to impose the stability assumption in order to 'discipline' the data

And United Nations (1983:132) has noted that,

> A history of declining mortality causes a departure from the stable age distribution but usually a rather limited departure

Beside the advantage of having lesser data requirements, another advantage that the Brass method has over the Preston and Coale method is the flexibility of eliminating data points believed to be faulty. Clearly more research is needed to modify the Brass method in such a way that the modification does not involve circularity or the need for additional data. In the interim, the Brass method has been used for the national figures and for the urban areas. For the other population sub-groupings, a weighting method has been used which takes into account the urban/non-urban disparity in South Africa.

In estimating completeness in death registration, there are some in-built checks for assessing the validity of the results and for safeguarding against having unrealistic life tables. For countries known to have incomplete recording of deaths, an estimate of completeness of 100% or more would be unacceptable. When the completeness of recording is very low and the estimated level of completeness is far higher than the actual level (a case of serious underestimation), the life expectancy derived from the life table would be very high. A comparison of the results with those of countries with accurate life tables could easily show the problem. On the other hand, when completeness of death registration is fairly high, to err in the exact level of incompleteness does not affect the life expectancy much. The mathematics show that if one comes to within 10 per cent of the true value in the corrected estimates of age-specific death rates, one should be within 1-4 per cent in the corresponding estimates of life expectancy at birth. For example, at life expectancy at birth of 55 years, the error introduced by a 10 per cent error in the estimate of the completeness translate to 0,5 to 2,2 years (Preston, 1984).

5.3 Outline of the Brass method

In general, for any closed population, the following equation applies:

$$r_{x+}^{T} = b_{x+}^{T} - d_{x+}^{T} \qquad (12)$$

where,

r_{x+}^{T} is the true partial growth rate for age *x*.

b_{x+}^{T} is the true partial birth rate for age *x* and is given as:

$$b_{x+}^{T} = \frac{{}_5N_x + {}_5N_{x-5}}{10 * N_{x+}} \qquad (13)$$

where

${}_5N_x$ is the number of persons recorded in the age interval from *x* to *x* + 5, and

d_{x+}^{T} is the true partial death rate for age *x*.

The growth balance method makes two major assumptions: a) the assumption of population stability and b) the assumption that the completeness of death registration is invariant to age within the range specified.

This means that

$$r_{x+}^{T} = r^{T} \qquad (14)$$

and

$$d_{x+} = C * d_{x+}^{T} \qquad (15)$$

where *C* is the completeness of death recording and d_{x+} is the partial death rate for age *x* based on recorded deaths.

Using these two assumptions, the closed population equation could be written as follows:

$$b_{x+}^{T} = r^{T} + \frac{1}{C} d_{x+} \qquad (16)$$

This is a linear equation with the parameters r^T as intercept and *1/C* as the slope. The reciprocal of the slope therefore gives the proportion of completeness of death recording.

The GRBAL program in PAS was used to fit the growth balance method to the data. The resulting correction factors were used to adjust the age specific death rates.

These adjusted age specific death rates were used as input into the LIFTB program of MORTPAK. Note that the Brass method is not seen as a perfect tool. It is however one of the very few tools available given the constraints outlined earlier. As a way out, it has been used only for South Africa as a whole and for urban areas. A weighting method is used for obtaining estimates for non-urban areas and the provinces. This method is elaborated upon below.

5.4 Method for incorporating the relative distribution of deaths and population in urban and non-urban areas

Upon applying the Brass method to national figures and those for all the population sub-groupings (urban, non-urban, and provincial breakdowns by gender), it emerged that that there were cases where the plots of the partial birth and partial death rates were clearly aberrant. The most aberrant cases were the following: non-urban males and females, Eastern Cape males and females and Northern Province males and females. According to the 1996 census, Eastern Cape and Northern Province are largely non-urban in their population composition, Eastern Cape being 63,4% non-urban and Northern Province being 89,0% non-urban. As such, it can be reasoned that because of the largely non-urban character of these provinces, registration of deaths becomes close to, or lower than, the minimum level required for application of these methods.

The strategies adopted for estimating the level of completeness of death registration for non-urban areas and for all the provinces are given in what follows.

5.4.1 For non-urban population

Given that the true number of deaths in South Africa is the sum of the deaths in the urban and non-urban areas, one has the following:

$$D_{RSA}^{T} = D_{ur}^{T} + D_{nur}^{T} \qquad (17)$$

where D^T stands for the true number of deaths and the subscripts *ur* and *nur* stand for urban and non-urban respectively.

As the true number of death, D^T , is a multiple, *k* or 1/C, of the registered deaths, *D,* one has that:

$$D^{T} = k * D$$

Hence,

$$k_{nur} = \frac{k_{RSA} * D_{RSA} - k_{ur} * D_{ur}}{D_{nur}} \qquad (18)$$

or,

$$k_{nur} = \left(\frac{D_{RSA}}{D_{nur}}\right) * k_{RSA} - \left(\frac{D_{ur}}{D_{nur}}\right) * k_{ur} \qquad (19)$$

In short, having obtained the more reliable correction factors for the national and for the urban population, the additive relationship between the number of the deaths in the sub-components and the whole, allows one to estimate the correction factors for the less reliable, or highly incomplete, non-urban data.

5.4.2 For all provinces

Similarly, for any given province, *i,* the total number of deaths in the province is a sum of the deaths in the urban and non-urban areas of that province.

$$D_i^T = D_{ur,i}^T + D_{nur,i}^T \qquad (20)$$

Hence, in terms of reported deaths, one has:

$$k_i * D_i = k_{ur,i} * D_{ur,i} + k_{nur,i} * D_{nur,i} \qquad (21)$$

or,

$$k_i = \left(\frac{D_{ur,i}}{D_i}\right) * k_{ur,i} + \left(\frac{D_{nur,i}}{D_i}\right) * k_{nur,i} \qquad (22)$$

Assuming that $k_{ur,i} = k_{ur}$ and that $k_{nur,i} = k_{nur}$

one has,

$$k_i = \alpha_i * k_{ur} + (1 - \alpha_i) * k_{nur} \qquad (23)$$

where

$$\alpha_i = \frac{D_{ur,i}}{D_i}$$

5.5 The life tables obtained and their plausibility

For the pre-1996 life tables (1985-1994), the summary of the life expectancies at birth obtained for the different population groups are shown in Table B. For the sake of brevity, the methodology used is not outlined here. The reader is referred to the above-mentioned Stats SA life table report for further details. Research on South African life tables has shown that death registration was incomplete (at varying degrees) for all population groups (Bah, 2000). As such, previously published life tables

cannot be used to assess the plausibility of the life tables derived for this work. Instead, we have constructed life tables at two extremes, the assumption of 100% completeness and the assumption of 50% completeness. The summary of the life expectancies at birth obtained are shown in Table C. For most of the provinces (with the exception of the largely non-urban provinces), the life expectancies of these two levels of completeness should bracket the life table obtained. Table D gives the summary of the life expectancies at birth using the estimated level of completeness.

Table B: Life expectancy at birth as derived from the 1985-1994 life tables, for South Africa as a whole and population groups

Population group	Male	Female
South Africa	54,12	65,38
African/Black	52,51	64,62
Coloured	57,36	65,02
Asian/Indian	60,95	68,90
White	65,22	73,08
Other and unspecified	59,75	66,04

Table C: Implied life expectancy at birth at two given levels of completeness of death registration, for South Africa as a whole, provinces and urban/non-urban residence, 1996

	Male		Female	
	100% completeness	50% completeness	100% completeness	50% completeness
South Africa	59,543	47,112	68,952	57,031
Western Cape	60,084	48,217	69,296	58,270
Eastern Cape	62,128	49,905	72,716	61,834
Northern Cape	56,083	43,335	64,538	51,893
Free State	56,126	43,121	62,885	49,552
KwaZulu Natal	56,154	43,186	67,388	54,745
North West	60,152	47,916	67,906	55,382
Gauteng	59,538	47,165	67,036	55,508
Mpumalanga	59,062	46,919	67,253	55,250
Northern Province	65,970	54,412	76,524	66,770
Urban	54,241	41,631	65,455	52,526
Non-urban	69,772	57,185	77,326	68,908

Table D: Life expectancies at birth as derived from the 1996 life tables, for South Africa as a whole, provinces and urban/non-urban residence.

	Male		Female	
	Percentage completeness of death registration	Life expectancy at birth e(0)	Percentage completeness of death registration	Life expectancy at birth e(0)
South Africa	66,90	52,11	65,40	61,60
Western Cape	79,55	55,75	81,34	65,68
Eastern Cape	64,43	54,16	64,18	65,76
Northern Cape	76,54	51,17	77,51	59,92
Free State	70,25	49,34	70,44	56,11
KwaZulu Natal	62,43	47,16	60,29	58,13
North West	56,51	50,00	53,30	56,47
Gauteng	81,26	55,54	83,83	63,87
Mpumalanga	58,65	49,56	56,12	57,20
Northern Province	49,26	54,10	45,06	65,14
Urban	84,30	50,85	86,7	62,58
Non-urban	39,10	52,76	35,7	63,92

Perhaps a surprising feature of Table D is the high sex differential in life expectancies at birth. This is comparable to experiences elsewhere. In Poland in 1990, Trovato and Lalu (1996) report that the male-female difference in life expectancy at birth was 9,14 years. For the same year, the male-female difference in life expectancy at birth was 8,89 years in Hungary and 8,41 years in France. In that study, Trovato and Lalu offered possible reasons for these differences as well as the differences observed in 27 other industrialised countries.

Another surprising feature is the slightly higher life expectancy in non-urban as compared to urban areas and the relatively high life expectancy at birth in some predominantly non-urban provinces. This can be situated within the experience of another Southern African country. In Zimbabwe, it has been found that mortality in the predominantly non-urban province of Matabeleland South was lower than that in other some predominantly urban provinces. Root (1997) further investigated this in a study and singled out population density as the possible reason for this differential. In the case of South Africa, more research is needed to fully establish the reality of this phenomenon.

6. Growth rates and interpolation

6.1 Overview

If population estimates are desired for quinquennial years, they can be produced by standard cohort-components method (with population and demographic rates grouped into five-year age groups) without need for population growth rates. If population estimates are desired for each year, they can be obtained by one of two approaches: first, by grouping population and demographic rates into single years and doing yearly projections; or second, by interpolating between successive quinquenial age distributions. The first approach is feasible in 'data-rich' countries and is utilised more in the context of projections produced by microsimulation. The latter is more widely used and is often done in the context of projections produced by macrosimulation (van Imhoff, 1998). The interpolation referred to is, in essence, making use of a growth rate (either linear, geometric, compound or exponential).

Using the census totals for 1996 and the projected totals for 2001, exponential growth of the population was assumed and the 'inferred' growth rate, *r,* was obtained as follows:

$$P_{t_2} = P_{t_1} * \exp(r * (t_2 - t_1)) \tag{24}$$

where t_1 is the initial time and t_2 is the final time.

Hence, $r = 0.2 * \log_e (P_{2001} / P_{1996})$ (25)

Regarding the population estimates for the period 1991-1996, one could either use growth rates derived for that period (using backward projection of the 1996 census to 1991 and inferring the growth rate from the totals) or assume that the growth rates for the period 1996-2001 would be applicable for the earlier period as well. Both methods were attempted but the latter procedure gave more plausible results and hence has been adopted.

For any given date, *tx* , either prior to 1996 or after 1996, the mid-year population estimate is obtained as follows:

$$P_{t_x} = P_{1996.775_1} * \exp(r * (t_x - 1996.775_1)) \tag{26}$$

where 1996.775 is the decimalised equivalent of the Census night, 10 October 1996.

Note that the 'inferred' growth rates should not be confused with intercensal growth rates obtained from data from two censuses. However, the more realistic the population projection model, the closer the two rates become. In passing, it must be noted that demographic literature also mentions another growth rate, the intrinsic growth rate (the rate that is achieved when the population becomes stable). Such a growth rate also becomes closer to the intercensal growth rate the closer the population is to stability. The growth rates calculated are shown in Table E. These inferred growth rates differ from the previously published ones since internal migration has now been taken into account and the life tables used are based on reported deaths, adjusted for under-registration.

6.2 Short proof of why male inferred rate of growth could be greater than that of female even though the female population may be higher

Using the same symbols as above, and adding the *m* superscript to indicate males and *f* for females, one has the following:

$$P_{t2}^m = P_{t1}^m \exp(r^m * (t2 - t1)) \tag{27}$$

$$P_{t2}^f = P_{t1}^f \exp(r^f * (t2 - t1)) \tag{28}$$

$$r^m - r^f = \frac{1}{(t2 - t1)} * \left[\log_e \left(\frac{P_{t2}^m}{P_{t2}^f} \right) + \log_e \left(\frac{P_{t1}^f}{P_{t1}^m} \right) \right] \tag{29}$$

$$r^m - r^f = \frac{1}{(t2 - t1)} * \left[\log_e SR_{t2} - \log_e SR_{t1} \right] = \Delta \tag{30}$$

where SR is the overall sex ratio in the population.

If $\Delta > 0$ then $r^m > r^f$

In short, it is the relative size of the male and female population at two points in time that determine whether the male or female rate of growth would be greater, not the absolute sizes of the male and females population at both time points.

7. Reconciling the totals with the sum of the components

In preparing mid-year estimates for the total population and for some of its components, one could either use a 'bottom-up' approach, a 'top-down' approach or a 'hybrid' approach. In the bottom-up approach, the national figures are obtained by aggregating the sub-national estimates. In this case, the problem of reconciling discrepancies does not arise. However, when the estimates for the national and the sub-national populations are obtained independently, there is often some discrepancy between the sum of the initial estimates of the components and the total population. There is need then to adjust for the differences to arrive at final estimates. This gives rise to the top-down approach, wherein the sub-national totals are forced to sum up to the national. This approach is also appropriate when the national figures are seen to be more reliable than the sub-national ones. The procedure for controlling is known variously as iterative proportional fitting, raking and rim-weighting. The hybrid approach attempts to combine the advantages of the top-down approach and the bottom-up approach.

In this report, a hybrid (bottom-up/top-down) approach has been used in arriving at the totals. Where internal migration has been included in the projection, a bottom-up approach was seen as more appropriate than a top-down approach and was therefore used. However in the case of the population groups and urban/non-urban estimates, a top-down approach was used.

In the case where reconciliation was done, it was done separately for males and females. The program used for doing this is CTBL32 in Population Analysis Software (PAS) by the US Bureau of Census. The urban and non-urban projections as well as the projections for the population groups were constrained to the national projection obtained by adding up the provinces. Note that the constraining of the population group totals to the national total ensures the population group projection has not been overestimated (by virtue of using life tables with lighter mortality).

8. Note to users who would like to prepare special population estimates

A) If a user needs to estimate the population of a given population group at the provincial level for a non-census date, one option is as follows. Make the assumption that the national growth rate for that population group applies at the provincial level and proceed with applying that rate on the provincial total of that population group. Other options are possible.

B) If a user needs to estimate the population at sub-provincial level for a non-census date, one option is as follows. Estimate the ratio of that sub-provincial total to the province total and apply that ratio on the provincial estimate at the desired date. Other options are possible.

9. Materials/software used

Several software programs were used in the course of preparing this report. Specifically, the following have been used:

1. Population Analysis System (PAS) from the US Census Bureau. The product is shareware and can be downloaded from the following site: http://www.census.gov/ipc/www/pas.html.
2. AIDSProj from Futures Group International. The product is shareware and can be downloaded from the following site: http://www.tfgi.com/aidsproj.asp.
3. The United Nations Software Package for Mortality Measurement (MORTPAK-LITE) and QFIVE. Both products have copyrights and can be purchased from the United Nations at the following address: Department of International and Economic and Social Affairs, United Nations, New York, New York 10017, USA.
4. Several linked Excel programs and FORTRAN programs developed at Stats SA.

Table E. Exponential growth rates inferred from the projected population, with and without taking into account additional deaths due to HIV/AIDS

Broad category	Unit	Gender	Exponential growth rates, *r*, for 1996-2002: Taking into account additional deaths due to HIV/AIDS	Exponential growth rates, *r*, for 1996-2002: Without taking into account additional deaths due to HIV/AIDS
Total	South Africa	Male	-	0,019841
		Female	-	0,019738
Urban/non-urban	Urban	Male	-	0,014981
		Female	-	0,018439
	Non-urban	Male	-	0,025898
		Female	-	0,023638
Population group	African	Male	-	0,028497
		Female	-	0,028660
	Coloured	Male	-	0,020119
		Female	-	0,020070
	Asian/Indian	Male	-	0,016957
		Female	-	0,018251
	White	Male	-	0,010179
		Female	-	0,011871
	Other and unspecified	Male	-	0,020193
		Female	-	0,018828
Province	Western Cape	Male	0,014222	0,014507
		Female	0,015921	0,016272
	Eastern Cape	Male	0,023244	0,023857
		Female	0,020179	0,020867
	Northern Cape	Male	0,008540	0,008974
		Female	0,010844	0,011379
	Free State	Male	0,013986	0,015046
		Female	0,014714	0,016074
	KwaZulu-Natal	Male	0,015969	0,017635
		Female	0,015588	0,017543
	North West	Male	0,014902	0,015993
		Female	0,015493	0,016884
	Gauteng	Male	0,015743	0,016928
		Female	0,018590	0,020166
	Mpumalanga	Male	0,020924	0,022137
		Female	0,020832	0,022341
	Limpopo	Male	0,032996	0,033383
		Female	0,026924	0,027356

Note: The categories used in the first column are non-overlapping since the provincial totals add up to the national total, the population group totals add up to the national total and the total for the urban and non-urban locations also add up to the national total.

References

Arriaga, E. (1983). Estimating fertility from data on children ever born by age of mother. *International Research Document. No. 11*. United States Bureau of the Census. Washington, D. C.

Bah, S. (1990). The application of multiregional demography to the estimation of migration using African data. Paper presented at the conference on the 'Role of Migration on African Development: Problems and Policies for the 1990s', Nairobi, Kenya, February 19-23, 1990.

Bah, S. (1995). The evaluation of the completeness of death registration in the presence of high out-migration: the case example of Mauritius. *International Institute for Vital Registration and Statistics Technical Papers.* No 61. August 1995.

Bah, S. (2000). A critical review of South African life tables. *Canadian Studies in Population* 27, 2:283-306.

Bennett, N. and S. Horiuchi (1981). Estimating the completeness of death registration in a closed population. *Population Index*, 47,2:207-221.

Brass, W. (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data.* Laboratories for Population Statistics, Chapel Hill, N.C.

Brass, W. and Hill, K. (1973). Estimating adult mortality from orphanhood. International population conference, Liège. *International Union for the Scientific Study of Population*. 3:111-123.

Daponte, B. E., Kadane, J. B., and Wolfson, L. J. (1997). Bayesian demography: Projecting the Iraqi Kurdish population, 1977-1990. *Journal of the American Statistical Association*. 92,440:1256-1267.

Daponte, B and Wolfson, L. (2000). Exploring Bayesian Demographic Approaches to Projecting the Population of South Africa. Consultancy report submitted to Nathan Associates and USAID, Pretoria.

Department of Health (2002). *Summary Report: National HIV and Syphilis Sero-Prevalence Survey of Women Attending Public Antenatal Clinics in South Africa, 2001*. Department of Health, Pretoria.

Doeve, W. (1984). *Operational Multiregional Demography in Developing Countries: An Application to Thailand.* Netherlands Interdisciplinary Demographic Institute (NIDI) Working Paper No. 48.

Hill, K. (1984). An evaluation of indirect methods for estimating mortality. In: *Methodologies for the Collection and Analysis of Mortality Data*. Edited by J. Vallin *et al.* IUSSP and Ordina Editions, Liège.

Hill, K. and Trussell, J. (1977). Further developments in indirect mortality estimation. *Population Studies*. 31, 2: 313-334.

Ledent, J. (1980). Multistate life tables: Movement versus transition perspectives. *Environment and Planning A 12*, 533-562.

Ledent, J. (1981). Constructing multiregional life tables using place-of-birth specific migration data. In: *Advances in Multiregional Demography*. Edited by A. Rogers. IIASA, Laxenburg, Austria.

Ledent, J. (1982). Transition probability estimation in increment-decrement life tables: Using mobility data from a census or a survey. In: *Multidimensional Mathematical Demography*. Edited by K. Land and A. Rogers. Academic Press, New York.

Ledent, J. (2000). Correspondence with the Directorate of Vital Statistics, Stats SA, Pretoria.

Ledent, J. (2001). Correspondence with the Directorate of Vital Statistics, Stats SA, Pretoria.

Ledent, J. and Rees, P (1980). Choices in the construction of multiregional life tables. WP-80-173. IIASA, Laxemburg, Austria.

Philpov, D. and Rogers, A. (1981). Multistate population projections. *IIASA Reports* 4,1:51-82.

Pollard, J. H. (1982). The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries* 109: 225-240.

Preston, S. (1984). Use of direct and indirect techniques for estimating the completeness of death registration systems. In: *Data Bases for Mortality Measurement.* Edited by United Nations, New York.

Preston, S. (2000). Correspondence with the Directorate of Vital Statistics, Stats SA. Pretoria.

Preston, S. and K. Hill (1980). Estimating the completeness of reporting of adult deaths in population that are approximately stable. *Population Studies*, 34,2:349-366.

Preston, S. *et. al*. (1980). Estimating the completeness of reporting of adult deaths in populations that are approximately stable. *Population Index* 46,2:179-202.

Rees, P. (1977). The measurement of migration from census data and other sources. *Environment and Planning A* 9, 247-272.

Rees, P. (1994). Estimating and projecting the populations of urban communities. *Environment and Planning A* 26, 1671-1697.

Rees, P. and Kupiszewski, M. (1999). Internal migration: What data are available in Europe? *Journal of Official Statistics,* 15,4:551-586.

Rogers, A. (1973) The mathematics of multiregional demographic growth. *Environment and Planning 5, 3-29*

Rogers, A. (1975). *Introduction to Multiregional Mathematical Demography*. John Wiley, New York.

Rogers, A. (1985). *Regional Population Projection Models.* Scientific Geography Series, vol 4. Sage Publications, Beverly Hills, Carlifornia.

Root, G. (1997). Population density and spatial differentials in child mortality in Zimbabwe. *Social Science and Medicine*. 44,3:413-421.

Rogers, A. and Castro, L. (1982). Model schedules in multistate demographic analysis: The case of migration. In: *Multidimensional Mathematical Demography*. Edited by K. Land and A. Rogers. Academic Press, New York.

Schmertmann, C. (1999). Estimating multistate transition harzards from last-move data. *Journal of the American Statistical Association*. Vol 94,445:53-63.

Simpson, S (ed). (1998). *Estimating with confidence*. Local Authorities Research and Intelligence Association. Wokingham.

Stats SA (1999). Recorded births, 1990-1997. Stats SA statistical release P0305. Pretoria.

Stats SA (2001a). Mid-year estimates, 2001. Stats SA statistical release P0302. Pretoria.

Stats SA (2001b). Advance release of recorded deaths, 1997-2000. Stats SA statistical release P0309.1. Pretoria.

Stover, S. (2000). Correspondence with the Directorate of Vital Statistics, Stats SA, Pretoria.

Stover, S. and Way, P (1998). Projecting the impact of AIDS on mortality. *AIDS* 12 (suppl 1): S29-S39.

Udjo, E. (1999). Recent evidence of levels, trends and differentials in fertility in South Africa. Paper presented at the workshop on ‘Fertility in Southern Africa’, School of Oriental and African Studies, University of London, 22-24 September 1999.

United Nations (1970). Manual VI: Methods of Measuring Internal Migration. United Nations, New York.

United Nations. (1983). *Manual X: Indirect Techniques for Demographic Estimation*. United Nations, New York.

United Nations. (1988). *MortPak-Lite: The United Nations software package for mortality measurement*. United Nations. New York.

United Nations. (1994). *AIDS and the Demography of Africa*. United Nations. New York.

United Nations. (1990). *Step-by-step Guide to the Estimation of Child Mortality*. United Nations. New York.

United Nations (1992). *Preparing Migration Data for Subnational Population Projections.* United Nations, New York.

van Imhoff, E. *et al* (1993). *Regional Population Projections in the Countries of the European Economic Area.* NIDI CBGS Publications, The Hague, Netherlands.

van Imhoff, E. (1998). Microsimulation methods for population projections. *Population* 10-1, 1998.

van Imhoff, E. *et al.* (1997). The selection of internal migration models for European regions. *International Journal of Population Geography,* 3: 137-159.

van Imhoff, E. (2000). Correspondence with the Directorate of Vital Statistics, Stats SA, Pretoria.

Willekens, F. (1982). Multidimensional population analysis with incomplete data. In: *Multidimensional Mathematical Demography*. Edited by K. Land and A. Rogers. Academic Press, New York.

Willekens, F. (2001). Correspondence with the Directorate of Vital Statistics, Stats SA, Pretoria.

Willekens, F. and Rogers, A. (1978). *Spatial Population Analysis: Methods and Computer Programs.* RR-78-18. IIASA, Laxenburg, Austria.