Centre for Actuarial Research (CARe)

A Research Unit of the University of Cape Town

Estimation of mortality using the South African Census 2001 data

Rob Dorrington
Tom A Moultrie
Centre for Actuarial Research
University of Cape Town

Ian M Timæus
Centre for Population Studies
London School of Hygiene & Tropical Medicine

CARe Monograph No. 11

ISBN: 0-7992-2281-X

Centre for Actuarial Research
University of Cape Town
Private Bag
7701 Rondebosch
SOUTH AFRICA

+27 (0)21 650 2475 (T)
+27 (0)21 689 7580 (F)
care@commerce.uct.ac.za (E)

Centre for Actuarial Research (CARe)

A Research Unit of the University of Cape Town

Estimation of mortality using the South African Census 2001 data

Rob Dorrington
Tom A Moultrie
Centre for Actuarial Research
University of Cape Town

Ian M Timæus
Centre for Population Studies
London School of Hygiene & Tropical Medicine

CARe Monograph No. 11

Summary

Estimation of mortality in South Africa has always been problematic. While reasonably accurate life tables have been computed for the Coloured, Indian and White population groups, mortality rates for the African population, and the South African population as a whole, have always been fairly approximate. Nonetheless it would appear from the official life tables and estimates of mortality derived from reconstruction of census populations that mortality in South Africa fell for many decades prior to the early to mid-1980s. Between then and the mid-1990s, rates appear to have levelled off, and even to have risen in some population groups.

The release of the 2001 census data provides an opportunity to update our estimates of mortality and decide if these past trends have continued through the intercensal period and, in particular, measure whether mortality has increased in line with predictions of the impact of HIV on mortality. Since the 2001 census was run exactly five years following the first census in democratic South Africa, one can make use of information from both censuses together to provide better estimates than can be made from a single census alone.

As part of the process of deriving estimates of mortality we interrogate the quality of the data, and in particular the edits performed by Statistics South Africa on the data, mainly to replace missing or "don't know" responses. This exercise has led us to the conclusion that, although often inconsequential in magnitude, most of the edits produced results which are inconsistent either with what might be expected or with the data that did not need editing. On the basis of these investigations, we recommend that the data be released without these edits.

Further, in the case of the data on survival of children ever born we investigated various combinations of edits and assumptions but are unable to produce any sequence of proportions surviving by age of mother that is remotely plausible. Thus, very reluctantly, we are forced to conclude that these data in the 2001 census are too poor to be usable.

As far as the data on adult mortality are concerned, these are used without the edits. In the case of deaths reported by household, the total numbers of deaths reported in the census by sex, population group and province were simply apportioned by age according to the age distribution of the reported deaths for which age was not imputed. In the case of the data on survival of parents, "don't know" and missing responses have been excluded.

The method used to produce rates by population group and for the country as a whole is to apply the generalization of Brass's Growth Balance method proposed by Hill to the numbers of deaths estimated, from the population register, to have occurred between the 1996 and 2001 censuses, the estimates of the population from each of the censuses, and an estimate of the intercensal migration derived from the reconciliation of the censuses. From this procedure the following estimates of completeness of death reporting are derived:

	African	*Coloured*	*Indian*	*White*	*National*	*National adjusted*
Males	63.9%	69.5%	64.8%	77.0%	83.5%	83.4%
Females	66.6%	69.7%	83.4%	78.5%	86.7%	84.5%

Inconsistencies between the completeness estimates by population group and for the country as a whole are due to the fact that some 20 per cent of the death certificates did not record the population group of the deceased.

Rates of mortality produced by grossing up the number of deaths by the extent of completeness were then graduated by fitting Brass's General Standard to the older ages and lightly smoothing the raw data at other ages. This process does not affect the size of the 'hump' in the rates for young adults which is indicative mortality due to HIV/AIDS.

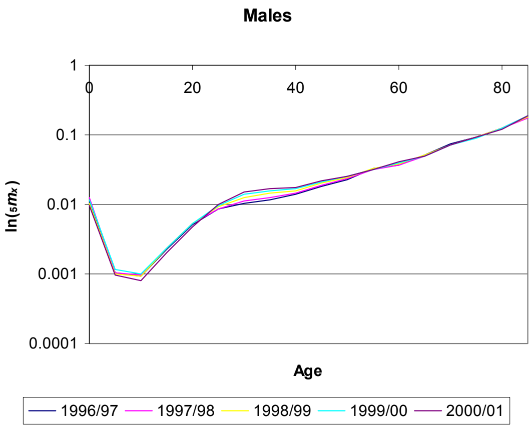
A check of these rates (${}_{15}q_{50}$ specifically) against those produced using the data on survival of parents reveals a fair degree of consistency, although it does suggest that mortality of White men and to a lesser extent White women may be exaggerated slightly by the estimates based on registered deaths.

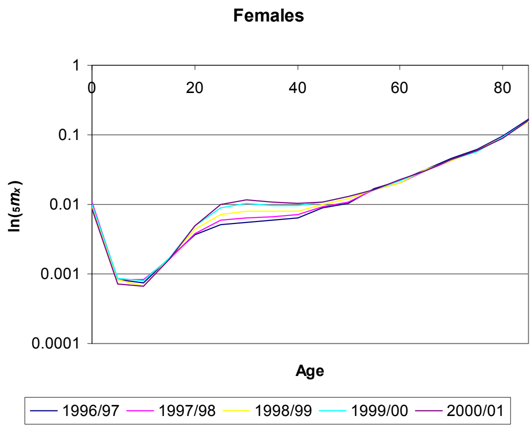
In order to produce mortality rates by province we first compare the number of deaths reported by households as occurring in the 12 months before the census with those expected on the basis of our estimates of national mortality, for each sex and population group. We then apportion the number of deaths expected on the basis of the estimates derived above, for each sex and population group, to the provinces in proportion to the number of deaths reported by households according to age, sex and population group. Rates are then calculated using these numbers of deaths and estimates of provincial population derived from the censuses, for the African population group and the provinces as a whole.

Although the rates produced are not inconsistent with expectations, the results do exhibit a number of features needing further investigation. In particular, rates of mortality at the older ages rise more rapidly with age in Western Cape and Gauteng than for the rest of the country. Once again the rates (${}_{15}q_{50}$ specifically) were compared with those produced using the data on survival of parents. This reveals a fair degree of consistency, particularly in the case of women, and particularly if one takes into account the fact that parents may not have died in the province in which their children were enumerated.

Comparison of the rates produced in this report with those from other sources suggests that, not only did adult mortality rates level off from the early to middle 1980s, but, in the case of the African (and hence South African) population, they appear to have increased rapidly since 1995, in line (as shown in the figure below for the whole population) with increases expected due to HIV/AIDS.

Since it is not possible to produce estimates of recent childhood mortality, it is not possible to produce reliable life tables. However, illustrative life tables are presented using ${}_{5}q_{0}$ estimated directly from the deaths in the last 12 months reported by households and the census estimates of the numbers of children under age five. In addition, maximum and minimum values of select indicators of mortality were estimated, suggesting that life expectancy for the country as a whole is unlikely to have been less that 49.6 years or greater than 59.2 years. The report concludes by identifying further work that needs to be done to improve on the estimates produced so far.





For the first time the question on deaths in the household over the past year also asked about whether the death was due to violence or occurred close to the time of birth. Although it is not the purpose of this report to analyse how well this question worked, a quick inspection suggests that the age distribution and the level of mortality due to external/violent causes for males are quite plausible (at around 18% of all deaths), but for females the rates appear to be too high, particularly at the older ages. The maternal mortality rate appears to be implausibly high at 575 per 100 000 births, however, this is only 6.5% of all deaths in the 15-49 age range which is well within the range of estimates from other sub-Saharan countries. On the other hand the high number could in part be attributable to the fact that a third of these deaths had age imputed, presumably on the basis of the cause of death, which might not have been universally correctly captured.

Authors' Note

This report was prepared for Statistics South Africa in terms of a contract awarded in October 2003 to prepare an analysis of the mortality data collected in the 2001 census. The authors are grateful to Statistics South Africa for this opportunity, as well as for all their assistance in making the data available to us.

Dr Ravai Marindo contributed to a preliminary draft of Chapter 2. However, all errors are our own and all opinions are those of the authors and do not necessarily reflect those held by Statistics South Africa.

Rob Dorrington, Tom A Moultrie and Ian M Timæus
Cape Town, November 2004

Table of Contents

List of Tables

List of Figures

1 Introduction

The overriding purpose of any census must be to provide as accurate as possible estimates of the numbers of people by major demographic sub-divisions of the population. In order not only to be able to verify and possibly improve on these estimates, but also to be able to project them to more recent time points, one needs to have reliable estimates of the components that determine the shape and size of the population, namely, fertility, mortality and migration.

In addition to this important role, these components are also important in their own right as indices of the demographic well-being of the country. As far as mortality is concerned, rates of infant and child mortality are key indicators of health and nutritional standards of living. Combined with estimates of adult mortality, these rates produce life expectancies from birth which are an important indicator of overall living conditions. Of course, both child and adult mortality rates are particularly important measures when the population is experiencing an HIV/AIDS epidemic of the magnitude of that being experienced by South Africa.

Fertility has been covered by the first of the two commissioned monographs (Moultrie and Dorrington 2004). This report concentrates on estimating mortality rates in South Africa, nationally, by population group, and by province.

There has always been uncertainty about the level of mortality in South Africa. Under apartheid, life tables were produced initially only for the White population then later for the Coloured and Indian population (the so-called South African Life Tables (SALTs)), but never for the African population. Deaths of Africans were only fully included as part of the vital registration system from 1979, and then only for deaths within the boundaries of what the government of the time deemed to be the Republic of South Africa. (In other words excluding deaths which occurred in the 'homeland' areas designated as Transkei, Bophutatswana, Venda and Ciskei once they were deemed to be 'independent'.) Thus, and because there was assumed to be a low level of reporting of (or capturing of reported) deaths, no attempt was made to produce official life tables for the African population.

The first attempts to make use of reported death data were those by Dorrington (1989; 1998) and Timæus (1993). Prior to that, mortality of the African population was generally estimated as a by-product of attempts to model and reconcile population projections with census counts (for example Sadie (1970; 1970; 1973; 1988; 1993) and van Tonder, Mostert and Hofmeyr (1987)).

The last set of full life tables to be published officially were those for 1984-86 for Coloureds, Indians and Whites. Between 1991 and 1997, population group was not recorded on the death certificate and thus (and possibly for other reasons) no attempt was made to produce national life tables around the time of the census in 1991. However, Statistics South Africa did publish abridged life tables by population group for 1985-94, using the data on survival of children ever born and of parents from the 1996 census, and by province (but not population group) for 1996, using the vital registration data and the 1996 census count (Statistics South Africa 2000). Unfortunately, the implausible change in rates over time implied by these two tables as well as the fact that the level of completeness of reporting of deaths of around 70 per cent implied by the rates for 1996 was significantly lower than the 85 per cent reported by Dorrington, Bourne, Bradshaw, et al (2001) calls into question the reasonableness of these tables.

The 2001 census provides a useful opportunity to reconsider this question. This report thus interrogates the census results with a view to deciding which of the data are usable and, if so, in what form. In particular, it inspects the various edits performed on the data to see which are acceptable and which need to be rejected as corrupting the data. Ultimately the purpose of this research is to produce estimates of childhood and adult mortality for the country as a whole, and by population group and province.

Given the history of South Africa, it is quite understandable that some question the continued use of population group as a means of stratifying the population. However, apart from the fact that there is no better measure of previous socio-economic deprivation than population group, sound methodological reasons exist for preferring demographic analysis at this level. Essentially these are based on the observation that the population groups are demographically very different. Thus, ignoring population group introduces a level of heterogeneity in the population that not only clouds the interpretation of the results but may well violate the underlying assumptions of the methods used. Further, this heterogeneity together with the geographic grouping of population groups in South Africa allows for a more nuanced approach to estimating rates at a sub-regional level.

Chapter 2 of this report interrogates the data from the 2001 census and the edits performed thereon to estimate adult mortality rates. First, it considers the deaths reported by households to have occurred in the 12 months prior to the census. This is followed by an analysis of the data on survival of parents of those enumerated in the census. Chapter 3 describes the method used to produce estimates of adult mortality, first nationally and then by province. This includes a description of adjustments to the number of deaths from the population register for the period 1998 to 2001 and the sample cause of death data for the years 1999 to 2001 (Statistics South Africa 2002) required to produce estimates of deaths over the whole of the intercensal period, nationally and by population group. Chapter 4 presents the estimates thus derived. Chapter 5 considers in some detail the data on children ever born and children surviving from the census data and draws the disappointing conclusion that these data cannot be used to produce estimates of childhood mortality. Chapter 6 considers past trends in mortality for the population as a whole and for each of the population groups and compares our estimates with those produced by other demographers in recent years. Finally, Chapter 7 draws conclusions about the reasonableness of the results produced and highlights a number of areas worthy of further research.

2 Usefulness of the census data on adult mortality

It is common for census data to be edited, even where populations are literate and data are collected with care. The 2001 South African census was no exception. Such edits take place because, no matter how much care is taken, censuses are prone to two major types of errors, namely coverage and content errors. Coverage errors arise from duplication or omissions of persons in a household (or, indeed, entire households) during enumeration. Generally, as far as coverage is concerned, manual or computer-assisted checks can often identify duplication of data and post-enumeration surveys are used to estimate the extent of omissions (albeit imperfectly).

Content errors occur when recorded characteristics of households or persons are incorrect because of errors or omissions by the person completing the census questionnaire or because errors were introduced during data processing. Content errors may take the form of omission of information, the choice of invalid values or codes, inconsistent entries or values of unreasonable magnitude.

In addition, content and coverage errors can be either random or systematic. Systematic errors are likely to arise from poor instrument design (leading to confusing questions), poor enumerator training, inadequate fieldwork to ensure that all parts of the country are covered, or conscious misreporting of certain information (e.g. income, citizenship, etc) due to a lack of trust in the confidentiality of the census data, or poor quality scanning. Systematic errors are relatively controllable while random errors are not. Systematic errors can further be either deliberate or (more commonly) inadvertent and can occur during data capture, data coding, and sometimes during office editing and tabulation.

Editing is required to adjust for random errors. Systematic errors are harder to correct, particularly if the errors have arisen from misunderstandings of the questions asked. Generally, in such circumstances, editing can simply remove the random error components from the systematic error. By removing random error, editing makes the raw data more useful and accurate. Heavily edited data will tend to have less in common with the raw data, and - more specifically - may deviate to a greater extent from the data that does not require editing.

2.1 Editing rules: imputation and hotdecking

Imputation involves the replacement of a value or blank response during editing based on meaningful use of other information captured, or some other reasonable techniques for assigning values to a particular field. Two forms of imputation can be applied: changing responses on the grounds of logical consistency (termed 'logical imputation') and – where logic cannot resolve a value – a hotdeck may be employed. Both procedures (and particularly the use of hotdecks) are contentious as there is a real danger that the edit rules can introduce systematic biases of their own into the data. Evidence of this phenomenon can be found in the discussion of the editing of mothers' month of birth presented in the report on fertility (Moultrie and Dorrington 2004).

A cold deck is simply an imputation rule which is independent of other data collected in the census, specifying that, if a certain value is missing and logical resolution is not possible, then a given value must be attributed to the response based on certain, pre-specified characteristics of the person in question. The dynamic or hot deck technique is a variant of this method. The deck is updated when new or other data which are deemed to be valid response combinations are encountered. The missing values for a particular case are then copied from those of another case with valid values and similar other characteristics. This process was also applied in some cases where data were not missing but were deemed to be incorrect.

The validity of this process depends on how representative the cases in the hotdeck are of those with missing data. The size of the hot deck is important, as it has to be large enough to prevent repetition of values, but not so large as to become difficult to update efficiently. The single biggest concern regarding the use of hotdecks must be that heterogeneity in hotdecked responses must be removed as far as possible through specification of the relevant dimensions (i.e. background characteristics used to draw a response) of the hotdeck table. Thus, for example, if a response is known or strongly believed to vary by a certain background characteristic, that characteristic must form one of the dimensions of the hotdeck table. As we shall suggest, Statistics South Africa (together with their advisors from the US Bureau of the Census) could occasionally have chosen better characteristics for the specification of hotdeck tables than they did.

At a meta-level, logical edits are not necessarily neutral in their effect on the data. A simple example suffices: in editing responses on whether a person's mother is alive, the edit rules actively seek a person in the household who may be the person's mother, possibly editing the enumerated relationship (or other) variables to accommodate this. Only where no person can be found, is a hotdeck applied, which includes responses of both "mother alive" and "mother dead". As a consequence, the entire editing procedure is biased towards finding a live mother in the household, a distortion which will become clear once we interrogate the data in section 2.3.

Nevertheless, Statistics South Africa must be commended for making available to us census data consisting of both raw and edited variables together with an imputation flag indicating whether the response was not edited and, if it was, whether it was edited logically or by means of a hotdeck, and whether the response had been missing or not before editing.

2.2 Cleaning and editing process applied to mortality data collected at a household level

Figure 2.1 shows the questions directly relating to mortality that were asked of every household and which constitute the household mortality record. These questions were directed to the respondent in the household only and were not asked of people enumerated using the institutional questionnaires. This may result in the introduction of unforeseen biases into the data collected. In addition, there can be no self-reporting of mortality so all information is collected by proxy and could contain errors due to imperfect knowledge on the part of the respondent. Particular problems can result from the dissolution of households as a result of the death of a last surviving member, the reconstitution of households in the event of other deaths, extended households leading to some deaths being reported by more than one household, as well as the serious problems associated with the omission of deaths in institutions.

The questionnaire available on the Statistics South Africa website is not the final version, and contains several errors. Other than grammatical errors, the exact specification of the reference period is incorrect, running as it does from October 10 2000 to October 10 2001, and not October 9 2001 as it should be. However, we are assured (Pali Lehohla, Statistician General, personal communication) that the version of the questionnaire finally printed did not have this error.

Figure 2.1 Wording of questions on household mortality in the last 12 months, Census 2001

INFORMATION REGARDING THE HOUSEHOLD						
ANYBODY DIED (H-31)	DECEASED (H-31a)					
Has any member of this household died in the past 12 months, i.e. between 10 October 2000 and 10 October 2001? Y = Yes N = No Dot the appropriate box. Y N	(If YES to H-30) What was the first name of the deceased?	What was the month and year of death? Write the month and year of death. Month Year	What is the sex of the deceased? M = Male F = Female Dot the appropriate box.	What was the age in years at death? For example, if 2 years of age write 0 0 2	Did (the person) die from an accident or through violence? Y = Yes N = No Dot the appropriate box.	If the deceased was a woman under 50 years, did (the person) die while pregnant or within six weeks after delivery? Y = Yes N = No Dot the appropriate box.
If YES, how many?		M M Y Y Y Y	M F		Y N	Y N
		M M Y Y Y Y	M F		Y N	Y N
Go to H-31a.		M M Y Y Y Y	M F		Y N	Y N
If NO, the questionnaire is completed.		M M Y Y Y Y	M F		Y N	Y N
		M M Y Y Y Y	M F		Y N	Y N

As with other variables collected in the census, responses to the questions on month and year of death, age at death, and sex were subject to imputation if they were implausible or missing. Unfortunately, however, the data provided by Statistics South Africa does not contain the full raw (pre-edit) data[1] on this question, and so the effect of the edits cannot be fully assessed. Thus we will only be able to determine the extent of use of different editing rules, and the edited responses, but not the responses given that gave rise to the need for editing in the first place.

In aggregate, data on 368 377 deaths were captured and deemed useable in Census 2001. After allowance for the undercount identified in the post-enumeration survey, and applying the derived household weights to these data, they represent an estimated 448 312 deaths in South African households.

Imputation rates were relatively low. Table 2.1 shows the proportions of all household deaths that were subjected to imputation/editing for month and year of death information, and age of deceased. Sex of the deceased was imputed in less than 2 per cent of cases.

[1] Presumably this is because the raw data included 2.5 million bogus mortality records, thought to be the result of scanning of dirty back pages of the questionnaire.

Table 2.1 Percentage of data on household deaths subjected to editing, Census 2001

	Per cent (n = 448 312)	
	Unchanged	*Edited*
Month of death	96.9	3.1
Year of death	94.0	6.0
Age at death	92.2	7.8

These proportions might suggest that the edit rules used to clean the data in Census 2001 are undeserving of close scrutiny. In some senses, this is true: the edited results are unlikely to affect the overall distributions and conclusions to any significant extent. However, these proportions mask a great deal of heterogeneity, by population group, province and age (amongst others). In addition, it is worth reflecting on the editing procedures employed in Census 2001 to identify any obvious flaws, and to ascertain whether they should be retained in future censuses.

The following sections examine each of the three variables in Table 2.1 in greater detail.

2.2.1 Month of death

The first variable examined is the reported month of death. As mentioned above, 3.1 per cent of responses were not accepted and were subjected to editing procedures. In the case of this variable, logical editing procedures cannot be applied as there is no additional data against which responses can be validated and – where necessary – corrected. Thus, only hotdeck procedures could be applied to missing or invalid data. The distribution of editing procedures by population group[2] is shown in Table 2.2.

Table 2.2 Distribution of editing procedures applied to household month of death variable, by population group, Census 2001

		No imputation	*Hotdeck (missing)*	*Hotdeck (non-missing)*	***Total***
Africans	N	379,332	9,793	2,700	**391,825**
	Per cent	96.8	2.5	0.7	**100**
Coloureds	N	25,443	608	214	**26,265**
	Per cent	96.9	2.3	0.8	**100**
Indians	N	7,149	98	79	**7,326**
	Per cent	97.6	1.3	1.1	**100**
Whites	N	22,315	383	203	**22,901**
	Per cent	97.4	1.7	0.9	**100**
Total	**N**	**434,239**	**10,882**	**3,196**	**448,317**
	Per cent	**96.9**	**2.4**	**0.7**	**100**

Note: Totals may not add to those provided earlier due to rounding errors in CS-Pro.

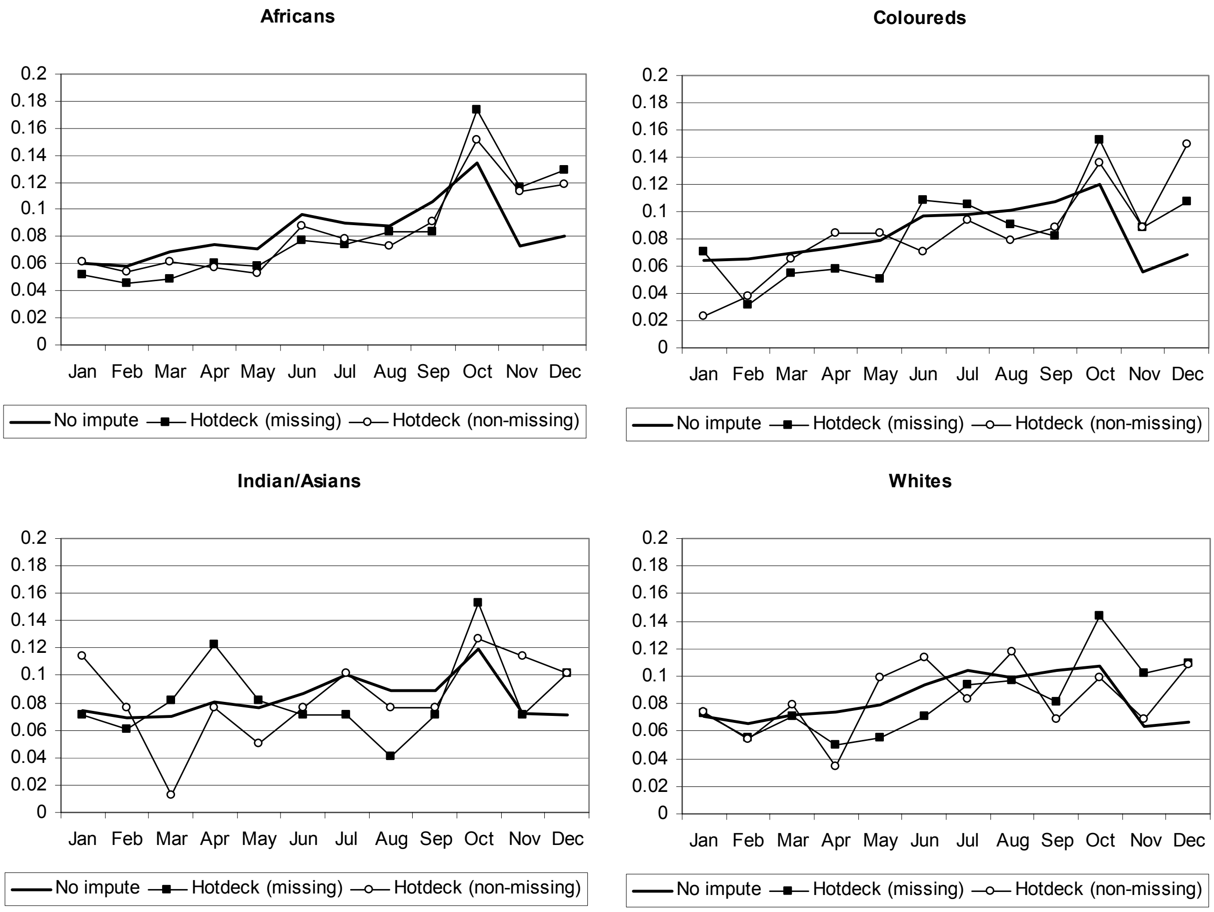
The data for Africans and Coloureds are marginally worse than those for Indians or Whites. Judging by the similarities in proportions hotdecked from non-missing data, this would appear to be a function of fewer missing values in these two groups.

Because information on day of death was not collected, the edit rules specified take a blunt approach to correcting data on month of death by assuming that all deaths after 9 October 2001

[2] In this section, when distributions of reported deaths in the household are analysed according to population group, this is done so by the population group of the majority of household members, a derived variable prepared by Statistics SA. We believe that this gives a better picture of the composition of the household than simply using the population group of the household head, although the difference is negligible.

and all those before 10 October 2000 were not reported by households. This is plainly not the case as can be seen from the distribution of the edited months of reported deaths in Figure 2.2.

Figure 2.2 Distribution of (edited) months of reported death by population group, Census 2001



Several features are apparent in these data. First, there is a spike in the (unedited) deaths reported as occurring in October. It is highest among Africans and Coloureds, with virtually no peak for Whites. Second, the hotdecked values correspond poorly with the unedited data for all population groups with the possible exception of Africans. Third, the hotdeck itself adds a peak to the data for October, this being no doubt in part due to the decision to accept October as a valid month of death in both 2000 and 2001. Fourth, the unedited data appears to have a bias against reporting deaths in November and December, something which is overcompensated for by the hotdeck.

In any event, there can be no doubt that the hotdecked values agree poorly with the unedited data, and certainly evince no pattern that could reasonably be deemed to be the roughly uniform distribution one would expect.

Thus we are not convinced that these edits improve the quality of the data.

2.2.2 Year of death

The second variable examined is the reported year of death. These data are significantly worse than those relating to month of death as discussed in the previous section. A year of death had to be imputed in 6 per cent of reported deaths. In part, this error is understandable since, with the exception of deaths occurring in October and the use of a reference period covering a year, respondents and enumerators may have felt that the year of death was self-evident and hence neglected to fill in an answer. A further consequence of this is that logical editing is possible, since valid responses to the question on month of death can help determine the year of death. By far most edits were based on logical grounds, when there was a response given. However, this editing rule contains an obvious bias.

Table 2.3 Distribution of editing procedures applied to household year of death variable, by population group, Census 2001

		No imputation	*Logical (missing)*	*Logical (non-missing)*	*Hotdeck (missing)*	*Hotdeck (non-missing)*	***Total***
Africans	n	367,828	1,537	18,025	1,952	2,485	**391,827**
	per cent	93.9	0.4	4.6	0.5	0.6	**100**
Coloureds	n	25,265	79	705	126	84	**26,259**
	per cent	96.2	0.3	2.7	0.5	0.3	**100**
Indians	n	6,927	32	317	25	26	**7,327**
	per cent	94.5	0.4	4.3	0.3	0.4	**100**
Whites	n	21,533	103	1,031	129	102	**22,898**
	per cent	94.0	0.4	4.5	0.6	0.4	**100**
Total	**n**	**421,553**	**1,751**	**20,078**	**2,232**	**2,697**	**448,311**
	per cent	**94.0**	**0.4**	**4.5**	**0.5**	**0.6**	**100**

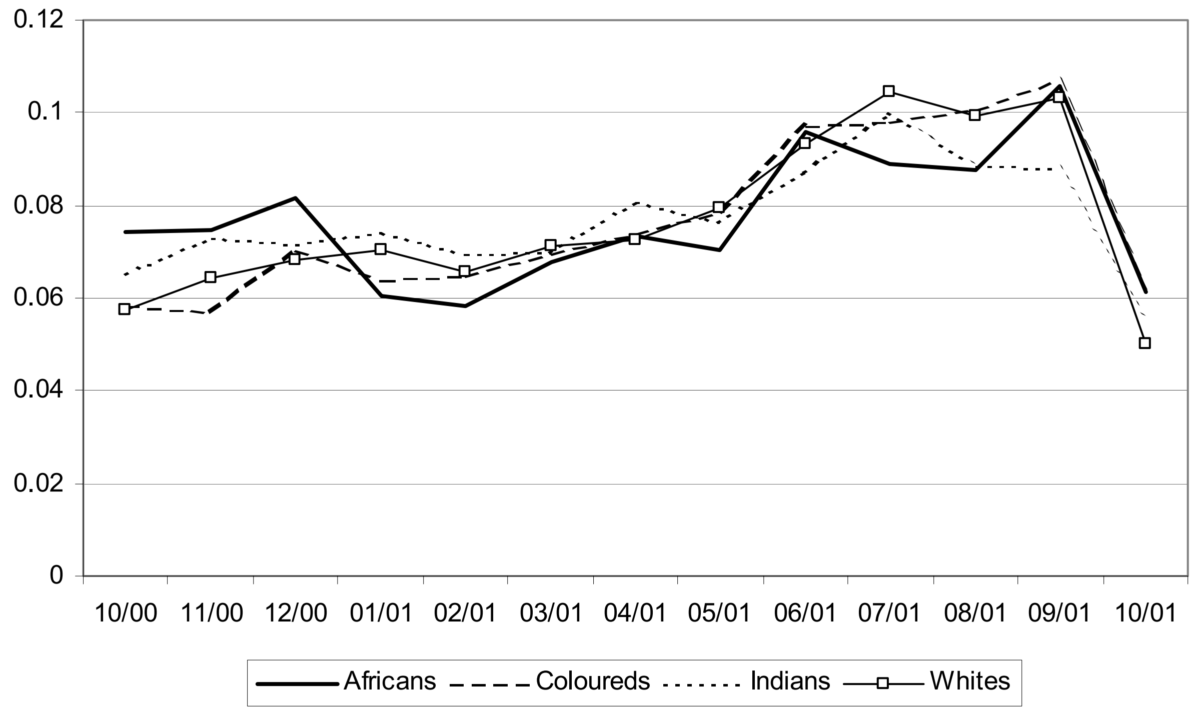
Note: Totals may not add to those provided earlier due to rounding errors in CS-Pro.

If a valid month of January to September is reported, but the year is invalid, the year of death is automatically recoded to 2001. Likewise, valid months of death from October to December are given a year of death of 2000. Thus no reference period error is assumed to exist. Since the raw variables for this section of the questionnaire have not been made available, the extent of the reference period error cannot be determined. In any event, these results are astonishing for their completeness. Women, for whom their last-born child was reported to have been born in 2000 or 2001, had imputation/hot-decking performed in more than a quarter of all cases. Here, household respondents, "reporting" recent deaths in the household only require imputation of year of death in less than 6 per cent of cases. Thus, despite the fact that one would expect mothers to be far more aware of their recent births than one would heads to recall deaths in the household, the level of imputation is more than four times less.

For some reason, imputation among Coloureds was notably less than among other population groups, which were similar in their extent of imputation. Finally, this rule is inconsistent with the rules for editing month of death (which made allowance for valid deaths in both October 2000 and October 2001) by deeming all reported October deaths, but without a valid year, to have occurred in 2000.

One can investigate the reported month and year of death variables further by examining the distribution of the edited data by both month and year. The results, by population group, are presented in Figure 2.3. The distribution of edited months and years of deaths show only slightly fewer deaths in October 2001 than those reported in October 2000. Given that the reference point was approximately one-third of the way through the month, one would anticipate that the number of deaths reported in October 2001 would be roughly half of that reported in October 2000.

Figure 2.3 Distribution of edited months and years of reported deaths in the household, by population group, Census 2001



While there is a clear drop-off in the reported numbers (and hence distribution) of deaths recorded as October 2001, the fall should have been much bigger. The temporal distribution of deaths for all population groups followed similar trends, increasing from January through September 2001. This does not correspond either with any supposed seasonality of deaths, nor with a roughly uniform distribution of deaths over time.

2.2.3 Age at death

Our concerns about the misreporting and possible biases in the reporting of dates of death are important for assessment the overall quality of the edits, but are not hugely significant given the methodology we have employed to estimate mortality levels from these data. Of far greater import, however, is the quality of the data on the age of the respondent at death, as these data bear directly on the estimation of the age pattern of mortality and, inevitably, on speculation about the scale of AIDS deaths in South Africa.

As shown in Table 2.1, these data were the worst collected of the three variables assessed so far. The age of the deceased had to be imputed in nearly 8 per cent of cases, and – given the absence of other confirmatory variables – this imputation had to be achieved solely through the use of a hotdeck. The extent of hotdeck usage by reason for use and population group is shown in Table 2.4. Implausible reported ages at death requiring the application of a hotdeck to non-missing data were found only in a handful of cases. These data are too scanty to support further analysis. Age at death was left blank in almost 5 per cent of deaths in (predominantly) White and Indian households, and in over 8 per cent in predominantly African households. These values were drawn from the same hotdeck used to resolve month and year of death.

Table 2.4 Distribution of editing procedures applied to age of deceased variable, by population group, Census 2001

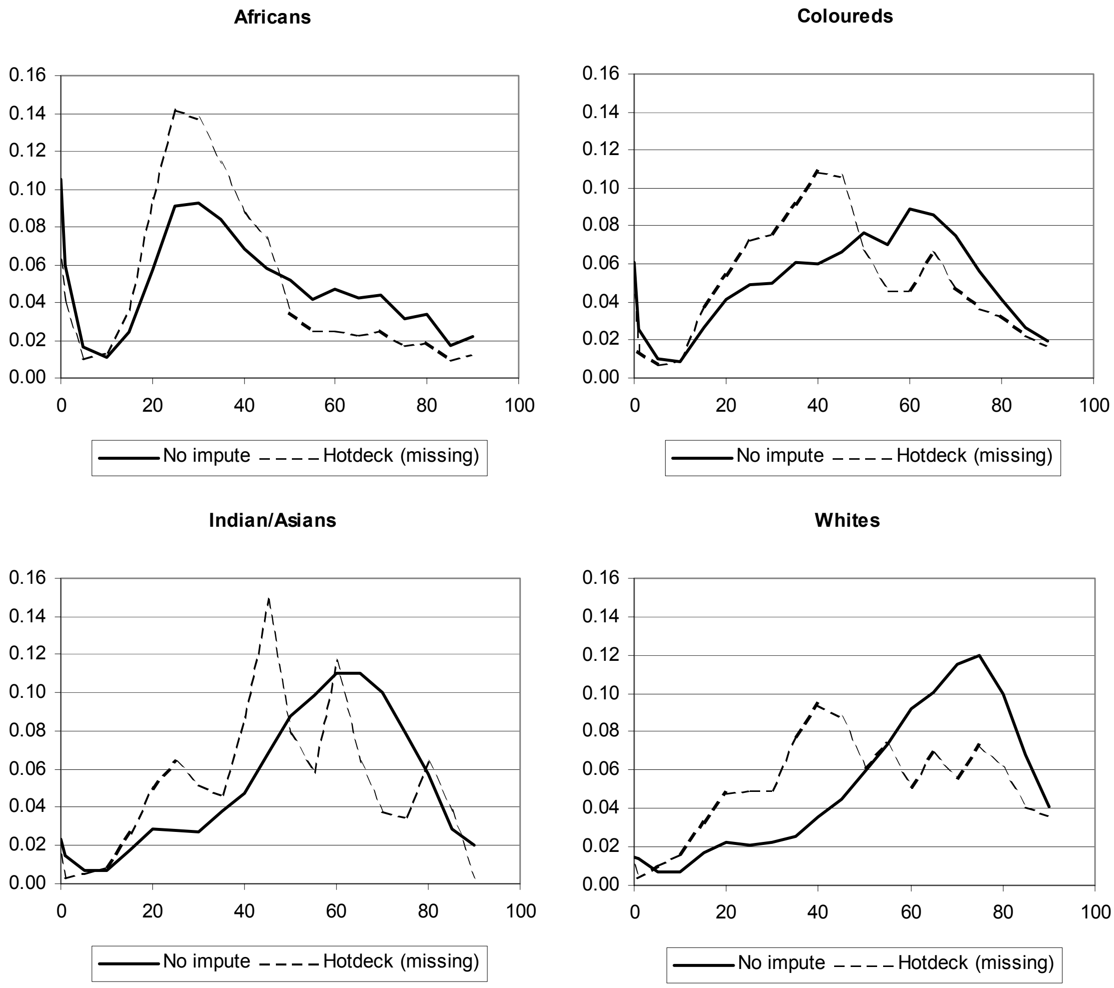
		No imputation	*Hotdeck (missing)*	*Hotdeck (non-missing)*	***Total***
Africans	n	359,855	31,868	107	**391,830**
	per cent	91.8	8.1	0.0	**100**
Coloureds	n	24,652	1,606	2	**26,260**
	per cent	93.9	6.1	0.0	**100**
Indians	n	6,981	344	1	**7,326**
	per cent	95.3	4.7	0.0	**100**
Whites	n	21,806	1,091	3	**22,900**
	per cent	95.2	4.8	0.0	**100**
Total	**n**	**413,294**	**34,909**	**113**	**448,316**
	per cent	**92.2**	**7.8**	**0.0**	**100**

Note: Totals may not add to those provided earlier due to rounding errors in CS-Pro.

The age distributions of the unedited data and the data drawn from the hotdeck in respect of missing values are shown in Figure 2.4.

Clearly something has gone awry with the construction and use of the hotdeck used to fill in missing values for the age of the deceased, as the age distribution does not match that observed in the data for any of the population groups, and tends to exaggerate death below age 40. Thus, for example, while the age distribution of unimputed deaths for the White population follows a pattern consistent with that observed in developed countries, the age distribution of the hotdecked data has a peak at age 40, and is roughly constant till age 80. For Africans, on the other hand, the hotdecked distribution is heavily weighted towards deaths among young adults, in a pattern strongly redolent of AIDS-associated mortality. While this pattern of mortality may well hold, artificial data of the kind produced by the hotdeck does not lead to rigorous or robust estimates of mortality, and we recommend, therefore, that these edits are discarded in their entirety.

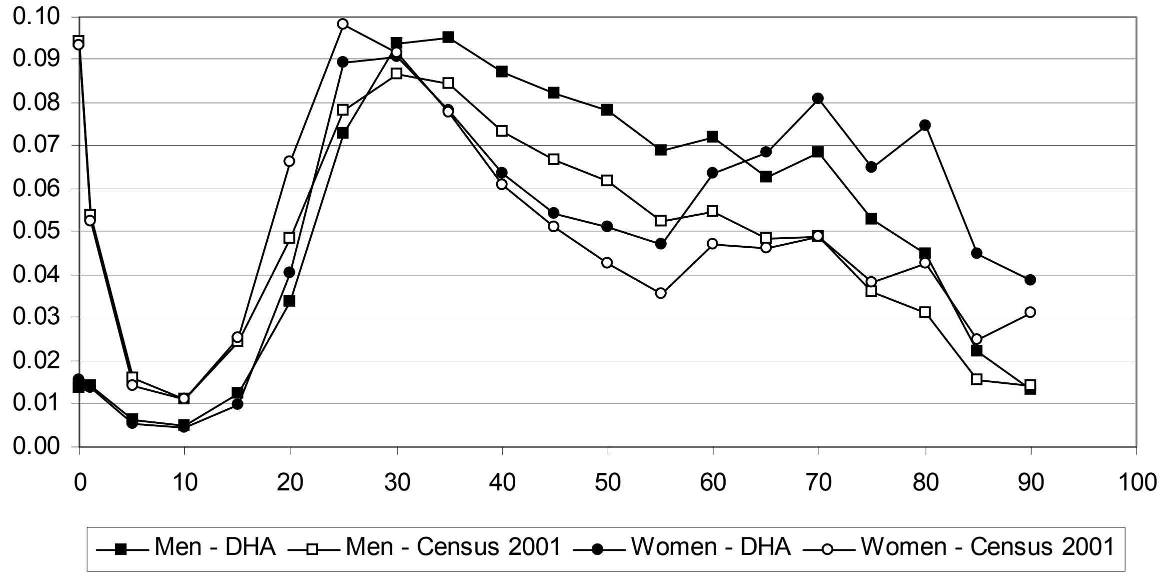
Figure 2.4 Age distributions of unimputed and missing ages imputed from a hotdeck by population group, Census 2001



The census data collected on adult deaths in the household can be compared with the numbers of deaths occurring in 2001 recorded on the population register maintained by the Department of Home Affairs (DoHA). Infant deaths, particularly, as well as those of children and young adults are less comparable because these deaths are unlikely to be fully recorded on the population register as the deceased are significantly less likely to have an identity number. In addition, the population group of the deceased is not captured on the population register, and hence the Home Affairs data cannot be stratified by population group.

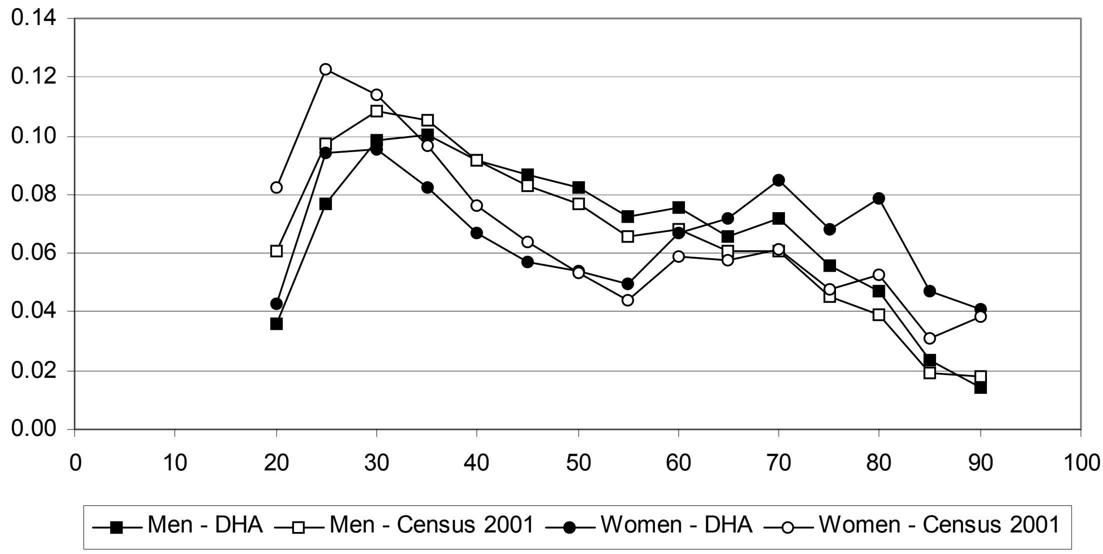
The possible magnitude of the non-reporting of infant deaths to the Department of Home Affairs is evident from Figure 2.5. According to the Home Affairs data, less than 2 per cent of all deaths occurring in 2001 were children under the age 1. By contrast, the census data suggest that infant deaths account for almost 10 per cent of deaths in the country.

Figure 2.5 Distribution of deaths by age, Census 2001 and Home Affairs deaths for 2001



If we examine the distribution of deaths of adults over the age of 20 (Figure 2.6), a more consistent indication of reporting of deaths to Home Affairs and in the census emerges, particularly for men. The data for women from the two data sources do not correspond at all well between ages 60 and 85, while the census data for both men and women suggest much higher numbers of young adult deaths in the census. Interestingly, no strong evidence appears in the census data to indicate that dissolution of households on the death of the oldest surviving inhabitant is a major force on the reported distributions. It may, however, be the case that dissolution of households does not occur on death (but possibly earlier – for example, on relocation to a retirement home, not covered by the questionnaire on household deaths), or does not occur to any great degree among African South Africans (and hence is not discernible in the national data presented here).

Figure 2.6 Distribution of deaths over age 20 by age, Census 2001 and Home Affairs deaths for 2001



2.2.4 Accidental deaths

In addition to the date of death and age of the deceased, the census questionnaire asked respondents whether the death was a consequence of an accident or violence. Again, no further data are available to corroborate the response given, and hence logical editing rules could not be applied. In addition, because respondents were only given a choice of yes or no, no editing using a hotdeck applied to non-missing (valid) values was required. The data, by population group of the majority of the household, are summarised in Table 2.5.

Table 2.5 Distribution of editing procedures applied to accidental death variable, by population group, Census 2001

		No imputation	*Hotdeck (missing)*	***Total***
Africans	n	371,158	20,670	**391,828**
	per cent	94.7	5.3	**100**
Coloureds	n	24,974	1,285	**26,259**
	per cent	95.1	4.9	**100**
Indians	n	6,984	342	**7,326**
	per cent	95.3	4.7	**100**
Whites	n	21,896	1,003	**22,899**
	per cent	95.6	4.4	**100**
Total	**n**	**425,012**	**23,300**	**448,312**
	per cent	**94.8**	**5.2**	**100**

Note: Totals may not add to those provided earlier due to rounding errors in CS-Pro.

Responses hotdecked from missing data were slightly more common among Africans than among other population groups. Also, with the exception of Whites, those subject to hotdecking were more likely to be attributed a violent or accidental death than those not subjected to hotdecking (Table 2.6).

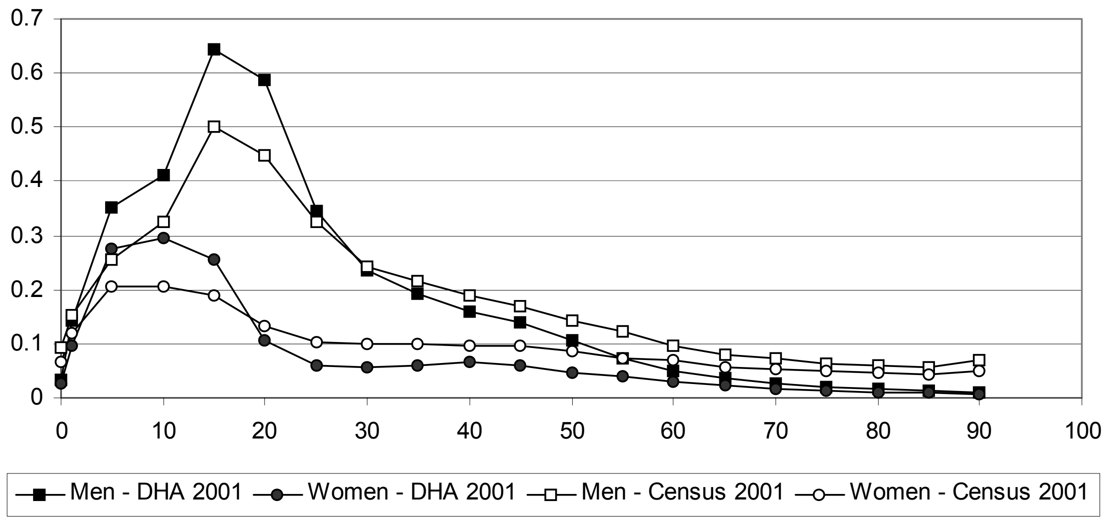
Table 2.6 Accidental or violent deaths by imputation status and population group, Census 2001

	No imputation – per cent	*Hotdeck (missing) – per cent*	***Total***
Africans	13.69	15.82	**13.80**
Coloureds	16.19	18.60	**16.31**
Indians	12.84	17.54	**13.06**
Whites	13.10	11.57	**13.04**
Total	**13.79**	**15.82**	**13.90**

Note: Totals may not add to those provided earlier due to rounding errors in CS-Pro.

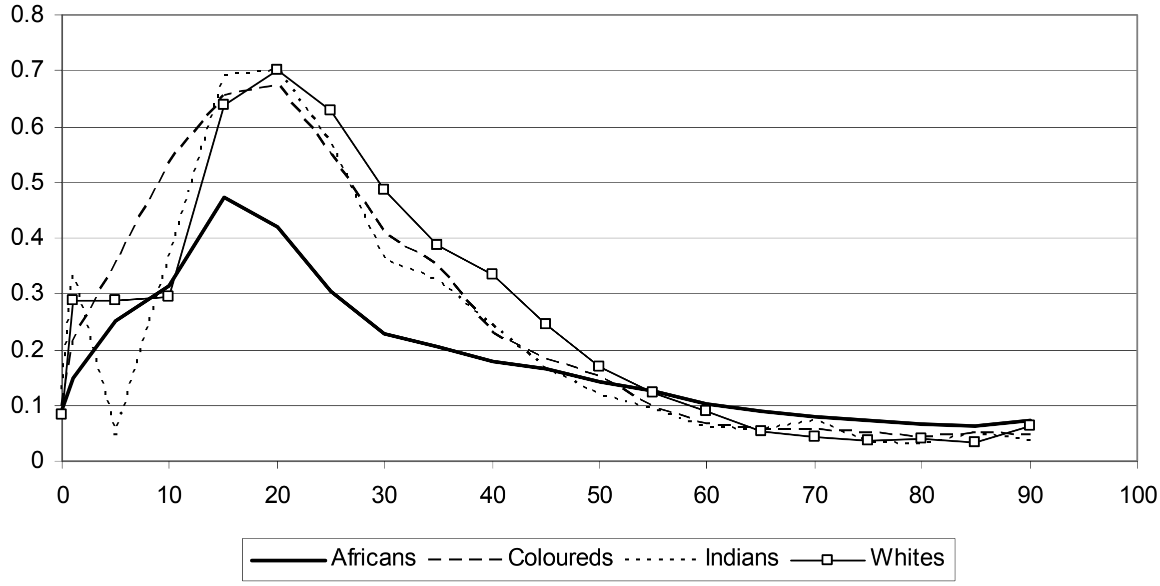
No doubt this pattern reflects in part the different age distributions across the population groups shown in Figure 2.4. The further dynamics of accidental death in South Africa are explored below. Compared with the data collected by Home Affairs, the census reports higher proportions of accidental deaths after age 35 for men and age 25 for women than is indicated by the Home Affairs data. At younger ages, by contrast, the census indicates fewer accidental deaths than indicated by the Home Affairs data. This discrepancy, however, is not altogether unexpected as the Home Affairs data are based on medical certification, whereas the census data are reported by household members of the deceased. All things considered, the data are remarkably consistent.

Figure 2.7 Proportion of deaths due to accidental causes, Census 2001 (edited data) and 2001 Home Affairs data



Analysis of cause of death is complicated by the fact that causes are competing: higher risks of dying from infectious or other natural causes reduce the proportions of death attributable to accidents. Where those other risks of natural death are low, the proportion of deaths attributable to accidents is high: thus, accidental deaths are highest between age 20 and 24 among White, Coloured and Asian men, accounting for almost seven out of every ten deaths in that age group, whereas accidental deaths are highest for African men aged 15 to 19, and account for less than half of all deaths in this sub-population.

Figure 2.8 Proportion of male deaths due to accidental or violent causes, by age group and population group, Census 2001 edited data

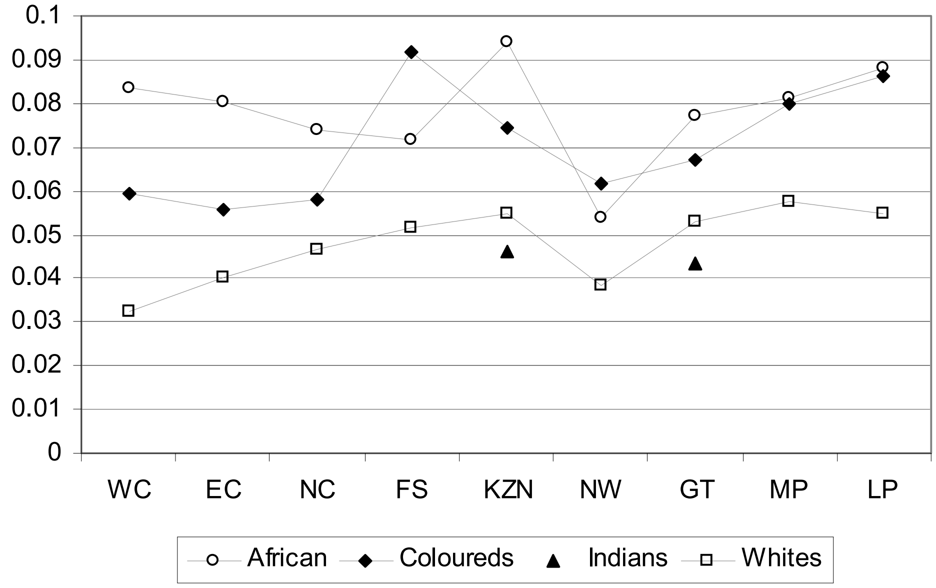


A similar, if more volatile, pattern is observed among female accidental deaths, although at their peak, accidents or violent deaths among women account for less than half the total number of deaths.

2.2.5 Household deaths by province

As our major purpose for these data is to use them to apportion an estimate of the number of deaths at a national level to the provinces by population group, sex and age, a major determinant of the usability (or otherwise) of these data is the plausibility and consistency of provincial distributions of mortality by these divisions. Figure 2.9 shows that the proportion of ages at death subjected to hotdecking is not constant by province, with the data from the North-West province being consistently less subjected to imputation than data from other provinces. Similar distributions (also showing clearly less imputation in the North-West) can be observed in respect of year and month of death. Furthermore, since we have already shown that the imputation by age contains significant biases and have demonstrated here that the proportion subjected to that imputation is not at all uniform across the provinces, we must conclude that the errors introduced by the imputed data will be more significant in some provinces than others.

Figure 2.9 Proportion of age of death imputed by population group and province, Census 2001



2.2.6 Conclusions: Deaths in the household

Typically where this question has been used it has not performed very well (for example, although this question was asked in the 1997 census in Swaziland it was not reported on since the data were felt to be of little use). Hence, users of these data must bear in mind that questions of this nature are relatively untried and untested in censuses globally, let alone those in developing countries. In the strongest possible terms, we caution against using these data to estimate directly current levels of mortality in South Africa. What we will try to do with these data is to check them against data from other sources (in particular vital registration systems), and, if consistent, gross them up to provide an indication of the extent of mortality in the country. With care, these data may be able to shed additional light on the recent patterns of South African mortality. Used carelessly, our concern is that these data could be used to demonstrate almost anything.

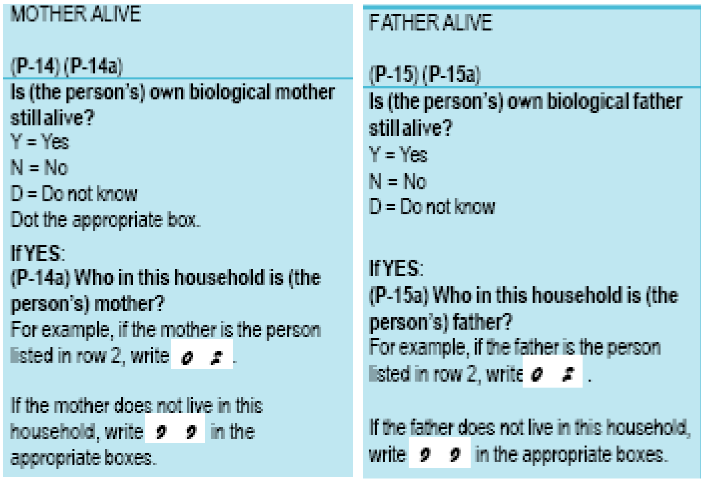
Thus, our conclusion is that these are specialist data and are of little use for estimating mortality directly. As such we recommend that the data be released only as part of the 10 per cent sample without the edits considered above. However, because we propose to use these data primarily to apportion the number of deaths nationally to the provinces and because the level of editing differs by province, we have elected to include all deaths reported by the households, but to assume that the data in need of editing were the same in all pertinent respects as the data that did not need editing.

2.3 Questions on the survival of biological parents

Questions on the survival of enumerated persons' parents (both mothers and fathers) have been asked routinely in developing country censuses ever since William Brass demonstrated that the proportions of mothers and fathers surviving by age of enumerated (together with additional information on the mean age at childbearing) could be used to derive life table measures of adult mortality. Questions on the survival of parents were asked about all household members in the 2001 census for this reason.

The questions asked relating to the survival of mothers are shown in Figure 2.10. Questions on survival status of biological father were phrased in exactly the same way as those for mothers.

Figure 2.10 Questions asked regarding maternal and paternal survival, Census 2001


MOTHER ALIVE

(P-14) (P-14a)

Is (the person's) own biological mother still alive?
Y = Yes
N = No
D = Do not know
Dot the appropriate box.

If YES:
(P-14a) Who in this household is (the person's) mother?
For example, if the mother is the person listed in row 2, write 0 2.

If the mother does not live in this household, write 9 9 in the appropriate boxes.

FATHER ALIVE

(P-15) (P-15a)

Is (the person's) own biological father still alive?
Y = Yes
N = No
D = Do not know

If YES:
(P-15a) Who in this household is (the person's) father?
For example, if the father is the person listed in row 2, write 0 2.

If the father does not live in this household, write 9 9 in the appropriate boxes.

As can be seen, respondents were offered an option of "Don't Know" to the question. However, the editing rules applied did not regard a response of "Don't Know" as valid – even though it is obvious that some respondents may truly be unaware of the survival (or otherwise) of their parents, and thus this is a valid response. Where a yes or no answer to the question of parental survival had been given the response was not subjected to further editing. Where no response was given, or where the respondent answered that they did not know about parental survival, both logical and hotdeck imputation procedures were used.

Given the phenomenon of paternal non-involvement, *a priori* we would expect non-response or "don't know" to apply particularly to enumerated persons' knowledge of whether or not their fathers were alive. This is indeed the case as can be seen from Table 2.7. Almost twice as many respondents of all population groups did not know the survival status of their father as they did that of their mother. Though it is not shown here, it is usually the case that women are (but only slightly) more aware of their parents' vital statuses than men.

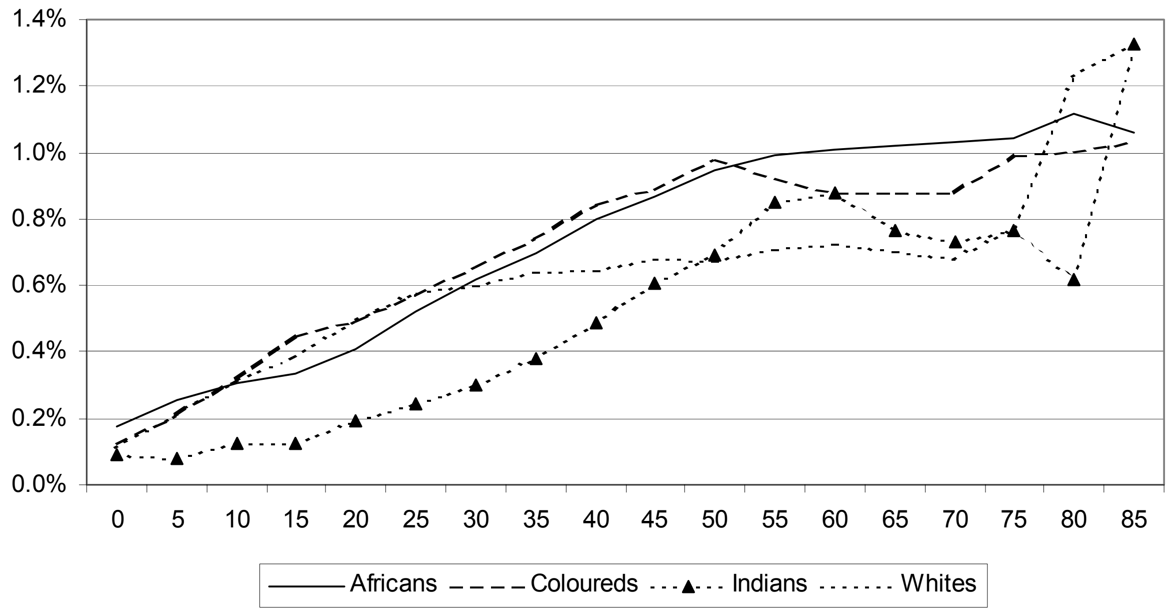
Table 2.7 Per cent responses to questions on parental survival subjected to editing by population group and parent, Census 2001

	African	*Coloured*	*Indian/Asian*	*White*
Mother alive - Don't Know	0.50	0.55	0.35	0.56
Mother alive - Missing	1.31	1.43	1.28	1.68
Total maternal imputation	1.81	1.97	1.63	2.24
Father alive - Don't Know	2.25	1.36	0.55	0.92
Father alive - Missing	1.27	1.53	1.32	1.68
Total paternal imputation	3.52	2.89	1.88	2.60
Ratio: Don't Know Father: Don't Know Mother	4.50	2.49	1.60	1.65

More significantly, Africans were 4.5 times more likely not to know the survival status of their father relative to that of their mother. For Coloureds, the figure was 2.5 times, while Indians and Whites were approximately 60 per cent more likely not to know the vital status of their father than that of their mother. Also significantly, when reporting on their parents' vital status, more respondents left the response blank than they replied that they did not know, with the single exception of Africans reporting on their fathers' survival.

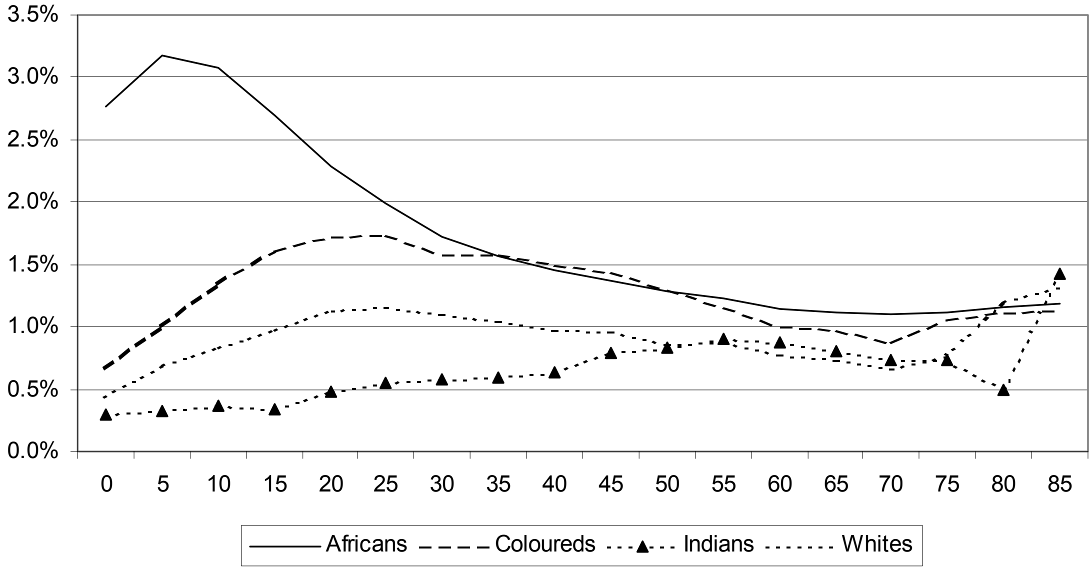
The proportion of respondents for whom data on maternal survival had to be edited is almost constant across all age groups. This, however, masks an interesting trend in that the proportion of "Don't Know" responses increases with age for all population groups up to age 60 and then levels off (Figure 2.11). A very different pattern is apparent in the responses on paternal vital status.

Figure 2.11 Proportion answering "Don't know" by age and population group: Mother alive?, Census 2001



Given the proportions indicated in Table 2.7, the effect of editing these data on the overall proportions will be trivial. Nevertheless, we recommend the rejection of these edits, both logical and otherwise, on several grounds.

Figure 2.12 Proportion answering "Don't know" by age and population group: Father alive?, Census 2001



First, the edited responses created using a hotdeck should be rejected, since the hotdeck on which they are based does not include population group as a stratification variable. By excluding population group, the implicit assumption is made that parental survival (controlling for the respondents' age, amongst other things) is independent of population group. We know that this is not the case.

Second, the logical edits should be rejected since they contain a subtle bias of their own. This bias arises from the fact that the edits seek – if none is identified by the respondent – a person in the household who might plausibly be the respondent's parent. If a suitable candidate can be found, a parent is attributed to the respondent. Thus, the edit rules are predisposed to identifying respondents' parents, and therefore potentially seriously understating adult mortality.

Third, comparison of patterns of imputed and non-imputed data suggest that the edit rules used to apportion responses of "Don't Know" between the two permissible parental survival states are severely biased.

Figure 2.13 (mothers) and Figure 2.14 (fathers) show the proportions of parental survival as indicated by the raw and the edited data, as well as by those data edited after being initially recorded as "Don't Know" or left blank. As would be expected, there is little substantive difference between the raw and the edited data at older ages. However, the difference between the raw and the edited data widens at younger ages, where – as a possible consequence of the positive bias referred to in the preceding paragraph – the edited data indicates much higher rates of parental survival. This is particularly the case for estimates of paternal survival.

Figure 2.13 Proportion of respondents' mother alive: raw and edited data for Africans by age, Census 2001

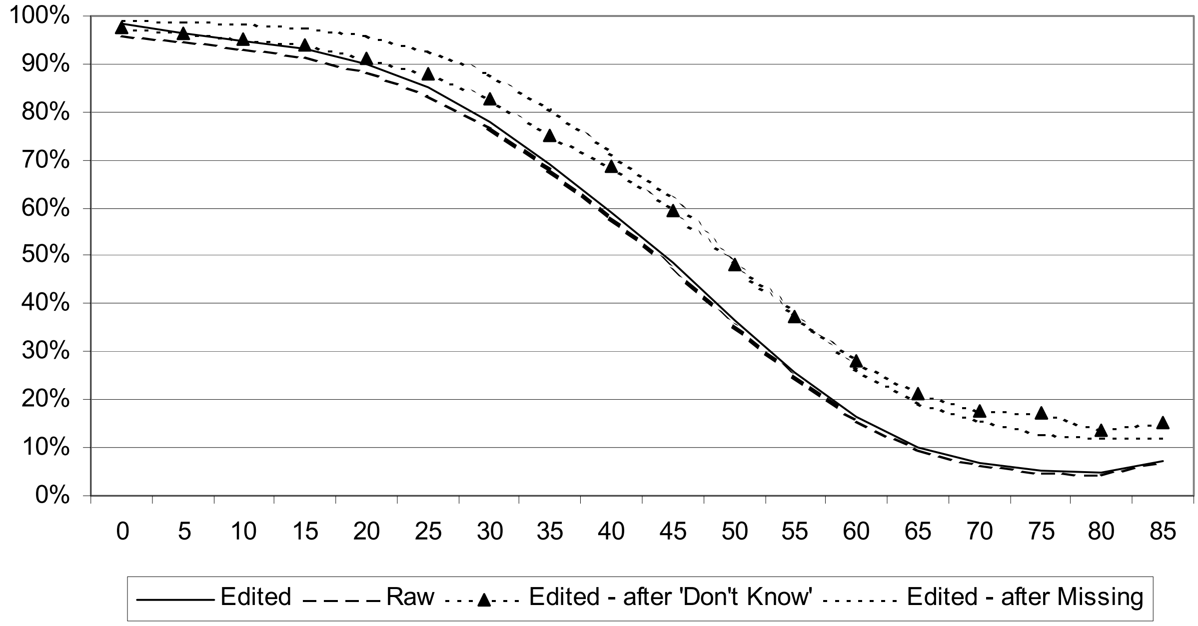
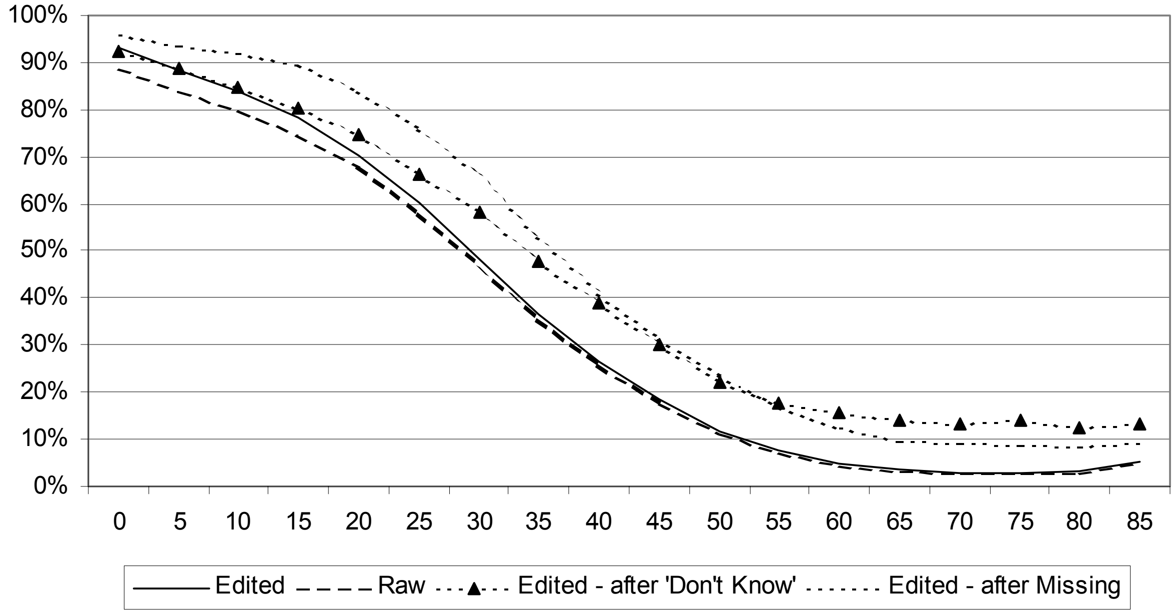


Figure 2.14 Proportion of respondents' father alive: raw and edited data for Africans by age, Census 2001



Far more tellingly, however, are the patterns of parental survival indicated by the data edited from "don't know" or missing responses. Among both maternal and paternal reports, the proportion of parents imputed to be alive at extreme old ages is implausible. The rules applied are patently wrong as they suggest that around 15 per cent of the parents of octogenarians are themselves alive. Thus, while the effect of these edits may be inconsequential, especially at the higher ages, it does not make sense to accept edits that produce such silly results.

A fourth reason to exclude the edits applied to these variables lies in the fact that these edits are heavily dependent on respondents' reporting of mother and father person numbers (MPNs and FPNs). A potential problem with this variable arises in the South African census, because the term "mother" is used in household relationships, while "biological mother" is specified under the questions on survival of parents. It is well known that the term "mother" is used very loosely by many communities, and can refer to grandmothers, foster parents or another woman who plays a mothering role, especially when the biological mother is absent. Unless clearly explained during data collection, this distinction may cause some error. Statistics South Africa themselves have pointed out the problematic nature of these two variables, and we see no reason to disagree with their conclusions (Statistics South Africa 2003).

For these reasons, we recommend the excision of all edits from these variables. In particular, it must be recognised that "Don't know" is a valid response to this question and these responses should not be edited out.

2.4 Conclusions

This section has interrogated the data on household as well as parental mortality collected in the 2001 South Africa Census. After thorough investigation, our conclusions are the same in every case, namely that the edit rules applied to the data by Statistics South Africa assisted by the US Bureau of the Census distort the results more than they add to the quality of the data.

Poor enumerator training is surely to blame for many of the errors, but perhaps it is appropriate for us to raise another concern here. A decision was taken in the planning of the 2001 census to use extensive editing to ensure that there were no missing data. It is possible that awareness of the editing procedures gave some in the organisation a false sense of confidence that editing rules could compensate for poorly collected data. If so, hopefully the interrogation of the data in this report (and in the report on fertility (Moultrie and Dorrington 2004)) will dissuade these people of this notion in future.

3 Method for estimating adult mortality rates

Since the data on deaths in the household from the census are problematic in a number of respects which prevent one being able to estimate mortality rates even using indirect techniques (in particular because it is clear that the percentage of all deaths that are reported cannot be assumed to be constant for all ages, as is necessary to assume for these methods) we have to resort to alternative data to produce rates at a national level. However, these data in turn do not lend themselves to being used to produce rates at a provincial level (since deaths are not necessarily registered in the province of residence). To overcome these problems, we have devised a method to make use of both sources together to produce province-specific estimates.

3.1 Overview

Before describing the method in detail it will be useful to give an overview of the strategy lest the reader gets bogged down in the fine detail of the many adjustments necessary to create usable data and meaningful outputs.

Essentially the strategy comprises four steps. The first is to estimate the adult mortality rates nationally for each of the sexes and then for each of the population groups for each of the sexes by applying the Generalised Growth Balance method (Hill 1987) to estimates of the population from the 1996 and 2001 census and the registered deaths.

The next step is to use the data on survival of parents from the censuses to estimate the level of mortality for these sub-divisions of the population at sufficiently high ages as to avoid having to find a standard table that allows for HIV/AIDS (say, ${}_{15}q_{50}$) These estimates are then compared with similar estimates derived from the rates produced from the registered deaths. If these agree, we can accept the estimated rates; if they do not, further analysis is necessary to decide on the best estimate of the level of mortality.

The third step is to estimate the number of deaths that occurred in the year preceding the 2001 census corrected for under-registration. These are then divided by the number of deaths in households reported in the census to have occurred in the 12 months prior to the census in 2001 to provide adjustment factors by population group, sex and age group. These adjustment factors indicate the extent of under-reporting of deaths by respondents in the household questionnaire.

The final step is to apply these adjustment factors to the household deaths from the 2001 census by province to produce adult mortality by province, population group, sex and age. Where the data are considered too scanty to produce reliably different mortality rates, provinces will be grouped together by population group to provide more robust estimates.

As a final check, we apply the orphanhood method first proposed by Brass (Brass and Hill 1973) to the provincial data from the census and check the rates of survival thus produced against those produced by our method.

3.2 Method in detail

3.2.1 Estimation of adult mortality nationally, by population group and sex

3.2.1.1 Registered deaths

The data used to adjust the reported household deaths from the census for under- (or over-) reporting are taken from two sources. The first, numbers of deaths from the population register maintained by the Department of Home Affairs, were supplied by Ria Laubscher, of the Medical Research Council (MRC) Rapid Mortality Surveillance project. These data cover the number of deaths by sex and age for each of the calendar years 1998 to 2001 inclusive. Population group is not captured as part of the population register.

The second source comprises deaths from the cause of death sample (of all deaths for which a death certificate was issued) drawn by Statistics SA for the calendar years 1997-2001 inclusive (Statistics South Africa 2002). These deaths are grouped by sex, age and population group (although in about 20 per cent of the cases no population group was specified on the form). For a number of years, population group was not captured as part of the death certificate, but with the introduction of a new form of death certificate during 1998 it was reintroduced as an item on the confidential form. Thus, we only have information on deaths by population group from this sample for the years 1999, 2000 and 2001.

3.2.1.2 Estimating the total reported deaths

The deaths from the Statistics SA sample were matched first against those from the population register to provide an estimate of the percentage of deaths with death certificates which do not appear on the population register (because they didn't have identification numbers or birth certificates). These estimates, derived for each age group and sex for each of the four years 1998 to 2001, were then used to derive an estimate of the numbers of reported deaths in each of these years from the numbers of deaths registered on the population register.

Next the numbers of reported deaths for each of the five years between the 1996 and 2001 censuses were estimated by:

- projecting backwards the numbers by age and sex for the calendar years 1998-2001 to give the numbers for the 1996 and 1997 calendar years
- apportioning the deaths in calendar years to each of the five intercensal years starting from 10 October 1996.

The obvious approach for estimating the number of deaths in the earlier years, where one can assume that the mortality rate is not changing very rapidly, is to calculate, for each age and sex, an average annual number of deaths over the years 1998-2001 and then back project that number using the annual population growth rate between the censuses for that age group and sex. This is the most commonly adopted approach. However, where there is reason to believe that mortality rates have not remained roughly constant over time, such as in certain age groups due to the AIDS epidemic, then a better approach is to back-project the number of deaths using the trend in the number of observed deaths over time, in this case, fitting an exponential trend to the number of deaths occurring between 1998 and 2001.

However, in order to avoid extrapolating spurious trends, we confined this approach to ages 25-64 for males and 20-59 for females, and applied it only where there was a strong correlation (R^2>0.85) between the data and the fitted line.

To estimate the number of deaths reported in each of the population groups, the national deaths for the calendar years 1999, 2000 and 2001 were apportioned by year, sex and age to population groups in the same proportions as found in the cause of death sample that had population group recorded on the death certificate.

The numbers of reported deaths by population group between the two censuses was then estimated from the numbers for the years 1999-2001 by back projection and apportioning by a method similar to that used for the national numbers.

The results of these manipulations appear in Appendix 1.

3.2.1.3 Deriving of mortality rates using the General Growth Balance (GGB) method

The estimates of the number of reported deaths derived in this way, together with the estimates of the population by population group, sex and age from each of the 1996 and 2001 censuses[3] were used with the general growth balance (GGB) method proposed by Hill (1987) to produce estimates of the mortality rates by population group, sex and age and for the country as a whole. Essentially, the method relies on the balancing equation applied to the population of persons over a given age. Thus, for a population closed to migration:

$$N(x) - \left[P_2(x+) - P_1(x+)\right] = D(x+)$$

where $P_1(x+)$ and $P_2(x+)$ represent the numbers of persons aged x and over in the population at the first and second censuses respectively, $D(x+)$ represents the number of deaths during the intercensal period to persons aged x and over, and $N(x)$ represents the number of persons reaching exact age x during the intercensal period.

Dividing through by the number of person years lived during the intercensal period by persons aged x and over produces the following relationship:

$$n(x) - r(x+) = d(x+)$$

which when the true quantities are replaced with the observed quantities[4] allows us to estimate not only the extent of completeness of the reporting of deaths, but also the completeness of one census relative to the other.

In order to improve the estimates the method was adapted in two important ways.

---

[3] In the case of the 1996 census those with unknown population group were excluded for these populations but included in the estimate of the national population. Those with unknown age were excluded from both the population group and national population estimates.

[4] In other words $P_i(x+) = P_i^o(x+)/k_i$ and $D(x+) = D^o(x+)/c$ where P^o and D^o are observed numbers and k_i and c are measures of completeness, assumed to be constant over all (adult) ages.

First, the method was adapted to allow for migrants, as substantial migration can distort the numbers of people living in different provinces at different ages. Numbers of migrants were estimated as follows:

- Net immigration of foreigners was estimated from the change in the numbers of foreign-born between the 1996 and 2001 censuses, by population group, sex and age.
- Comparison of the recorded emigration (most presumed to be South African-born) with the change in the numbers of South Africans recorded in recent censuses in the main receiving countries (UK, US, Canada, NZ and Australia) was used to estimate the extent of hidden emigration. These were all assumed to be White (which appears to be a reasonable assumption according to the UK census).
- Projections of the population to 2001 using these estimates of migration and the 1996 population corrected for undercount resulted in more Africans than were estimated by the 2001 census and fewer Whites, both in the 20-29 age group. It was assumed that this excess/shortfall was further hidden immigration/emigration and the estimates changed accordingly.

Although these are not completely accurate estimates of migration, they serve the purpose of ensuring that some allowance is made for the impact that the pattern of migration might have on the level of mortality estimated by this method, and certainly represent an improvement on the conventional assumption that net migration is zero. With the exception of the White population, the numbers are so small relative to the non-migrant population as to have a minimal effect on the estimates of completeness of death reporting produced.

As an alternative we applied an adaptation of the Generalised Growth Balance method suggested by Ken Hill (personal communication) designed to provide in addition, given a pattern of migration rates, estimates of the level of migration.

Second, since the GGB method allows one to estimate the 'undercount' of one census relative to the other, these estimates were used to scale up the relatively underestimated population. An extinct generation method (originally proposed by Bennett and Horiuchi (1981; 1984) was used as a check on the estimate of under-recording of deaths derived from the GGB method. Although a description of this method entails moderately complicated formulae, the basic idea originally proposed by Carrier (1958) is simple, namely that where there is no error caused by lack of completeness in the data, the number in the population at a specific age at a point in time must be equal to the number of deaths in future years arising out of that cohort. These numbers of deaths in turn can be estimated, on the assumption that mortality remains constant over time, from the number of deaths recorded in an interval by noting that deaths at any particular age will grow at the population growth rate at that age.

Interestingly these methods suggest that the 1996 census population estimate for adults was undercounted by some 5 per cent relative to the 2001 census estimate of the population. The relative under count was 7.5 per cent, 6 per cent and 3 per cent for the African, Coloured and Indian populations respectively with virtually no relative undercount of the White population[5].

[5] It should be noted that we are talking here of the undercount of one census relative to another. This says nothing of the overall level of the count, or indeed, which of the two censuses is closer to the truth. It should also be noted that part of the relative undercount (around 1% absolute) can be accounted for by those for whom either population group or age were unknown.

The raw mortality rates derived from the numbers of deaths, increased to allow for the under-reporting of deaths, and the census population estimates corrected for relative undercount, were then graduated using the logit relation postulated by Brass (1968):

$$\lambda\left(l(x)\right) = \alpha + \beta \cdot \lambda\left(l^s(x)\right)$$

where $\lambda\left(l(x)\right) = \text{logit}\left(1 - l(x)\right) = 0,5\ln\left(\frac{1-l(x)}{l(x)}\right)$, and

$l(x), l^s(x) =$ the proportion surviving to age *x* in the fitted and a 'standard' life table respectively

It is imperative that the standard table used has the right shape, in particular allowing, where appropriate, for the appropriate level of excess mortality due to HIV/AIDS. Unfortunately such standard tables do not exist and while it would appear from work creating such life tables ((INDEPTH (International Network for the continuous Demographic Evaluation of Populations and their Health) 2002; Murray, Ahmad, Lopez *et al.* 2000; Timæus 2004)) quite possible, through the introduction of one additional parameter, to allow the magnitude of the 'AIDS hump' to be determined by the data, such a system has yet to be created. Thus we decided to accept the raw rates as far as possible and only to graduate the rates at the oldest ages (60-85 for Africans, 70-85 for Coloured and 80-85 for Indian and White populations), using Brass's logit relation and General Standard. In the interests of smoothness the rates at some ages (one age group each for African males and females, and Indian males, and two age groups for Indian females) were interpolated between the rates at adjacent ages, where they appeared to be out of line. *Since some may look to the estimates of mortality produced in this report for evidence of the impact of HIV/AIDS on mortality, it should be emphasized that use of these tables as standards does not exaggerate the hump in mortality schedules observed between ages 25 and 45.*

3.2.1.4 Deriving the level of mortality using data on survival of parents of 'respondents' in the census

Brass (Brass and Hill 1973) first proposed a method to estimate adult mortality based on data on the proportions of respondents whose mother (father) were reported as still living. Since then the method has been improved by, for example, Hill and Trussell (1977) and Timæus (1992). Since we have data on the survival of parents from two censuses, cohorts in the first census can be identified in the second and survivorship over the intersurvey period can be estimated by constructing a hypothetical or synthetic intersurvey cohort of respondents. However, orphanhood data are liable to distortions due to the so-called 'adoption' effect (confusion of de facto and biological parents) at younger ages. We have therefore used a variation of the method and coefficients devised by Timæus (1991) which uses the responses of those aged 20 and above only.

Again, the choice of standard is important if one is to infer a level of mortality over a particular age range, particularly so if this range includes young adults in the context of an AIDS epidemic. Our approach is to use as standard a table produced from Brass's General Standard using Brass's logit relation with β set to an appropriate value. Earlier work with vital registration data up to 1996 suggests that when estimating ${}_{15}q_{50}$ for the country as a whole (and presumably also for the African population) β could best be set to 1. However, β was (somewhat arbitrarily) set to 1.4 for the other population groups to reflect the lighter (but more steeply rising) mortality rates in this age range in these groups.

In order to apply the orphanhood method one needs an estimate of the average age of mothers and fathers at the birth of their children. This average age does change over time but a reasonable approximation is the average age at the time when the births of those aged 15-59 occurred on average, which is roughly in the mid-1960s.

These average ages were estimated as follows:

- The average age of African parents (mothers and fathers) has been observed by Timæus, Dorrington, Bradshaw et al, (2001) to be 26.8 (mothers) and 33.7 (fathers) from the SADHS data.
- The average age of Coloured, Indian and White parents was estimated from published data on the average age of mothers and fathers at the birth of their children for selected years in the 1960s (Buro (sic) of Statistics 1968; 1972). The average ages at maternity were 27.0, 27.0 and 26.8 for Coloured, Indian and White mothers, and 30.0, 31.0 and 30.5 for Coloured, Indian and White fathers.
- The difference between the estimated age at maternity were small relative to the uncertainty and all slightly higher than the estimate for all population groups combined derived from the SADHS, so it was decide to use 26.8 for all groups individually and combined.
- For age at paternity we used the estimates derived above, namely, 33.7, 30.0, 31.0 and 30.5 for the African, Coloured, Indian and White population groups and 33.00 for the country as a whole.

In the case of the provincial rates a weighted average, weighted by the relative size of the populations aged 50 and over, of the mean age at childbirth and of the betas for the various population groups, was used. The estimates of ${}_{15}q_{50}$ derived from the life tables produced above were compared with those derived using the orphanhood data in order to check consistency of the level of mortality being produced by the two methods.

3.2.1.5 Provincial mortality rates

Unfortunately, the vital registration data cannot be used to produce reliable estimates of mortality at anything below the national level, since many deaths are recorded not at the place of residence but at the place where the death was registered, thus violating the required correspondence between the deaths and the population exposed to the risk of death necessary in order to estimate mortality rates. Thus, in order to estimate mortality rates at a provincial level, it is necessary to make use of both the estimates of mortality derived from the vital registration data as well as the number of deaths reported by households in the census. To do this we need to make the assumption that the relationship between deaths reported by households and the estimates derived according to the method outlined above, apply in each of the provinces.

Specifically, this has been achieved by first estimating the mortality rates applicable for the year immediately preceding the 2001 census. The number of deaths in this year is estimated from the estimated number reported in that year, derived as explained in section 3.2.1.2, by assuming that the completeness of death registration changed (increased) linearly over the period such that the mortality rate of those over 65 shows no trend (the slope is zero). The reasonableness of this assumption was checked by comparing the trend in adult mortality rates thus produced with the trends suggested by past estimates. No evidence was found to suggest that this assumption is unreasonable. We then calculate the ratios of the expected number of deaths based on the national estimates of mortality derived above to the number reported by households in the census by population group, sex and age. The resulting ratios are then used to correct the number of deaths reported at a provincial level for under- (or over-) reporting. Finally, the rates produced by these estimates of the number of deaths in the province have been inspected for reasonableness and the implied ${}_{15}q_{50}$ compared with that estimated from the orphanhood data for the provinces.

This is implemented in the next chapter.

4 Results – adult mortality rates

4.1 Completeness of death reporting (relative to the census enumeration)

Table 4.1 shows the completeness of the estimates of deaths derived from the population register and the sample of death certificates drawn by Statistics South Africa when compared to the population estimated by the censuses adjusted for any relative undercount. The 'national adjusted' is the level that would be necessary to produce rates equivalent to the weighted average of the rates of the population groups (using the most recent census numbers as weights).

Table 4.1 Completeness of death reporting (relative to the census enumeration)

	African	*Coloured*	*Indian*	*White*	*National*	*National adjusted*
Males	63.9%	69.5%	64.8%	77.0%	83.5%	83.4%
Females	66.6%	69.7%	83.4%	78.5%	86.7%	84.5%

The completeness at a population group level is somewhat lower than that of the population as a whole because around 20 per cent of recorded deaths were not classified by population group. Bearing this in mind, it appears that reporting of death is better for women than men and is most complete for the White population group and for Indian women, and lowest for the Indian and African men. However, it is difficult to understand, without further investigation, the large differences between the estimated completeness of Indian and males and females. If one assumes that close to 100 per cent of Coloured, Indian and White deaths are recorded, then the figures in Table 4.1 for these population groups are measures of the completeness of recording of population group on the death certificate while those for the African population reflect both failure to report some deaths and failure to record population group on certificates.

The estimate of completeness based on national data and that implied by a weighted average of the population group estimates are not the same but nevertheless quite close. In part this difference could be due to the heterogeneous nature of the population with respect to reporting and hence a violation of the assumption that completeness is the same for all ages (the completeness of registration is unlikely to be equal for all the population groups and the proportions of the groups are not the same at all ages). Thus there are reasons for preferring the estimate that is consistent with the weighted average of the estimates derived for each population group.

Interestingly the adaptation of the Generalised Growth Balance, which also produces estimates of migration, produced similar estimates of completeness for African, Coloured and Indian populations, probably because the level of migration was low in these three population groups. In the case of the White population, the method estimated the level of completeness of death registration to be higher (and the mortality rates to be lower). However, the estimate of migration produced by this method for the White population was clearly inconsistent with the pattern of numbers counted in the censuses, and thus it was decided to prefer the above estimates. Also of interest is the fact that in the case of the African, Coloured and Indian populations the method suggested net emigration, where our estimates suggest net immigration, at least among young adults.

4.2 Graduated mortality rates

Table 4.2 and Table 4.3 present the graduated (curve fitted to the observed rates) mortality rates for each of the population groups and nationally, together with the weighted average of the population group-specific rates. Figure 4.1 compares these rates with the raw rates derived from adjusting the estimate of the number of reported deaths for under-reporting and adjusting the census populations for relative under-enumeration.

From these, we see that the rates appear to meet with expectations in terms of the ranking of population groups, with, the mortality rates of the African population group being generally higher than those of the Coloured group, and the rates for both these groups being somewhat higher than those of the Indian and White groups. Also, the mortality rates of males, in each group, are higher than those of the females in that group. However, within this overall ranking, one can note several points of interest. First, mortality after the age of 65 for men and 50 for women appears to be highest among the Coloured population. Second, the mortality of Coloured men aged 15-19 is at least as high as that of African men of the same age group. Third, mortality of African women in the 20-29 age range is nearly as high as that of African men in this age range. Finally, at ages over 70 the mortality of the Indian population appears to be higher than that of the African population, although, it is not clear how much reliance can be placed on this result.

We reiterate, as was mentioned above, that the weighted average of the population group-specific rates are slightly higher than the rates derived from the deaths ignoring population group for females.

From Figure 4.1 we see that the fit of the graduated curve to the raw estimates, although appears to be reasonably good with the exception of ages where the graduated rates have been deliberately set to differ from the raw rates.

It is important to note, further, that it is clear from these comparisons that the graduation process in no way introduces or enhances any possible 'AIDS' hump in the data.

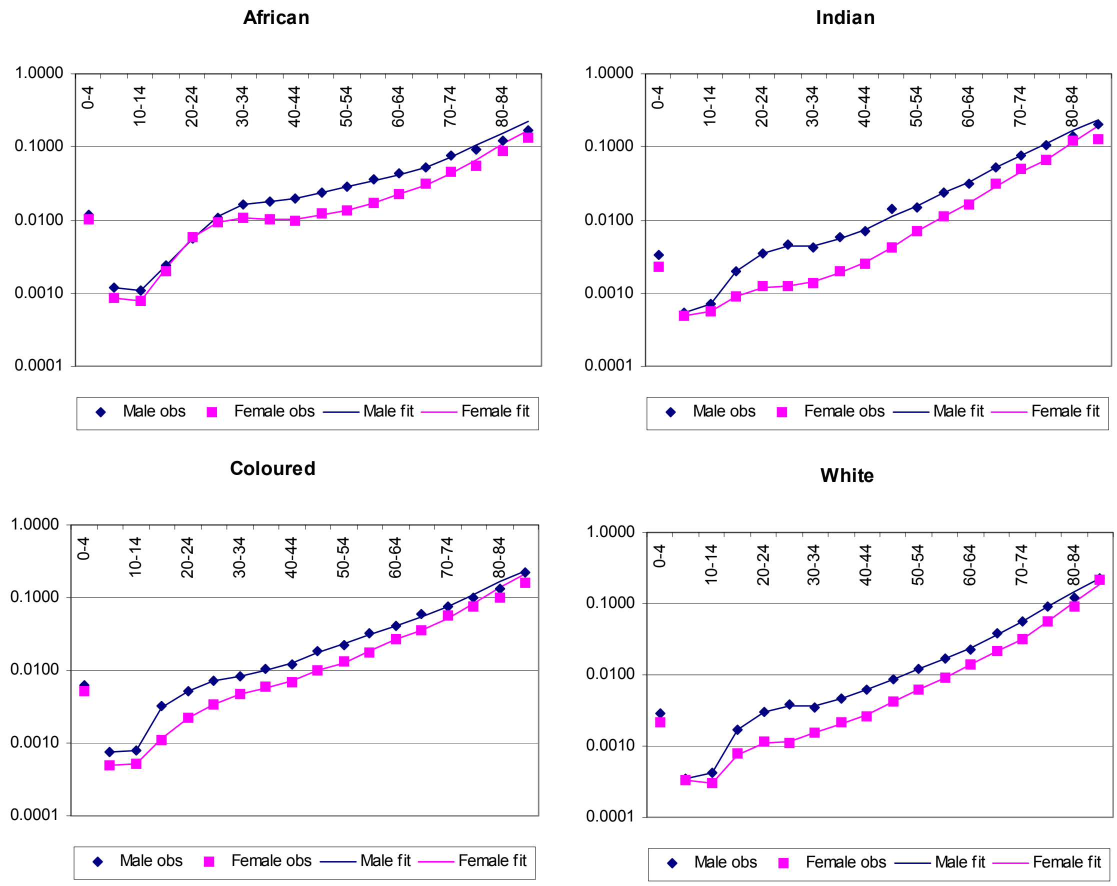
Table 4.2 Graduated mortality rates: males

	African	*Coloured*	*Indian*	*White*	*National*	*National Weighted average*	*Ratio of National to weighted*
5-9	0.0012	0.0008	0.0005	0.0003	0.0010	0.0011	94%
10-14	0.0010	0.0008	0.0007	0.0004	0.0009	0.0010	96%
15-19	0.0024	0.0033	0.0020	0.0017	0.0023	0.0024	94%
20-24	0.0057	0.0052	0.0036	0.0031	0.0052	0.0054	95%
25-29	0.0116	0.0076	0.0048	0.0039	0.0094	0.0104	90%
30-34	0.0167	0.0084	0.0043	0.0036	0.0127	0.0142	90%
35-39	0.0185	0.0108	0.0059	0.0048	0.0143	0.0158	91%
40-44	0.0203	0.0123	0.0071	0.0064	0.0160	0.0172	93%
45-49	0.0250	0.0185	0.0120	0.0088	0.0200	0.0215	93%
50-54	0.0289	0.0229	0.0153	0.0123	0.0244	0.0249	98%
55-59	0.0362	0.0326	0.0247	0.0173	0.0321	0.0314	102%
60-64	0.0428	0.0418	0.0323	0.0240	0.0388	0.0382	102%
65-69	0.0531	0.0562	0.0529	0.0382	0.0496	0.0499	100%
70-74	0.0747	0.0757	0.0774	0.0555	0.0683	0.0702	97%
75-79	0.1076	0.1121	0.1083	0.0896	0.1003	0.1031	97%
80-84	0.1559	0.1651	0.1648	0.1462	0.1483	0.1542	96%
85+	0.2224	0.2354	0.2388	0.2240	0.2151	0.2237	96%

Table 4.3 Graduated mortality rates: females

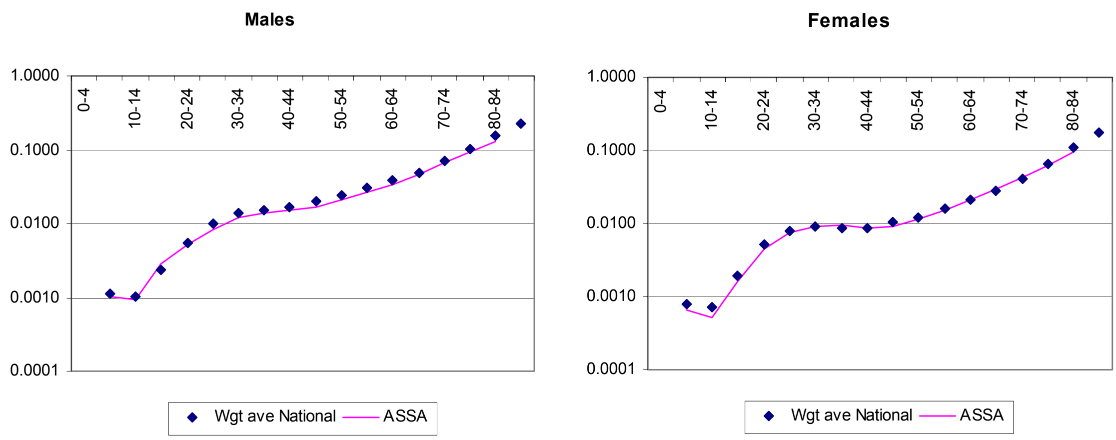
	African	*Coloured*	*Indian*	*White*	*National*	*National Weighted average*	*Ratio of National to weighted*
5-9	0.0009	0.0005	0.0005	0.0004	0.0008	0.0008	98%
10-14	0.0008	0.0005	0.0006	0.0003	0.0007	0.0007	102%
15-19	0.0020	0.0011	0.0009	0.0008	0.0016	0.0018	88%
20-24	0.0061	0.0023	0.0013	0.0012	0.0043	0.0053	81%
25-29	0.0098	0.0034	0.0013	0.0011	0.0074	0.0084	89%
30-34	0.0112	0.0047	0.0014	0.0016	0.0083	0.0093	89%
35-39	0.0107	0.0060	0.0021	0.0022	0.0082	0.0090	91%
40-44	0.0103	0.0071	0.0026	0.0027	0.0083	0.0088	94%
45-49	0.0126	0.0103	0.0042	0.0043	0.0098	0.0109	90%
50-54	0.0137	0.0133	0.0070	0.0062	0.0117	0.0122	96%
55-59	0.0182	0.0178	0.0116	0.0094	0.0159	0.0162	98%
60-64	0.0227	0.0277	0.0168	0.0145	0.0210	0.0215	98%
65-69	0.0297	0.0351	0.0290	0.0221	0.0288	0.0288	100%
70-74	0.0441	0.0519	0.0465	0.0329	0.0407	0.0426	96%
75-79	0.0683	0.0843	0.0660	0.0552	0.0646	0.0662	98%
80-84	0.1088	0.1369	0.1145	0.1024	0.1056	0.1089	97%
85+	0.1719	0.2118	0.1903	0.1819	0.1706	0.1769	96%

Figure 4.1 Comparison of the graduated rates with the observed rates (log scale)



Further, as shown in Figure 4.2, comparison of the national rates show a shape of mortality that is entirely consistent with that projected by models of the AIDS epidemic, such as the ASSA2000 model.

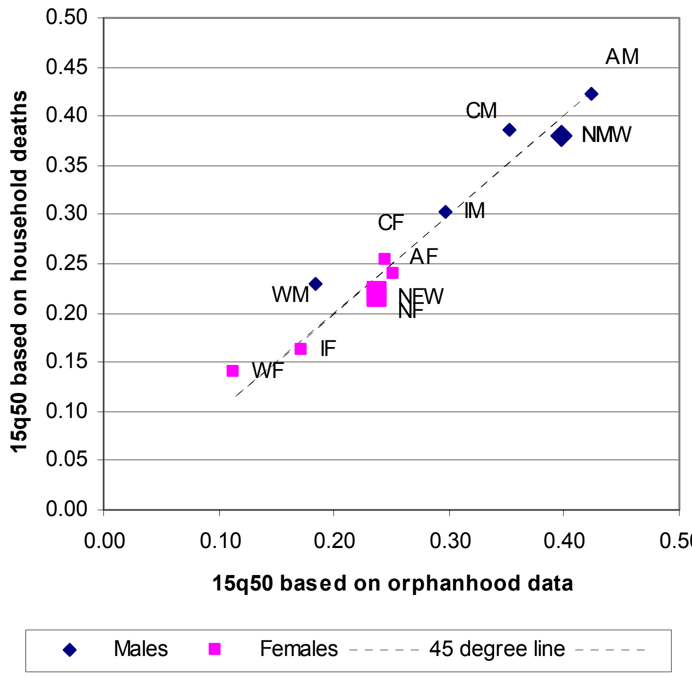
Figure 4.2 Comparison of the weighted average national mortality rates against those projected using the ASSA2000 model (log scale)



4.3 Comparison of $_{15}q_{50}$ with that estimated using the orphanhood data

Figure 4.3 compares estimates of $_{15}q_{50}$ derived from the rates shown above with those estimated on the basis of the data on survival of parents of those enumerated in the census.

Figure 4.3 Comparison of $_{15}q_{50}$ with that estimated using the orphanhood data



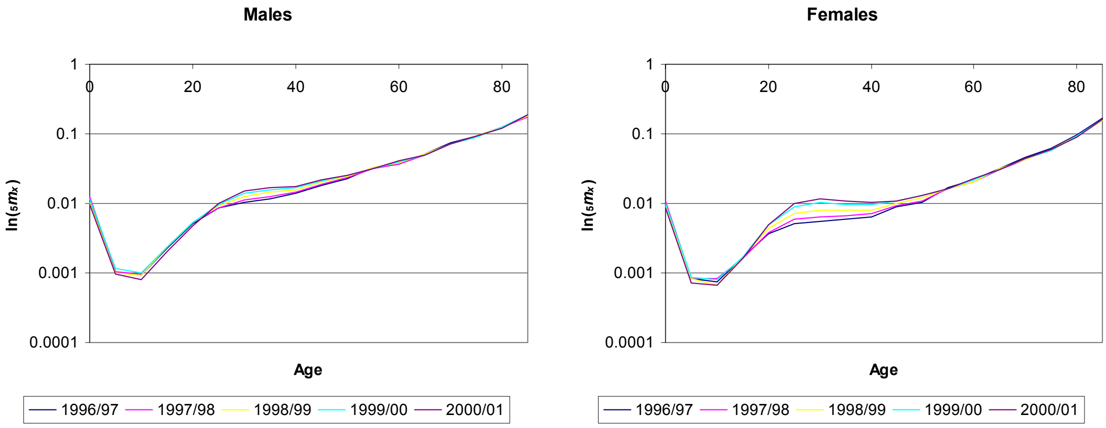
AM – African Males, AF – African Females, etc, NM – National Male, NMW – National Female Weighted, etc.

By and large, there appears to be a high degree of consistency between the estimates, with the weighted average of the population group-specific rates being close to the estimates of national mortality using the orphanhood data. However, the White male (and to a relatively lesser extent, White female) rates produced using the vital registration data appear to be too high and further investigation may be needed to decide which of the two estimates of the level of mortality is correct.

4.4 Mortality rates over time

Figure 4.4 shows the change in mortality over time. The clear trend of rising mortality over time in the 15-55 age range is found exclusively in the African population group, with no trend in mortality over the intercensal years discernable in the other population groups.

Figure 4.4 National mortality rates showing rising trend over time



4.5 Correcting the number of household deaths to agree with the estimates of mortality

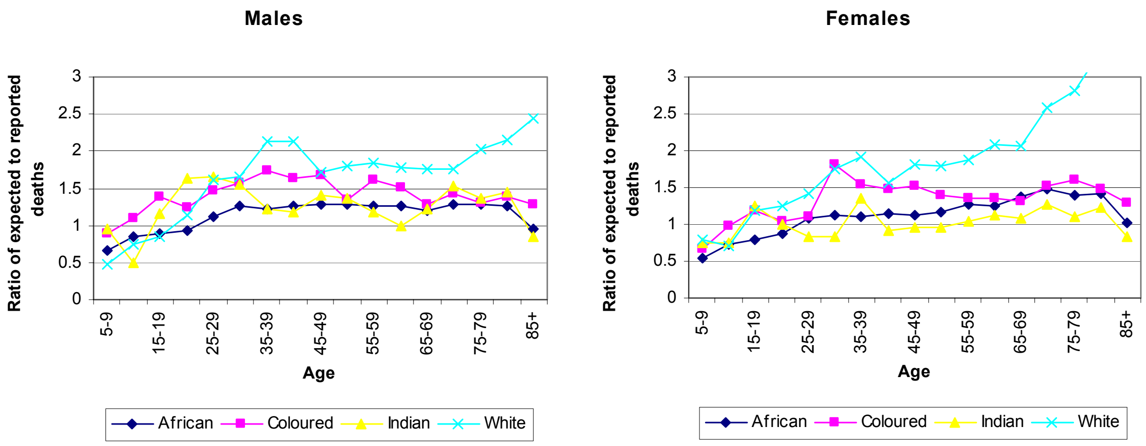
Figure 4.5 shows the extent of the adjustment to the number of deaths reported by households described in section 3.2.1.5 by sex, population group and age. The first thing to notice from this figure is that the adjustment at the peak of what might be the 'AIDS hump' (in the age range 30-39 for males and 25-34 for females) is minimal. Second, with the exception of the White population group the ratio is pretty constant but rises with age for African women. However, for the White population group the extent of under-reporting increases with age, probably because of a much higher proportion either dying alone or in institutions. Third, the adjustment factor is below 1 for the African, Indian and White children, possibly due to 'extended' or 'stretched' households (where some individuals are regarded as living in different places for part of the time) reporting the death of an individual at more than one place at the time of the census, but undoubtedly also because the under-registration of deaths of children is worse than that of adults.

Table 4.4 shows the extent of adjustment to deaths of those aged five and over. Perhaps not unexpectedly the adjustment is higher for male deaths than for female deaths and highest for the White population group. However, the adjustment for Indian females is somewhat counter-intuitive. This is undoubtedly largely attributable to the poor fit of the graduated rates at the older ages.

Table 4.4 Adjustment to the number of deaths reported by household, overall

	African	*Coloured*	*Indian*	*White*
Male	1.19	1.46	1.28	1.85
Female	1.14	1.40	1.07	2.60

Figure 4.5 Extent of adjustment to the number of deaths reported by households



4.6 Mortality rates by province

Figure 4.6 and Figure 4.7 show the estimated mortality rates for the African population by province for males and females respectively. These are contrasted with the national rates for the African population for the year prior to the 2001 census based on the vital registration data.

Unfortunately these results are not every encouraging. First, while it is quite plausible to find that KwaZulu-Natal has the highest rates for much of the young adult age range, closely grouped with the Eastern Cape and Free State, and that the Western Cape and Limpopo (represented by NP or Northern Province in 2001 census data) have the lowest, it is somewhat surprising, given the level of HIV prevalence in the province, to see rates of mortality for Gauteng so close to these low rates, or possibly such a big difference between the groups of high and low rates. Second, rates based on the household deaths for Western Cape and Gauteng, increase at the older ages at a far more rapid rate than those of the other provinces. It is highly unlikely that old people in Western Cape and Gauteng experience significantly higher mortality than those in the other provinces. Equally it seems unlikely that the mortality of older people in Limpopo would be so much lighter than in the other provinces. Finally, although one should not read too much into the data below age 20, the rates for the Western Cape and Gauteng men are very different from the rest, and it seems unlikely that the mortality of 10-14 year olds is lightest in Limpopo.

The highly erratic results for the other population groups provide no grounds for producing province-specific estimates for these population groups. However, for completeness the raw rates for all population groups by province are presented in Figure 4.8 and Figure 4.9.

Figure 4.6 Estimated mortality rates of African males by province, Census 2001 (log scale)

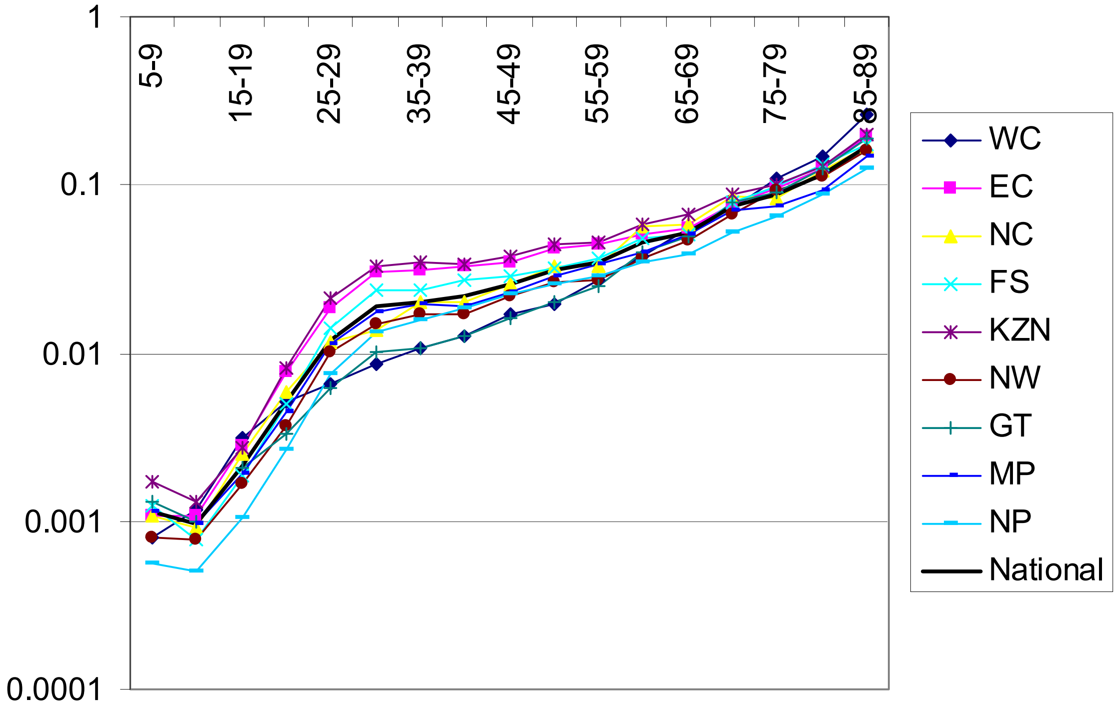


Figure 4.7 Estimated mortality rates of African women by province, Census 2001 (log scale)

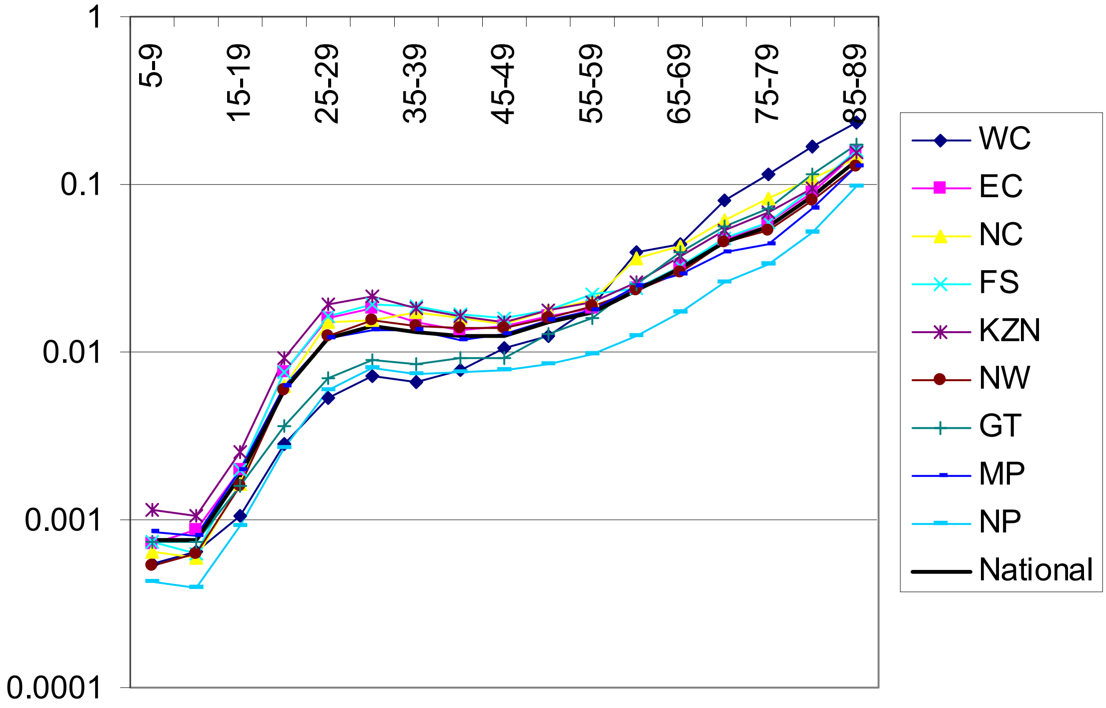


Figure 4.8 and Figure 4.9 show the same thing but for all population groups combined.

Figure 4.8 Estimated mortality rates of all males by province, Census 2001 (log scale)

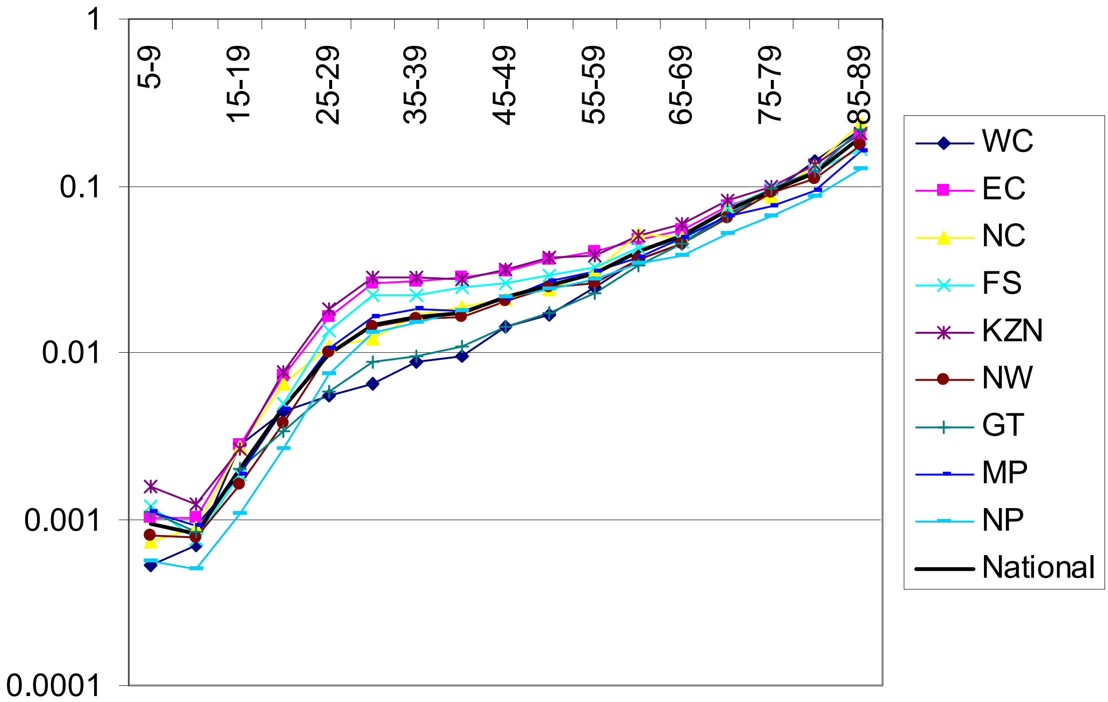
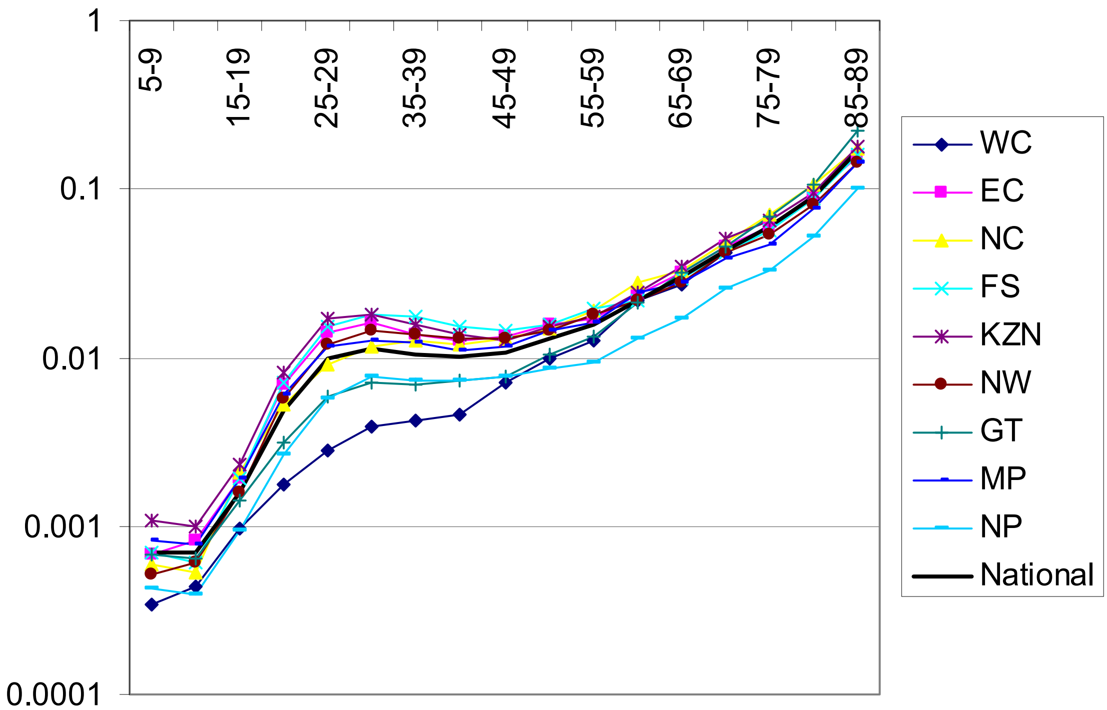


Figure 4.9 Estimated mortality rates of all females by province, Census 2001 (log scale)



4.7 Comparison of $_{15}q_{50}$: Provinces

Table 4.5 compares the ranking of the provinces by the mortality rates derived using the data on survival of the parents against that based on the rates estimated above, while Figure 4.10 shows the same data graphically.

Although there is, with the clear exception of the Northern Cape, a fair degree of consistency between the rankings of the province by female mortality there is less consistency as far as the males are concerned. However, keeping in mind that in a (probably significant) number of cases parents do not live in the same province as their children, the orphanhood data can be regarded as providing some corroboration of the results derived from the household deaths. Figure 4.10 reveals that the rates based on the survival of parents by and large are higher (lie below the 45 degree line) than those estimated using deaths reported by households. To a large extent this probably merely reflects the relationship at the national level.

Table 4.5 Comparison of the ranking of the provinces on the basis of $_{15}q_{50}$ (from lowest to highest)

	Males		*Females*	
Ranking	*Orphanhood*	*Household deaths*	*Orphanhood*	*Household deaths*
1	Northern Cape	*Gauteng*	**Limpopo**	**Limpopo**
2	*Gauteng*	*Western Cape*	*Gauteng*	*Western Cape*
3	*Western Cape*	*North West*	*Western Cape*	*Gauteng*
4	*North West*	*Limpopo*	**North West**	**North West**
5	*Limpopo*	Mpumalanga	**Mpumalanga**	**Mpumalanga**
6	**Free State**	**Free State**	**Eastern Cape**	**Eastern Cape**
7	Mpumalanga	Northern Cape	Northern Cape	KwaZulu-Natal
8	**Eastern Cape**	**Eastern Cape**	**Free State**	**Free State**
9	**KwaZulu-Natal**	**KwaZulu-Natal**	KwaZulu-Natal	Northern Cape

Bold indicates consistent ranking, italic indicates rankings within one position of one another.

There is a need to investigate the reason for the rapid increase in mortality rates with age in Gauteng and Western Cape, the impact of assuming that parents live in the same province as their children over 20, misfit of the rates for the Northern Cape, and possible reasons for the fact that the rates based on the orphanhood data are on average higher than those based on deaths in the household. However, in the mean time Table 4.6 and Table 4.7 present graduated rates for the provinces using methods very similar to those described already with regard to the national and population group estimates.

Figure 4.10 Comparison of $_{15}q_{50}$ with that estimated using the orphanhood data: provinces

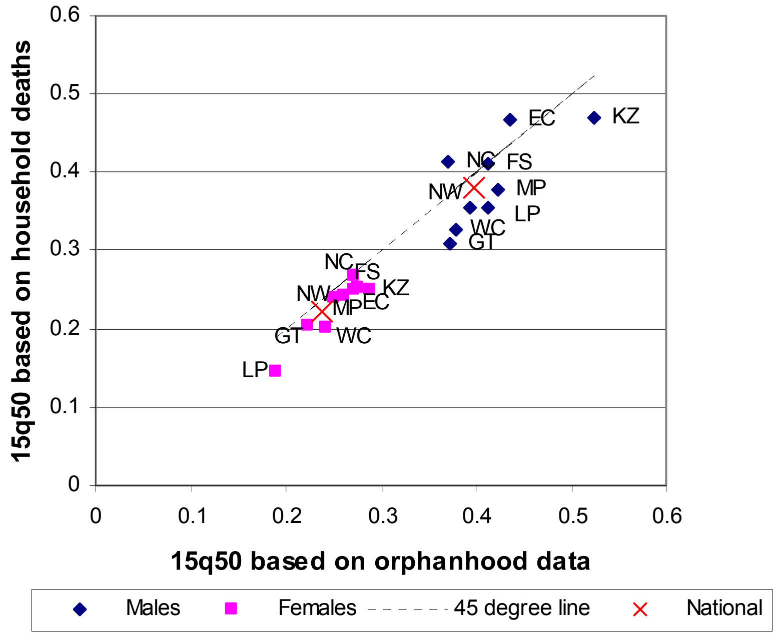


Table 4.6 Graduated mortality rates, by province: males

	WC	*EC*	*NC*	*FS*	*KZ*	*NW*	*GT*	*MP*	*LP*
5-9	0.0005	0.0010	0.0007	0.0012	0.0016	0.0008	0.0011	0.0011	0.0006
10-14	0.0007	0.0010	0.0009	0.0007	0.0012	0.0007	0.0008	0.0009	0.0005
15-19	0.0027	0.0028	0.0027	0.0018	0.0026	0.0016	0.0020	0.0018	0.0011
20-24	0.0044	0.0074	0.0066	0.0049	0.0077	0.0038	0.0034	0.0045	0.0027
25-29	0.0057	0.0167	0.0111	0.0137	0.0189	0.0101	0.0059	0.0109	0.0075
30-34	0.0067	0.0264	0.0123	0.0224	0.0286	0.0145	0.0088	0.0167	0.0134
35-39	0.0089	0.0274	0.0169	0.0226	0.0291	0.0163	0.0096	0.0185	0.0153
40-44	0.0097	0.0289	0.0189	0.0247	0.0280	0.0166	0.0111	0.0178	0.0181
45-49	0.0145	0.0310	0.0209	0.0263	0.0317	0.0204	0.0144	0.0212	0.0217
50-54	0.0170	0.0368	0.0242	0.0295	0.0376	0.0248	0.0172	0.0266	0.0243
55-59	0.0255	0.0402	0.0315	0.0329	0.0396	0.0268	0.0236	0.0312	0.0285
60-64	0.0362	0.0419	0.0449	0.0379	0.0449	0.0323	0.0298	0.0368	0.0317
65-69	0.0445	0.0556	0.0502	0.0513	0.0596	0.0452	0.0442	0.0473	0.0361
70-74	0.0644	0.0755	0.0713	0.0714	0.0809	0.0652	0.0674	0.0611	0.0513
75-79	0.0991	0.1056	0.1038	0.1020	0.1125	0.0966	0.1047	0.0887	0.0757
80-84	0.1517	0.1501	0.1519	0.1475	0.1587	0.1441	0.1604	0.1311	0.1147
85+	0.2234	0.2128	0.2186	0.2117	0.2227	0.2109	0.2342	0.1931	0.1740

Table 4.7 Graduated mortality rates, by province: females

	WC	*EC*	*NC*	*FS*	*KZ*	*NW*	*GT*	*MP*	*LP*
5-9	0.0004	0.0007	0.0006	0.0007	0.0011	0.0005	0.0007	0.0008	0.0004
10-14	0.0004	0.0008	0.0005	0.0006	0.0009	0.0006	0.0006	0.0007	0.0004
15-19	0.0010	0.0019	0.0021	0.0019	0.0023	0.0016	0.0014	0.0019	0.0010
20-24	0.0018	0.0070	0.0053	0.0073	0.0084	0.0058	0.0032	0.0061	0.0027
25-29	0.0028	0.0146	0.0094	0.0157	0.0173	0.0122	0.0060	0.0118	0.0059
30-34	0.0039	0.0165	0.0120	0.0185	0.0185	0.0147	0.0072	0.0129	0.0079
35-39	0.0043	0.0139	0.0126	0.0175	0.0156	0.0138	0.0069	0.0125	0.0073
40-44	0.0047	0.0128	0.0123	0.0155	0.0140	0.0132	0.0075	0.0111	0.0075
45-49	0.0071	0.0134	0.0131	0.0147	0.0129	0.0131	0.0079	0.0117	0.0079
50-54	0.0099	0.0157	0.0149	0.0161	0.0155	0.0148	0.0106	0.0147	0.0087
55-59	0.0130	0.0182	0.0201	0.0190	0.0186	0.0180	0.0146	0.0162	0.0100
60-64	0.0221	0.0223	0.0235	0.0216	0.0243	0.0203	0.0194	0.0245	0.0116
65-69	0.0263	0.0310	0.0335	0.0294	0.0335	0.0281	0.0294	0.0272	0.0165
70-74	0.0398	0.0448	0.0496	0.0417	0.0481	0.0407	0.0466	0.0359	0.0250
75-79	0.0674	0.0676	0.0765	0.0619	0.0719	0.0618	0.0770	0.0542	0.0403
80-84	0.1160	0.1051	0.1200	0.0953	0.1104	0.0971	0.1283	0.0853	0.0692
85+	0.1916	0.1640	0.1855	0.1490	0.1699	0.1540	0.2036	0.1370	0.1223

5 Childhood mortality rates

This chapter is a summary of the investigations carried out on the data that could be used to estimate child mortality using indirect techniques based on women's responses to questions on the numbers of children ever born (CEB) and the numbers of those children surviving (CS). In part, the material builds on earlier work relating to the average numbers of children born to women, which are required to estimate levels of fertility using indirect techniques. This chapter extends the analysis to examine women's reports of their numbers of surviving children and the resulting ratios of surviving to ever born children.

5.1 Overall data quality

The report on fertility in South Africa based on the 2001 census (Moultrie and Dorrington 2004) has already pointed out the significant flaws in the data on children ever born (CEB), both in terms of the extent of imputation (particularly at younger ages), and in terms of the overall level of childbearing implied by the data (which are logically inconsistent with previous census data at older ages).

At the outset, we must emphasise that higher proportions not subjected to hotdecking and imputation do not carry with them any suggestion that the underlying data are 'better' or more reliable: the absence of imputation or hotdecking simply implies that there are no substantive grounds for rejecting the raw data. Misunderstandings and enumeration errors can still render the responses offered meaningless.

In aggregate, the data are of very poor quality. Table 5.1 shows the proportion of women whose responses on their numbers of children ever born and children surviving were not subject to any form of imputation or hotdecking.

Table 5.1 Percentage of cases where no editing of children ever born and children surviving data was required, by population group and age group

	African		*Coloured*		*Indian/Asian*		*White*	
Age	*CEB*	*CS*	*CEB*	*CS*	*CEB*	*CS*	*CEB*	*CS*
12-14	65.2	34.5	53.5	27.2	61.4	19.7	46.2	22.6
15-19	73.5	44.0	63.7	37.2	68.8	24.6	55.9	28.9
20-24	82.5	62.5	78.5	59.5	79.1	40.9	73.9	44.5
25-29	88.2	75.6	87.6	75.4	88.0	64.3	85.4	63.6
30-34	90.9	81.2	91.2	82.0	92.2	78.3	90.2	76.6
35-39	91.9	83.2	92.6	84.5	93.5	82.9	91.3	81.3
40-44	91.4	83.3	92.5	84.7	93.3	83.6	91.5	82.7
45-49	89.9	82.3	91.3	83.7	91.9	82.6	90.4	82.2

At all ages, and across all population groups, the data on children surviving (CS) are demonstrably worse than those on children ever borne. Some of this error is attributable to the logical dependence of the second question on the first: A woman who has not had children, for example, may not have been asked about her surviving children and the enumerator may have left the relevant entries on the enumeration form blank (rather than entering zeros), thereby necessitating editing of the data.

A similar phenomenon is at play in the data on children ever born: evidently childless women may not be enumerated as being childless (i.e. recorded as being of parity zero), with the enumerator failing to record any response on the enumeration form. As mentioned in the report on fertility (Moultrie and Dorrington 2004), edit rules were derived by Statistics South Africa in association with staff from the US Bureau of the Census and applied to the data. These rules, we argued, did not fully capture the extent of the error of non-reporting for those for whom a question is deemed "irrelevant", originally described by el-Badry (1961). The conclusion reached there suggested that the edits applied to the data distorted the parity data more than they improved them.

Furthermore, it was shown that the data on children ever born, no matter how heroically manipulated, were logically inconsistent with the results from the 1996 census. An estimate of the corrected average parities in 2001 was derived, based on the estimated levels of fertility between 1996 and 2001 and the average parities in 1996. Unfortunately, the adjustment technique applied there is applicable only to the data on children ever born, since correction of the data on children surviving requires estimates of child mortality, the very thing we are trying to measure.

Nonetheless, the extent of editing and hotdecking in the data on children surviving is startling: at best, the data for every sixth woman (among Coloured women aged 40-44) had to be inferred or hotdecked. At the other extreme, the reported numbers of children surviving among Indian women aged between 12 and 14 were hotdecked or subjected to logical imputation in four out of every five cases. Such reliance on editing and hotdecking rules does not inspire confidence in the data.

Of the four possibilities open to edit and correct the data, the use of a hotdeck applied to non-missing data was the least used in the case of the children surviving data (Table 5.2). This is a relief, since the possibility that unintended biases can be introduced into the data makes any reliance on hotdecking as a form of data correction suspect. Of great concern, however, is the failure of the logical imputation system to deal with evidently childless women at younger ages. This can be seen by the preference for hotdecking over imputation at younger ages except (for some reason) among Indians. In any event, the extent of missing data on children surviving among women younger than 25 is readily apparent, and is indicative of a systemic failure in the collection of these data. The most likely explanation must be that enumerators were not adequately trained in the collection of these data, and were not appraised of the importance of these data in the derivation of crucial demographic variables.

Table 5.2 Percentage of edited children surviving cases subjected to different forms of editing, by population group and age group

	Logical – Data missing				*Logical – Data not missing*				*Hotdeck – Data missing*				*Hotdeck – Data not missing*			
Age	*A*	*C*	*I*	*W*	*A*	*C*	*I*	*W*	*A*	*C*	*I*	*W*	*A*	*C*	*I*	*W*
12-14	30.8	26.4	41.7	23.6	0.3	0.2	0.2	0.2	34.2	46.1	38.2	53.3	0.2	0.2	0.2	0.2
15-19	28.9	25.9	44.1	27.1	1.2	0.9	0.4	0.4	25.6	35.8	30.7	43.5	0.3	0.2	0.1	0.1
20-24	17.5	16.8	37.5	29.1	3.5	3.0	1.6	1.3	15.9	20.4	19.7	24.9	0.5	0.3	0.2	0.2
25-29	9.1	9.1	21.7	21.0	5.1	4.4	3.5	2.8	9.5	10.7	10.2	12.3	0.7	0.4	0.3	0.3
30-34	5.5	5.7	11.2	12.0	6.0	5.1	4.6	3.8	6.4	6.8	5.5	7.2	0.9	0.4	0.4	0.4
35-39	4.4	4.4	7.6	8.0	6.5	5.4	5.0	4.4	5.0	5.2	4.2	5.9	1.0	0.4	0.4	0.4
40-44	3.9	4.0	6.7	6.5	6.4	5.5	4.9	4.5	5.2	5.3	4.4	5.8	1.2	0.5	0.4	0.5
45-49	3.7	4.1	6.6	6.2	6.3	5.3	4.6	4.3	6.5	6.4	5.7	6.8	1.3	0.5	0.5	0.6

Note: A-Africans; C-Coloureds; I-Indians; W-Whites

Likewise (and again, with the exception of Indians) the use of a hotdeck was chosen over logical means (in other words, the answer could not be deduced logically) among women over 40. Again, this points to a fundamental flaw in the raw data.

The known errors in the parity data notwithstanding, the extent of the editing and manipulation of the data indicated by Table 5.1 and Table 5.2 should be sufficient to alert users of these data that they are not to be relied upon.

The report on fertility suggested that one explanation for the lower, and logically inconsistent, parity data indicated by the 2001 census might be that women excluded their dead children from the numbers of children ever born (while reporting the correct numbers of children surviving), thereby biasing their responses towards a greater proportion of their children still surviving. The rest of this chapter considers the data on children surviving in greater detail to determine whether a plausible set of corrections can be applied to these data so that, when divided by the (corrected) estimates of average parity, plausible ratios of the proportion surviving to ever born are derived, thereby allowing the production of plausible estimates of childhood mortality from standard indirect techniques. Several different scenarios are investigated.

5.1.1 Using completely unedited data

Given our uncertainty regarding the usefulness and reliability of the editing procedures used, we first examined the average parities and numbers surviving as coded by enumerators. In the first variant, discussed here, we exclude all 'missing' cases from the denominator only. In effect, this makes the assumption that those women whose data are missing are the same as those for whom it is not. Unreasonable as this assumption may be, the results are presented below for completeness.

Table 5.3 Estimated average CEB and CS, based on raw data, missing excluded from the denominator

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.030	0.203	0.767	1.479	2.305	3.072	3.591	3.946
	Coloureds	0.025	0.201	0.763	1.434	2.057	2.535	2.838	3.065
	Indians	0.018	0.068	0.388	1.084	1.803	2.252	2.457	2.572
	Whites	0.033	0.065	0.315	0.883	1.511	1.924	2.108	2.205
CS									
	Africans	0.043	0.306	0.911	1.553	2.312	3.008	3.443	3.696
	Coloureds	0.038	0.315	0.930	1.540	2.112	2.551	2.823	3.013
	Indians	0.044	0.162	0.690	1.371	1.971	2.356	2.544	2.663
	Whites	0.051	0.113	0.479	1.107	1.663	2.017	2.172	2.263
Ratio									
	Africans	1.421	1.505	1.189	1.050	1.003	0.979	0.959	0.937
	Coloureds	1.518	1.569	1.219	1.074	1.027	1.006	0.995	0.983
	Indians	2.507	2.366	1.780	1.265	1.093	1.046	1.035	1.035
	Whites	1.543	1.722	1.520	1.254	1.100	1.048	1.030	1.026

As would be expected (since the number of women with no CS data are likely to exceed the number with no CEB data), the ratios are clearly implausible, implying that (for the most part) more than 100 per cent of those born survive.

The second possible assumption regarding the missing data is to include them in the denominator. This, then, is the same as making the assumption that all women whose children ever born data are missing, and all those whose data on children surviving are missing have no children ever born and no children surviving, respectively (Table 5.4).

Table 5.4 Estimated average CEB and CS, based on raw data, missing included in the denominator

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.020	0.151	0.643	1.331	2.142	2.895	3.378	3.665
	Coloureds	0.014	0.129	0.606	1.273	1.903	2.382	2.670	2.852
	Indians	0.011	0.047	0.310	0.965	1.684	2.133	2.328	2.411
	Whites	0.016	0.037	0.235	0.763	1.380	1.783	1.961	2.034
CS									
	Africans	0.015	0.139	0.607	1.264	2.035	2.727	3.129	3.319
	Coloureds	0.011	0.121	0.584	1.235	1.848	2.305	2.562	2.698
	Indians	0.009	0.041	0.295	0.934	1.642	2.080	2.262	2.335
	Whites	0.012	0.033	0.220	0.739	1.343	1.737	1.905	1.968
Ratio									
	Africans	0.759	0.922	0.943	0.950	0.950	0.942	0.926	0.906
	Coloureds	0.777	0.937	0.964	0.970	0.971	0.968	0.960	0.946
	Indians	0.816	0.860	0.954	0.968	0.975	0.975	0.972	0.969
	Whites	0.763	0.899	0.938	0.969	0.973	0.975	0.971	0.968

These data are more plausible than those presented in the previous table, as they never exceed one, and show (at least after age 25) roughly consistent differentials in proportions surviving by age and population group. Two factors, however, make the use of these data inadvisable. First, scrutiny of the raw data shows (particularly at younger ages) examples of improbable numbers of children born and children surviving to individual women are reported. At the very youngest ages, where childbearing is rare, a single woman reporting 90 or more births can distort the average parities very easily. (A simple example of this can be found among White women aged 12-14: Based on the unedited (but weighted) responses, 1 110 children have been born to these women, 131 (11.8 per cent) of which arise from a single woman who reported 99 births, possibly as a result of the enumerator using 99 for "unknown"). Second, the assumption will obviously bias the proportions both born and surviving downwards, meaning that even if plausible ratios are derived, the underlying parity and survival data are no longer remotely consistent with our best estimates of actual average parities, and hence cannot be relied upon.

Neither of the approaches discussed in this section are recommended for further analysis: they are included here simply for reference purposes.

5.1.2 Removal of implausible parities from the CEB and CS data

A second approach would be still to work with the raw data, but systematically limit the possible range of CEB and CS that women of different ages could plausibly have. To avoid imposing unnecessary constraints on these data, the limits chosen were at the outer reaches of biological plausibility, increasing from 5 to women aged 12-14, 9 to women aged 15-19 through to 22 for women older than 35. Cases where more than the 'permissible' numbers were recorded were set to zero at each parity, and the number of missing cases increased correspondingly so as to keep the total numbers constant. In effect, then, implausible answers are recoded as 'missing'. These missing cases are excluded from the denominator in the derivation of average parities (Table 5.5).

Table 5.5 Estimated CEB and CS, based on raw data with implausible parities removed, missing excluded in the denominator

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.016	0.195	0.762	1.477	2.304	3.071	3.589	3.943
	Coloureds	0.016	0.196	0.761	1.434	2.056	2.534	2.837	3.064
	Indians	0.012	0.062	0.385	1.084	1.803	2.252	2.457	2.572
	Whites	0.016	0.059	0.306	0.880	1.508	1.923	2.107	2.204
CS									
	Africans	0.026	0.297	0.907	1.552	2.311	3.008	3.442	3.694
	Coloureds	0.029	0.309	0.928	1.540	2.110	2.550	2.823	3.013
	Indians	0.031	0.149	0.688	1.371	1.968	2.356	2.544	2.660
	Whites	0.030	0.102	0.470	1.106	1.659	2.016	2.172	2.262
Ratio									
	Africans	1.627	1.521	1.190	1.050	1.003	0.980	0.959	0.937
	Coloureds	1.773	1.581	1.221	1.074	1.026	1.007	0.995	0.983
	Indians	2.468	2.420	1.786	1.266	1.091	1.046	1.035	1.034
	Whites	1.856	1.738	1.535	1.256	1.100	1.049	1.031	1.026

From these data we notice that at the older ages, the effect on average numbers born and surviving (relative to Table 3) is trivial, as would be hoped, while that at younger ages is large (again, as would be hoped). However the data are still incoherent, producing survival ratios of greater than 100 per cent. Clearly this is not a reasonable approach to adopt in the circumstances. The main reason for showing these data is simply as an intermediate stage towards the next possible solution to the problem of poor quality data on children born and children surviving.

5.1.3 Removal of inconsistent parities, and use of an el-Badry correction applied to the missing data

Rather than simply including or excluding the missing data en masse from the data, an el-Badry correction can be applied to the data on parity to apportion the reported missing data between "true" childless and "true" missing states. This correction has an impact on the data on children surviving, since women with no children ever born must, by definition, have no children surviving. Hence, an el-Badry correction was applied to the parity data, and the revised estimates of childless women were used as a constraint in the estimation of women with no reported children surviving (thereby allowing some women whose children surviving were captured as 'missing' to be reallocated to zero). As with all applications of the el-Badry correction, the residual 'missing' data are excluded from the denominator in all average parity or average children surviving calculations.

Table 5.6 Estimated average CEB and CS, based on raw data with implausible parities removed, el-Badry correction applied and missing excluded from the denominator

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.011	0.150	0.664	1.382	2.225	3.008	3.508	3.806
	Coloureds	0.009	0.128	0.614	1.295	1.935	2.422	2.715	2.901
	Indians	0.008	0.044	0.317	0.994	1.735	2.197	2.397	2.483
	Whites	0.008	0.035	0.241	0.802	1.452	1.879	2.068	2.144
CS									
	Africans	0.009	0.141	0.639	1.338	2.148	2.874	3.297	3.501
	Coloureds	0.008	0.121	0.601	1.276	1.906	2.377	2.643	2.789
	Indians	0.006	0.039	0.306	0.979	1.718	2.178	2.371	2.445
	Whites	0.007	0.032	0.230	0.789	1.438	1.863	2.047	2.114
Ratio									
	Africans	0.869	0.938	0.963	0.968	0.965	0.955	0.940	0.920
	Coloureds	0.908	0.948	0.978	0.985	0.985	0.981	0.973	0.961
	Indians	0.729	0.881	0.967	0.985	0.990	0.992	0.989	0.985
	Whites	0.918	0.908	0.954	0.984	0.990	0.992	0.990	0.986

These data, which we term Minimal Edit data, seem more reasonable than any presented earlier. However, these data, too, cannot be correct insofar as they demonstrate lower proportions of children surviving (and hence higher child mortality) among White and Indian women aged 15-24 than among African and Coloured. While it may be possible, our understanding of the historical and current differentials in access to medical care, living conditions and poverty by population group make this result exceedingly unlikely, especially in the light of the very high survival ratios shown in these two groups at older ages. In section 5.2 we compare these data with the edited data (as per the edits produced by Statistics South Africa) and the 1996 census data.

5.1.4 Data edited according to the procedures adopted in the report on fertility

Table 5.7 shows the effect of editing the data according to the procedures adopted to derive a reasonable estimate of fertility from the 2001 census data (Moultrie and Dorrington 2004). The approach adopted was to examine only those data for women for whom none of the responses to the fertility questions had been subjected to editing, imputation or hotdecking. In the case of the work on fertility, this resulted in obviously low estimates of the level of fertility (which could be corrected for), but a very good approximation to the distribution of fertility by age.

Two versions of the children ever born data are presented. The first shows that estimated from the 2001 census data, while the second shows that estimated using the 1996 parity data and the estimates of recent fertility derived from the 2001 census data. We must stress that we believe that the second series is the 'better' of the two as the first is logically incompatible with that from the 1996 census.

No equivalently 'revised' series can be derived for the children surviving data, since this would require estimates of child mortality over the intercensal period. Similarly, no revised estimate of average parity for the youngest age group could be derived either, since this would have required data to estimate youngest-young fertility in the intercensal period. These data are not available, and equally assuming no youngest-young fertility is unreasonable.

Table 5.7 Estimated average CEB and CS, based on data edited according to the procedures adopted in the fertility report, original and revised parity estimates

CEB (original)	*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
Africans	0.002	0.122	0.608	1.336	2.184	2.937	3.367	3.591
Coloureds	0.002	0.108	0.565	1.232	1.865	2.343	2.623	2.795
Indians	0.000	0.028	0.282	0.951	1.693	2.168	2.374	2.478
Whites	0.000	0.018	0.200	0.747	1.384	1.790	1.997	2.128
CEB (revised)								
Africans		0.175	0.669	1.464	2.261	3.068	3.629	4.083
Coloureds		0.147	0.574	1.325	1.972	2.509	2.885	3.197
Indians		0.064	0.328	1.043	1.858	2.288	2.465	2.674
Whites		0.041	0.249	0.811	1.540	1.950	2.113	2.233
CS								
Africans	0.002	0.120	0.604	1.324	2.156	2.882	3.262	3.423
Coloureds	0.002	0.107	0.566	1.247	1.886	2.358	2.612	2.737
Indians	0.000	0.028	0.281	0.947	1.684	2.152	2.353	2.444
Whites	0.000	0.018	0.199	0.744	1.377	1.779	1.979	2.101
Ratio (original CEB)								
Africans	0.977	0.989	0.993	0.991	0.987	0.981	0.969	0.953
Coloureds	0.979	0.990	1.002	1.012	1.011	1.006	0.996	0.979
Indians	0.737	0.995	0.995	0.995	0.995	0.993	0.991	0.986
Whites	1.000	0.996	0.996	0.995	0.995	0.994	0.991	0.987
Ratio (revised CEB)								
Africans		0.686	0.903	0.904	0.953	0.939	0.899	0.838
Coloureds		0.727	0.986	0.941	0.957	0.940	0.905	0.856
Indians		0.440	0.855	0.908	0.907	0.941	0.955	0.914
Whites		0.434	0.800	0.917	0.894	0.912	0.936	0.941

The first set of ratios (those based on the original CEB) are not plausible, indicating survival ratios very close to or greater than one, and a generally too-high probability of child survival at the older mothers' ages. (By the same token, however, the closeness of these results to unity does point towards women systematically reporting their number of children ever born as their number of children surviving at the time of enumeration. Such an error could easily have been introduced as a result of inadequate or incorrect enumerator training). The second series indicates a schedule of child survival that is plausible at the oldest ages (roughly credible ratios and differentials by population group) but which is unusable at the younger ages (showing again significantly worse child survival among young White women than among any other population group). In any event, it is unlikely that only four fifths of children born to young white women are still alive.

Hence, thus far, none of the procedures adopted have resulted in a series of estimates of child survival ratios that is inherently plausible. The next section looks at the estimates arising from Statistics South Africa's editing rules.

5.1.5 Data edited according to Statistics South Africa's editing rules

The estimates of average parity and average numbers of children surviving are shown in the table below.

Table 5.8 Estimated average CEB and CS, based on Statistics South Africa's edits

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.009	0.185	0.739	1.452	2.278	3.049	3.571	3.923
	Coloureds	0.011	0.195	0.737	1.405	2.029	2.510	2.813	3.037
	Indians	0.005	0.050	0.345	1.036	1.770	2.228	2.428	2.550
	Whites	0.006	0.046	0.264	0.837	1.479	1.899	2.089	2.187
CS									
	Africans	0.008	0.180	0.713	1.398	2.183	2.892	3.328	3.576
	Coloureds	0.011	0.193	0.727	1.383	1.994	2.456	2.731	2.908
	Indians	0.005	0.049	0.342	1.029	1.757	2.208	2.401	2.511
	Whites	0.006	0.045	0.261	0.831	1.468	1.883	2.066	2.153
Ratio									
	Africans	0.964	0.972	0.965	0.963	0.958	0.949	0.932	0.912
	Coloureds	0.979	0.987	0.987	0.984	0.982	0.979	0.971	0.958
	Indians	0.984	0.991	0.992	0.993	0.992	0.991	0.989	0.985
	Whites	0.964	0.993	0.992	0.993	0.993	0.991	0.989	0.984

At first glance, these data appear to be the most consistent of them all. They indicate Africans have lower proportions of children surviving (at least among women over 14), and clearly lower survival ratios at older ages.

However, we have already argued that the average parities indicated by the census data have been derived according to edit rules that are hard to justify, that the edits try to apply adjustments at an individual level that could have been made better by demographers applying methods at an aggregated level, and that the extent of the edits applied to the data make them unreliable (Moultrie and Dorrington 2004). In addition, the ratios indicated by these data do not show the well-documented decline in the proportion of children surviving to younger (below 20) mothers[6]. By contrast, the data presented above in respect of Africans shows a monotonic decline from age 15 for Africans and Coloureds, and no discernible dip for Indians and Whites. In

[6] The reasons for the existence of the lower proportion surviving at younger ages are many: *inter alia* poorer maternal health and child-care skills; fewer children ever borne, meaning that a single child death can reduce the proportion surviving to 50 per cent or zero with great ease

addition, these proportions surviving are inconsistent with (being far higher than) those used to derive estimates of child mortality from the data in the 1996 census, as discussed below.

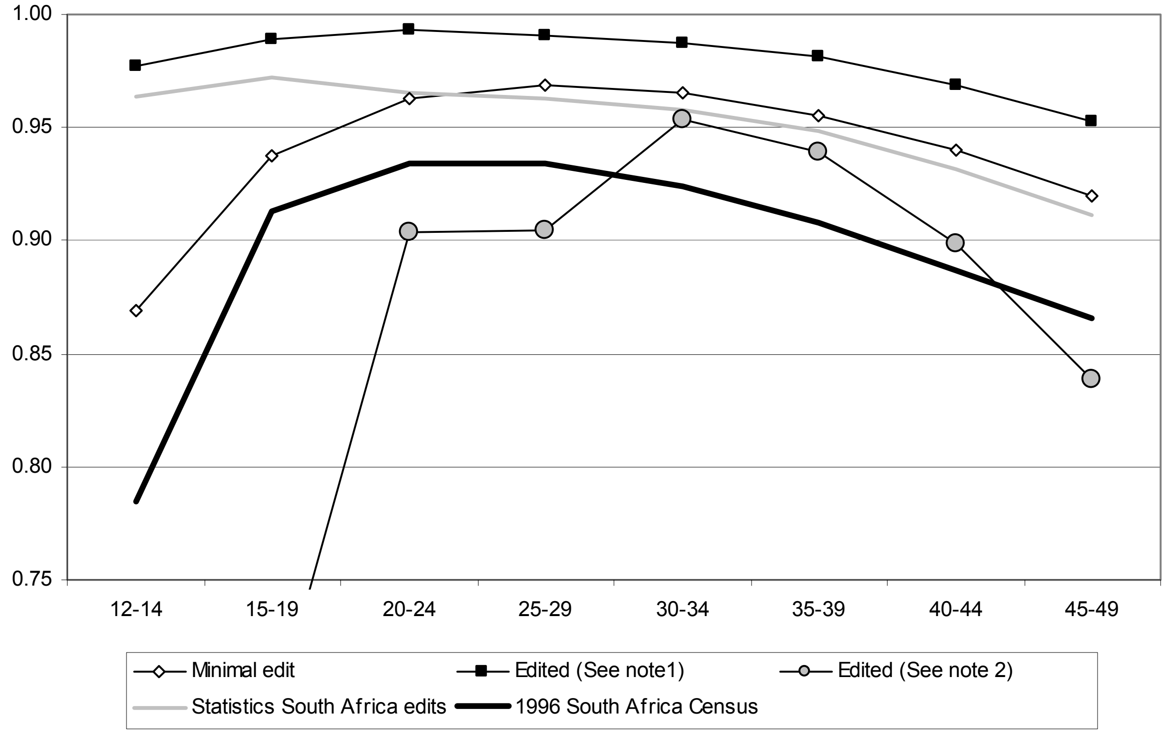
5.2 Data from the 1996 census

Our best estimates of the numbers of children born and surviving from the 1996 census data are shown in the table below, and are derived from previous work prepared by staff and associates of the Centre for Actuarial Research (CARe – see particularly Moultrie and Timæus (2002; 2003) and Dorrington, Moultrie and Nannan (2002)).

Table 5.9 Estimated average CEB and CS, based on 1996 census data

CEB		*12-14*	*15-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*
	Africans	0.024	0.164	0.743	1.554	2.496	3.267	3.922	4.333
	Coloureds	0.023	0.136	0.679	1.389	2.106	2.681	3.123	3.597
	Indians	0.021	0.047	0.417	1.198	1.940	2.320	2.609	2.782
	Whites	0.012	0.045	0.287	0.918	1.577	1.963	2.154	2.309
CS									
	Africans	0.019	0.150	0.694	1.452	2.305	2.965	3.479	3.751
	Coloureds	0.019	0.127	0.648	1.340	2.040	2.569	2.919	3.274
	Indians	0.018	0.041	0.400	1.168	1.895	2.248	2.516	2.643
	Whites	0.011	0.040	0.276	0.894	1.535	1.902	2.085	2.215
Ratio									
	Africans	0.785	0.913	0.934	0.934	0.924	0.908	0.887	0.866
	Coloureds	0.845	0.931	0.955	0.965	0.969	0.958	0.935	0.910
	Indians	0.859	0.872	0.960	0.975	0.977	0.969	0.964	0.950
	Whites	0.845	0.893	0.964	0.975	0.973	0.969	0.968	0.959

Figure 5.1 Ratio of children surviving to children ever born showing effect of different editing approaches and compared with equivalent data from 1996 South Africa Census, African South African women by age



Notes 1: These data are those for Africans presented in the first part of Table 5.7, based on a best estimate of the average numbers of children surviving and children born derived from data not subjected to editing or hotdecking.

2: These data are those for Africans presented in the second part of Table 5.7, based on corrected estimates of average children ever born based on average parities in the 1996 census and best estimates of the intercensal trends in fertility.

In contrast to the edited data from the 2001 census, the 1996 data are far more curved, having a maximum in all population groups between 25 and 34. The proportions reported surviving in 1996 are also much lower than the edited data from 2001 (and in fact all other data presented with the exception of the data derived using the corrected parities and algorithm used in the fertility report presented in the second panel of Table 5.7.)

The proportions surviving for African women by age derived using all the methods presented above are shown in Figure 5.1. The solid dark line represents the data from the 1996 census, while the solid grey line represents those from the 2001 census data. Ignoring for the time being the huge inconsistency between these data series below age 20, it might be tempting to ascribe the remaining differences to improved mortality. However, this is very unlikely in an era of HIV/AIDS. Also the implied 2.8 per cent absolute increase in survival ratios for African women aged 40-44 in 1996 relative to women aged 45-49 in 2001 (the same cohort) is only mathematically possible if there has been at least a 27 per cent increase in the total number of births to women in this cohort over the five year period with ALL of those additional births surviving. At lower rates of survival (say at 95 per cent of new births surviving) the increase in the total (lifetime) number of children ever born would have to be of the order of 60 per cent. Biologically, this is impossible. Thus, as with the data used in the estimation of fertility, relative to the 1996 data we must conclude that the edited data available from which to calculate estimates of child mortality are implausible too.

5.3 Conclusion

The analyses presented in this chapter lead to the disappointing conclusion that there is no way to use consistently derived and plausible estimates of the women's average numbers of children ever born and surviving to estimate child mortality with any credibility. The unedited data cannot be used, because they are distorted by biologically impossible answers. Removing these responses from data on both children ever born and children surviving results in estimated proportions of children surviving that cannot be made consistent with the equivalent data from the 1996 census or the 1998 DHS, despite the fact that the Minimal Edit data produces an age-pattern that most closely mimics that from 1996.

The proportions of children surviving arising from Statistics South Africa's edits closely follow the Minimal Edit results at older ages, but clearly overstate the proportion of surviving children born to younger women and are logically inconsistent with the results from the 1996 census. However, when our best estimates of average numbers of children ever born are used, estimates which attempt to adjust for underreporting of children ever born, the resulting survival ratios are clearly of little worth.

In aggregate, the results presented here point to significant flaws in the data collected on women's lifetime fertility in the 2001 census. Certainly, women tended to confuse the two questions indicating a weakness in enumerator training. However, this clearly is not the only error present in these data.

That it is not possible to derive reasonable, reliable and useful estimates from the 2001 South African census is not a conclusion that is easily reached. First, on epistemological grounds, it is impossible to prove that such estimates cannot ever be derived. Second, on policy grounds, a failure to derive estimates of child mortality leaves a gaping hole in our understanding of current population dynamics in the country, and will severely hamper efforts aimed at improving the quality of life of all South Africans.

We have documented here what we think are the most plausible approaches that could have been taken to derive estimates of child mortality from the 2001 census data. None of them have withstood close scrutiny. If nothing else, we hope that other demographers are able to take the work presented here and, cognisant of our findings, draw more useful conclusions from these data. We would be more than willing to contribute to such efforts, and to work with interested parties in an attempt to prove that our conclusions here have been too hastily drawn.

If we are indeed correct in our conclusions, the analysis presented here provides little in the way of comfort to demographers. What benefit there is lies in the clear lesson taught that huge and sustained efforts must be made to avoid any repeat of problems of this nature in future South African censuses.

Census instruments must be carefully evaluated; training of enumerators must be overhauled; editing and imputation rules must be subjected to the closest analysis. This is no easy task, but is one to which we are committed to contributing our skills and expertise.

6 Trends in mortality over time

Prior to 1994, official life tables for the Coloured, Indian and White populations were produced by the central statistical organization for the three-year periods centred on 1946, 1951, 1960, 1970, 1980 and 1985 (abridged life tables up to 1970 from van Eeden and van Tonder (1975), thereafter from Central Statistical Services (1985; 1987)). In addition, tables were also produced for the White population for three years centred on 1921 and 1936, and for the Coloured population for 1936. No official life tables were produced during this period for the African population, since deaths of Africans were only captured as part of the vital registration system from 1979 and even then were presumed to be so incomplete as to be useless (although Dorrington (Dorrington 1989, 1998; Dorrington, Bradshaw and Wegner 1999) has shown otherwise).

Population group ceased to be recorded on the death certificate in 1991, which together with the assumptions that the completeness of death registrations was too low and suspicions about the accuracy of the census count, meant that no life tables were produced for the period 1990-92.

Following the 1996 census, Statistics South Africa published official abridged life tables for two periods; 1985-1994 and 1996 (Statistics South Africa 2000). Tables for the first period were produced by population group (including a table for 'other and unspecified' population group) but not by province. Those for the second period were produced by province, but not population group.

Prior to the late 1990s mortality rates for the African population (and hence, though it was never represented as such, the population as a whole) were derived largely from survival factors that resulted from the reconstruction of census populations (Mostert, van Tonder and Hofmeyr 1987; Sadie 1970; 1970; 1973; 1988; 1993).

6.1 Adult mortality

Figure 6.1 compares the estimates of ${}_{35}q_{30}$ [7] from Table 4.2 and Table 4.3 for all population groups combined with those from four other sources. The first (Official+), covering the years from 1946 through to 1985, is a weighted average of the official South African Life Tables (SALT) for Coloureds, Indians and Whites, with estimates for the African population from Sadie (1973) as presented in van Eeden and van Tonder (1975) up to 1970, and Sadie (1988) for 1980 and 1985. Essentially the SALT tables were derived directly from the registered deaths and census estimates of the population. The rates for Africans were derived from the survival probabilities needed to reconcile estimates of the African population derived after correcting the census counts to conform to expectations.

The second, covering the more recent past from 1985 to 2000, comes from Timæus, Dorrington, Bradshaw *et al.* (2001). This paper presents the results of analysing all sources of data

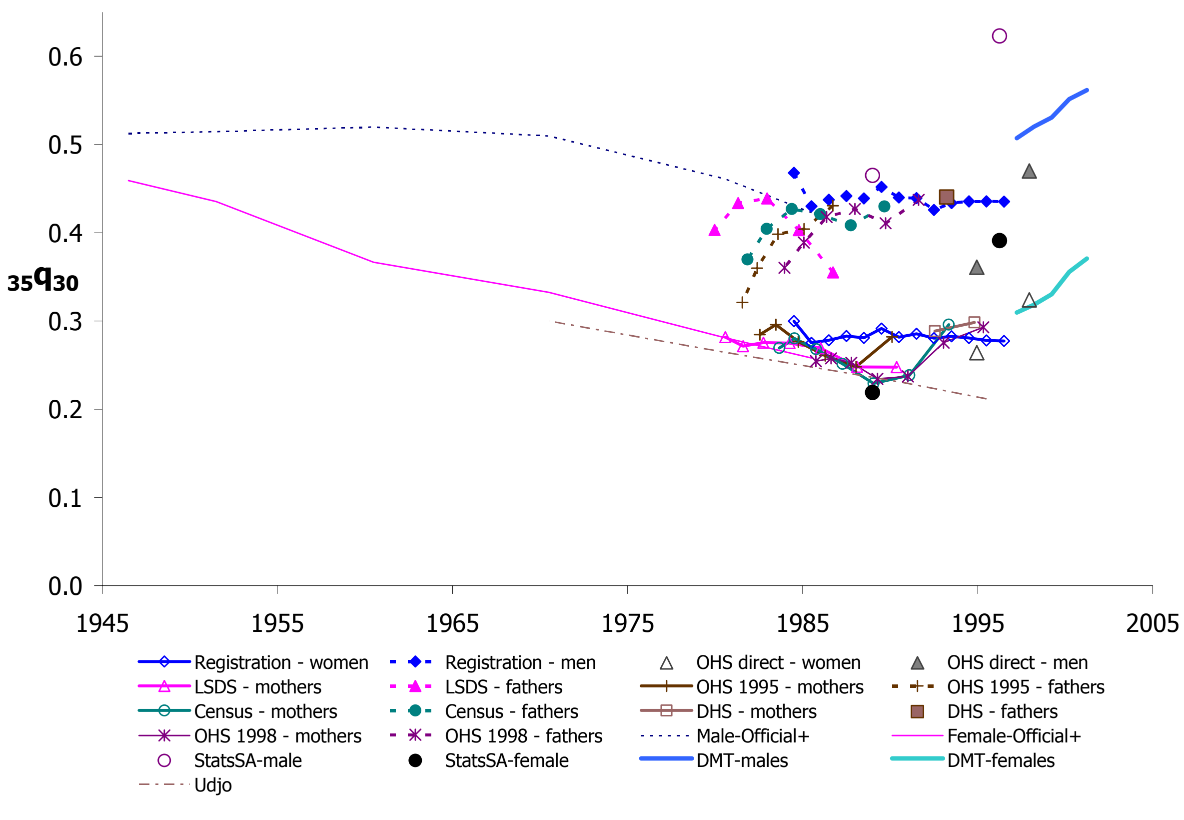
[7] The age intervals of these comparisons were chosen to correspond with those of Timæus, Dorrington, Bradshaw et al .

reflecting national mortality rates between 1980 and 2000, and as such is the most comprehensive of work on mortality in the two decades preceding the 2001 census.

The third source is the estimates based on the official abridged life tables published by Statistics South Africa (2000) covering two distinct periods namely one estimate (plotted as a point at the start of 1989) for the period 1985-1994 and one for the calendar year 1996. The first estimate is based on analysis of data on survival of parents and the proportion of children surviving out of children ever born from the 1996 census, the second on application of Brass's Growth Balance technique (Brass 1975) to registered deaths and the 1996 census population.

The final set of estimates are those produced by Udjo (1999) using the same data as was used by Statistics South Africa to produce the abridged life tables for 1985-1994. However, Udjo concluded that although the trend in male adult mortality was plausible, the level appeared to be too high to be used to produce reasonable life tables. Instead, he assumed male life expectancy at birth to be some six years below that of females. Unfortunately he did not publish his life tables and so we are only able to derive estimates of ${}_{35}q_{30}$ for females from his published estimates of alpha and beta.

Figure 6.1 Trends in the probability of dying between ages 30 and 65, South Africa



From Figure 6.1 we can observe the following:

The Statistics South Africa estimate for 1985-94 is too low for women, while it is difficult to say anything about the estimate for men. On the other hand, the estimates for 1996 are completely at odds with all the other estimates and must be rejected as being far too high.

As might be expected, Udjo's estimates are broadly similar to those from the abridged life tables around 1989. However, by and large the trend for earlier years is at odds with the 'official' estimates and the trend for later years at odds with all the other estimates.

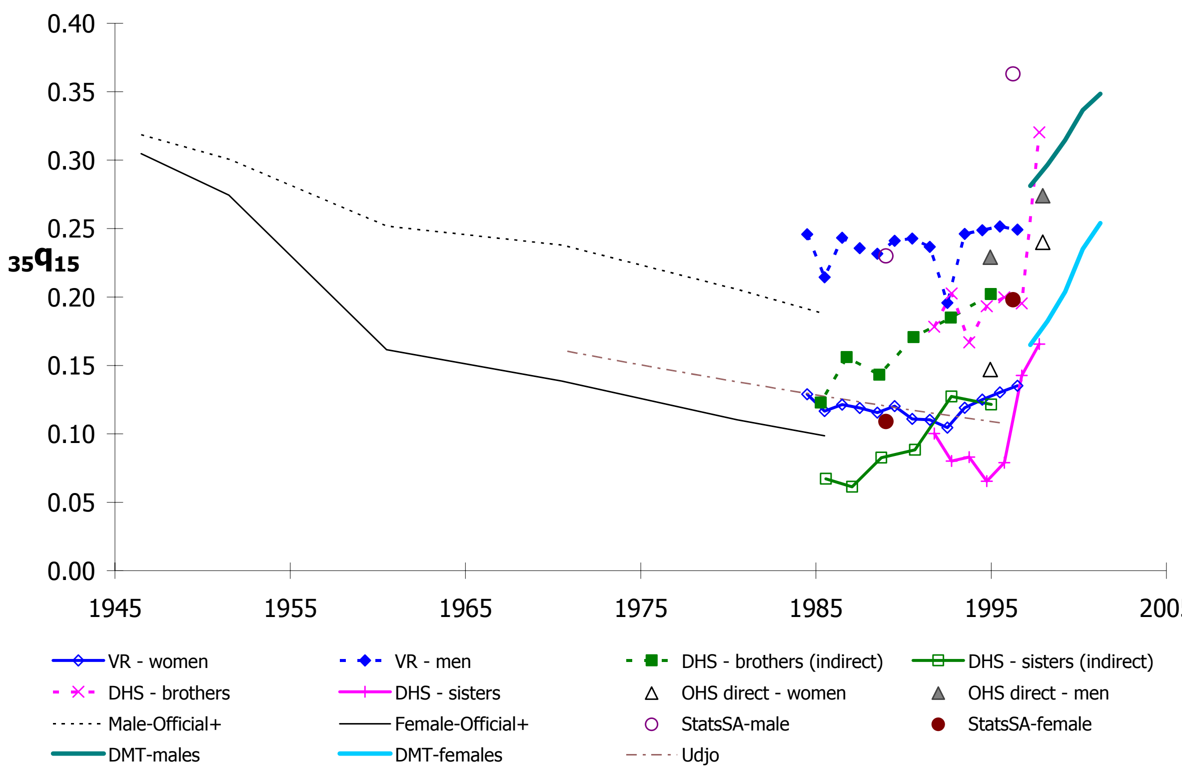
The 'official' estimates for 1985 appear to be a little on the low side.

Female mortality at ages 30 to 64 fell between 1945 and 1980 but appears to have been constant between 1980 and 1995, after which there has been a steady increase in mortality. Our estimates appear entirely consistent with those of Timæus, Dorrington, Bradshaw *et al* (2001) both in terms of level and trend.

As far as the rates for males are concerned these remained fairly constant until about 1970 then started falling but this trend was reversed in the early 1980s. However, as there is a discrepancy between our rates and those of Timæus *et al*, it is unclear whether rates between 1980 and 1995 remained level or whether (and this is the version supported by the results of this investigation) there had been a slight increase in rates since 1980 (due possibly to an increase in deaths due to violence and external causes) to be consistent with our rates.

The comparison in Figure 6.2 is the same as that in Figure 6.1 except it refers to the probability of dying in early adulthood (between ages 15 and 50).

Figure 6.2 Trends in the probability of dying between ages 15 and 50, South Africa



From Figure 6.2 we see the following:

Mortality fell between 1945 and 1985, first rapidly and then less rapidly. However, again the rates for 1985 appear to be on the low side.

In this age range Udjo's estimate appear to be quite consistent with those from vital registration data from 1985 to 1990, but again his rates for earlier years are at odds with the 'official' estimates.

There is some consistency both in level and trend in rates between our rates and those of Timæus *et al*, which lends a certain credence to our estimates. We can conclude that, consistent with the trends observed above, mortality in this age range remained fairly constant for women from 1980 to the mid-1990s and then started to rise rapidly, while for men it has been rising gradually since 1980 and more sharply since the late 1990s. The recent rise in the rates for females has been more rapid than that for males.

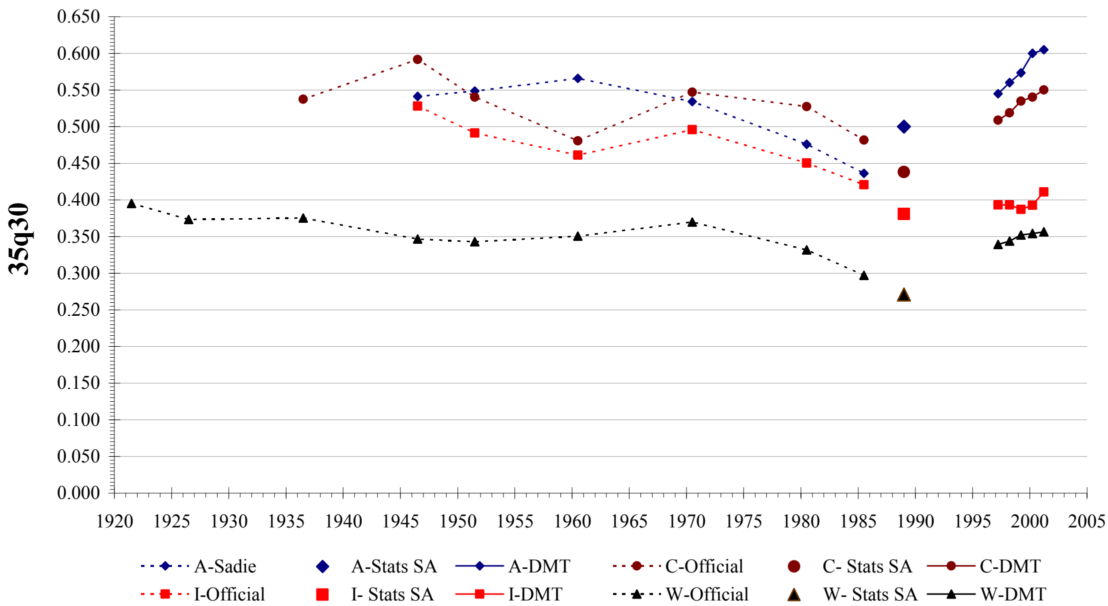
The Statistics South Africa rates for 1985-94 are quite consistent with this trend, but those for 1996 are once again far too high.

Figure 6.3 compares our rates for the four population groups with the official estimates (SALTs and the abridged life tables for 1985-94), and with those of Sadie for Africans prior to 1985.

From Figure 6.3 we can observe the following:

While it might be tempting to extrapolate the trend in the SALT to connect with estimates from the abridged life tables for Coloureds, Indians and Whites, this would produce an implausibly rapid change in trend in rates for both the Coloured and White population groups if we accept our estimates. Thus it might be more plausible to argue that both the SALT and the abridged life table values are too low for these groups and mortality has since 1980 remained roughly constant for White men, fallen to a plateau for Indian men, and fallen and then risen again for Coloured men. In keeping with this it would thus appear that mortality of African men has been rising since 1980, more rapidly since 1995.

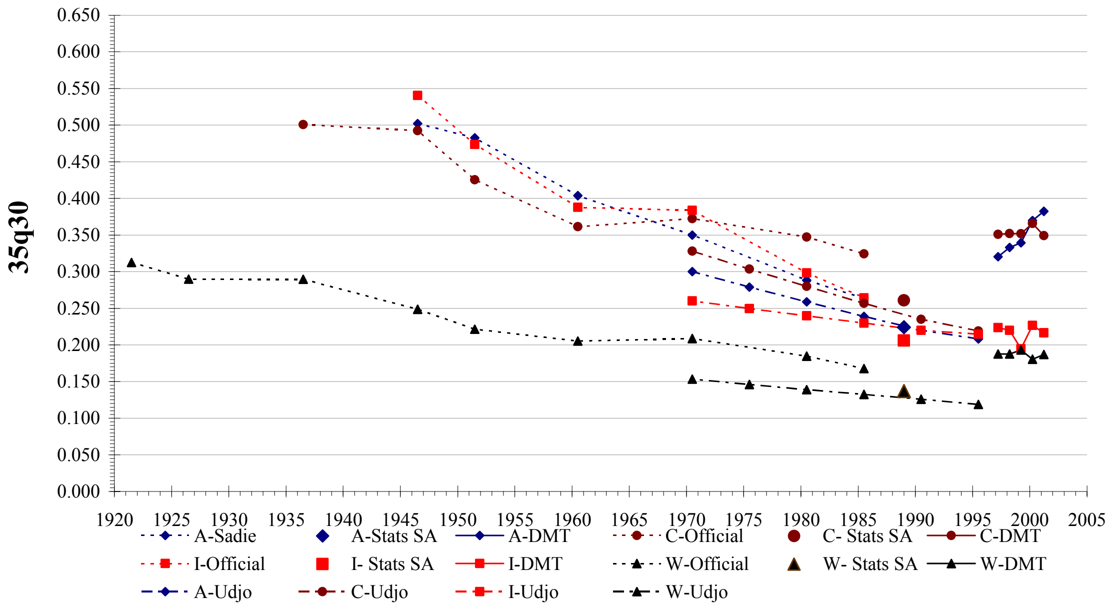
Figure 6.3 Trends in the probability of dying between ages 30 and 65: Males by population group



Note: A-African; C-Coloured; I-Indian; W-White; DMT-this paper

Figure 6.4 s the same comparison as Figure 6.3 but for women. It also includes estimates derived from Udjo (1999).

Figure 6.4 Trends in the probability of dying between ages 30 and 65: Females by population group



Note: A-African; C-Coloured; I-Indian; W-White; DMT-this paper

From Figure 6.4 we observe the following:

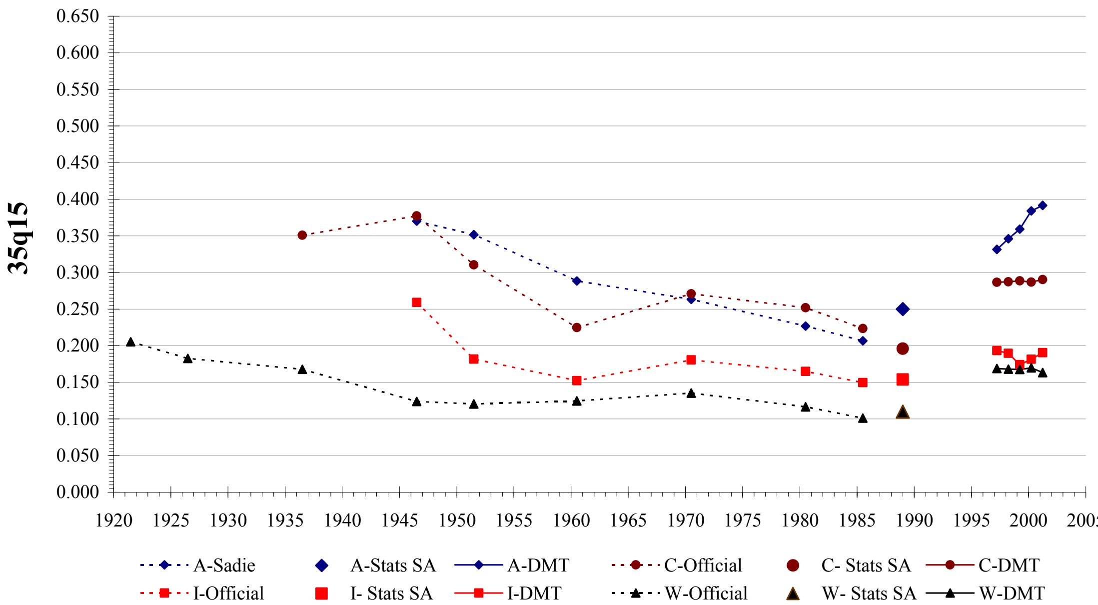
We can draw similar conclusions about the patterns of falling mortality in the past, with the possibility that the estimates from the SALT and abridged life tables, as well as those from Sadie are too low in the middle to late 1980s. In the case of women we would conclude that our estimates suggest that mortality has remained roughly constant since 1980 for White and Coloured women, fallen to a plateau for Asian women, and after remaining constant to about 1995 started rising rapidly for African women.

Once again Udjo's estimates are broadly consistent with those derived from the Statistics SA abridged life tables (as might be expected) and hence subject to similar criticisms. However, again his estimates for the earlier and later years are at odds with the results of others. In addition his estimates for African, Coloured and Indians seem implausibly close.

Figure 6.5 and Figure 6.6 present the same comparisons to those in Figure 6.3 and Figure 6.4 but for the probability of dying between ages 15 and 50.

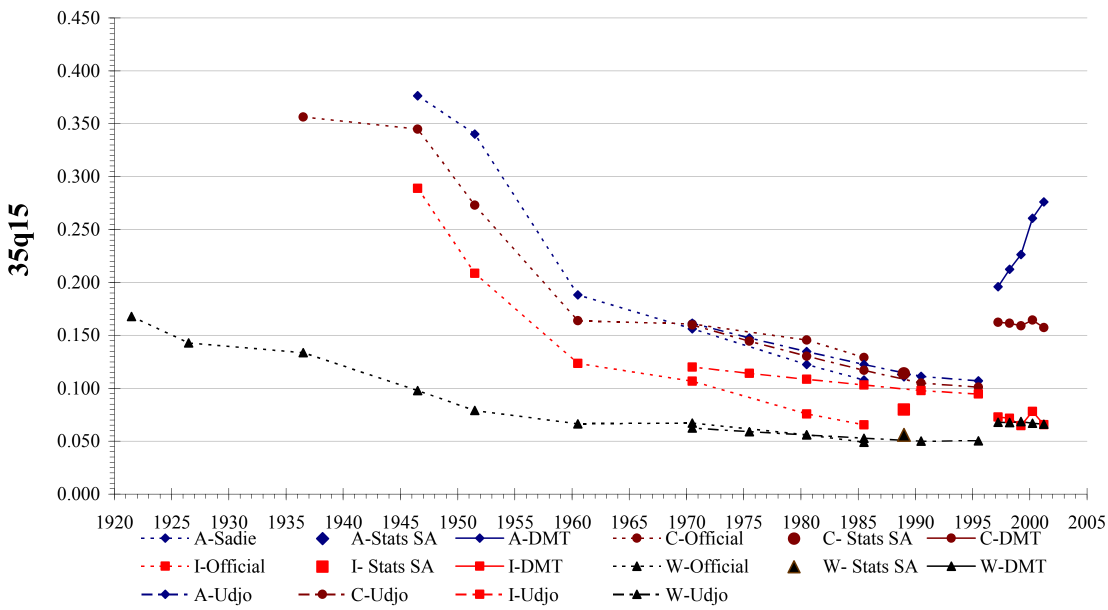
From Figure 6.5 and Figure 6.6 we can conclude that after falling until 1980, mortality of men in this age range appears to have increased for men in all population groups, significantly so for Africans. Whereas for women rates appear to have remained constant for the Coloured, Indian and White populations but to have increased for African women since 1980 with increasing rapidity since 1995. Once again the estimates based on the SALTs, Sadie's rates for Africans and Statistics South Africa's abridged life tables for 1985-94 appear to be too low with the possible exception of African men and White and Indian women. Again Udjo's estimates seem to be at odd with those of others.

Figure 6.5 Trends in the probability of dying between ages 15 and 50: Males by population group



Note: A-African; C-Coloured; I-Indian; W-White; DMT-this paper

Figure 6.6 Trends in the probability of dying between ages 15 and 50: Females by population group



Note: A-African; C-Coloured; I-Indian; W-White; DMT-this paper

6.2 Childhood mortality

Figure 6.7 presents rates of childhood mortality (the probability of dying between birth and age five) from a number of sources. The first of these are those based on a weighted average of the rates from the SALTs and Sadie as was described above for the adult rates. These cover the period 1946 to 1985.

The second are various estimates produced by Nannan, Timæus, Moultrie *et al* (2001). The estimates derived indirectly from the reports on the number of children ever borne and the number surviving from the 1996 census are on the high side. Below them, and coinciding with the indirect estimates based on data from the 1998 South African Demographic and Health Survey (Department of Health 2002), are estimates based on the census data corrected for the number of still births assumed to have been incorrectly included in the census responses[8]. Below that are the direct estimates from the SADHS data and slightly higher, an estimate based on these data but disregarding data from three provinces which are outliers.

The third set are those from Udjo (1999), which were derived from the children surviving and children ever born data from the 1996 census. Unfortunately there is some confusion about whether or not these estimates represent only girls or both sexes combined. The 1996 census did not collect data allowing calculation of child mortality by sex and Udjo does not say how he derived sex-specific rates but these estimates are published as part of a table indicating they represented female mortality.

Finally the figure includes the estimates based on the abridged life tables published by Statistics South Africa, and those from the ASSA2000 model have also been included.

From Figure 6.7 we can see that estimates based on the 1996 census data without correcting for the still births are likely to overestimate childhood mortality, which is the case with the estimates from the abridged life tables for the period 1985-94. On the other hand the estimate from the abridged life tables for 1996 appears to be too low, and together the two abridged life tables suggest an implausible fall in childhood mortality between the two periods.

Although it is difficult to know where within the range of estimates the true rates lie, and one's preference would probably lean towards the direct estimates from the DHS data, two further conclusions can be drawn from these comparisons. The first is that mortality rates fell through to about 1995, first rapidly and then less so, with the 'official' estimates being entirely consistent with those from Nannan *et al*. The second is that rates appear to reach a trough in the early 1990s and started to increase after that. This of course makes it impossible to project rates into the future and hence even more unfortunate that Census 2001 was unable to provide any useful data from which estimates of more recent mortality could be estimated.

[8] An estimate of the extent of this was derived by a comparison of the census data with those from the DHS, which was assumed through its more thorough set of questions to have excluded still births.

Figure 6.7 Trends in childhood mortality (the probability of dying before age 5)

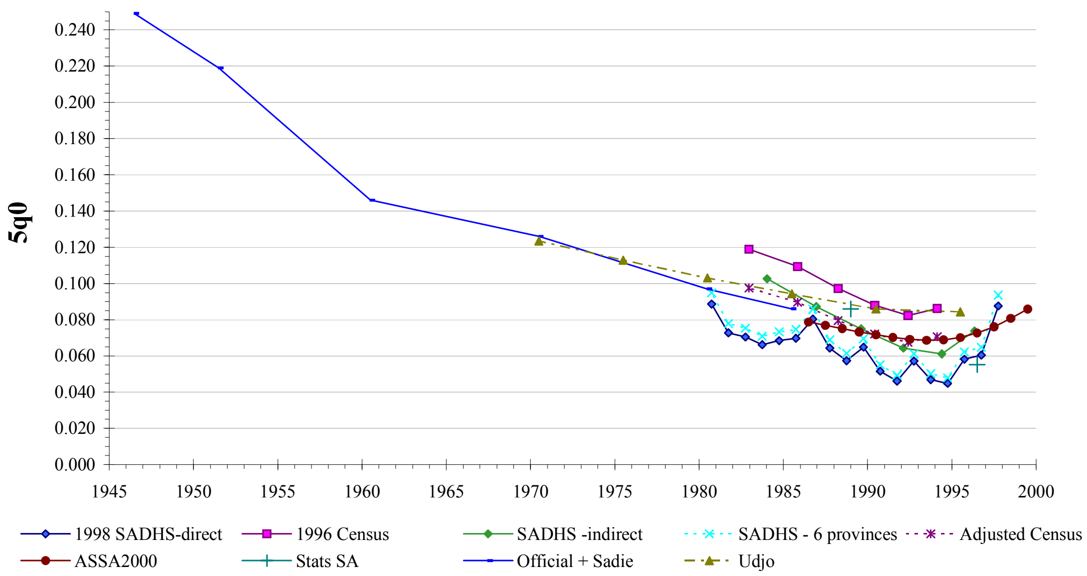


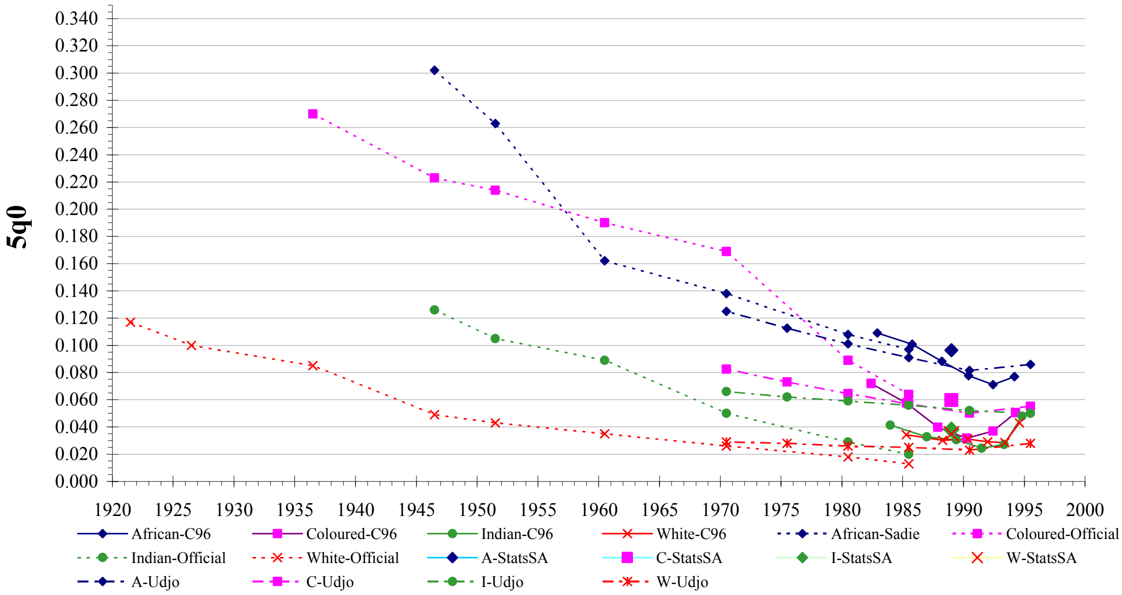
Figure 6.8 resents estimates by population group for the same sources as those presented in Figure 6.7. Here the results from Nannan *et al.* are based on the 1996 census corrected for still births.

From Figure 6.8 we observe a remarkable consistency between the rates derived by Sadie and the Coloured SALTs on the one hand and the rates derived by Nannan *et al.* However, such consistency does not exist in the case of the White and Indian populations. Although we cannot be too categorical about it, it would appear possible that the 1985 and possibly, in the case of the White population, the 1980, life tables underestimate childhood mortality. This conclusion is reinforced by a comparison of these rates with those of the population of England and Wales in 1980-82 (Office of Population Census and Surveys 1987), which showed the childhood mortality rates of Whites, as estimated by the SALT, to be as low, or lower, than those in England and Wales at around the same time, which does not seem too plausible.

Further, the up-tick in rates is found in all population groups, which is somewhat puzzling, and given the levels of prevalence of HIV observed in the different population groups is unlikely to be due to the virus in all cases. The estimates derived from Statistics South Africa's abridged life tables appear entirely consistent for Whites and Indians but too high for Africans and Coloureds (which probably is a result of not correcting for the inclusion of still births in the data from the census).

Udjo's estimates are surprisingly different from those of Nannan et al. While the difference for Africans could be explained by the fact that Udjo did not correct for the inclusion of still births, this cannot explain why his rates for Coloureds and Indians are so much higher, while his rates for Whites so much lower.

Figure 6.8 Trends in childhood mortality (the probability of dying before age 5), by population group



6.3 Provinces

As the provinces only came into being in 1994, Statistics South Africa's abridged life tables for 1996 were amongst the first attempts at estimating mortality by province. Figure 6.9 and Figure 6.10 compare our estimates (DMT) of ${}_{45}q_{15}$ for each of the provinces with those of (Dorrington, Moultrie, Nannan *et al.* In press) for 1996 and those derived from the official abridged life tables for 1996 (Statistics South Africa 2000). From these comparisons we see that not only are the official estimates too high, particularly in the case of males (given that mortality has been increasing rapidly over the past five years), but that there is not much consistency in ranking between the three sets of estimates. However, the points of consistency suggest that mortality is high in KwaZulu-Natal and the Eastern Cape, and low in the Western Province and Gauteng. Interestingly the correlation coefficient (R^2) of our estimates between men and women for various provinces is a high 70 per cent. The correlation coefficients (R^2) of our estimates of rates derived in this report and those derived by Dorrington *et al* (In press), derived entirely independently by a different method, are 69 per cent for men and 67 per cent for women.

Figure 6.9 Comparison of the probability of dying between the ages of 15 and 60 by province: Males

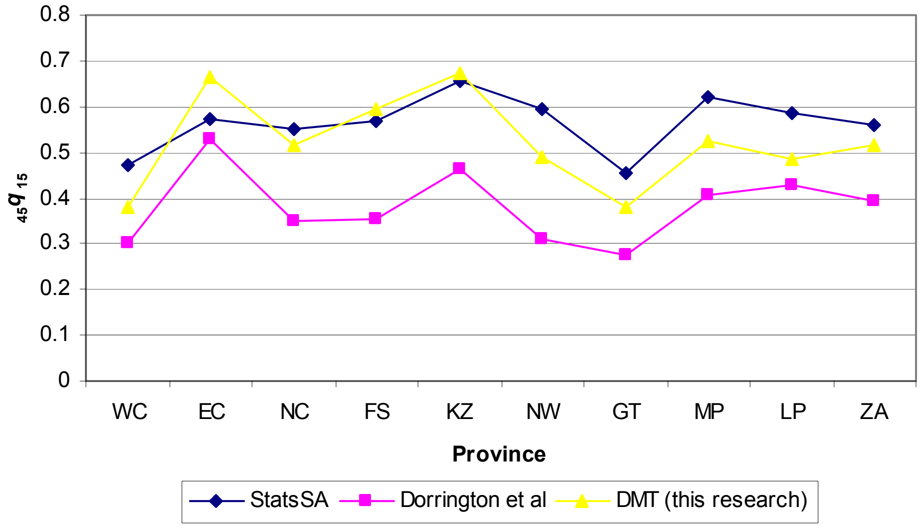
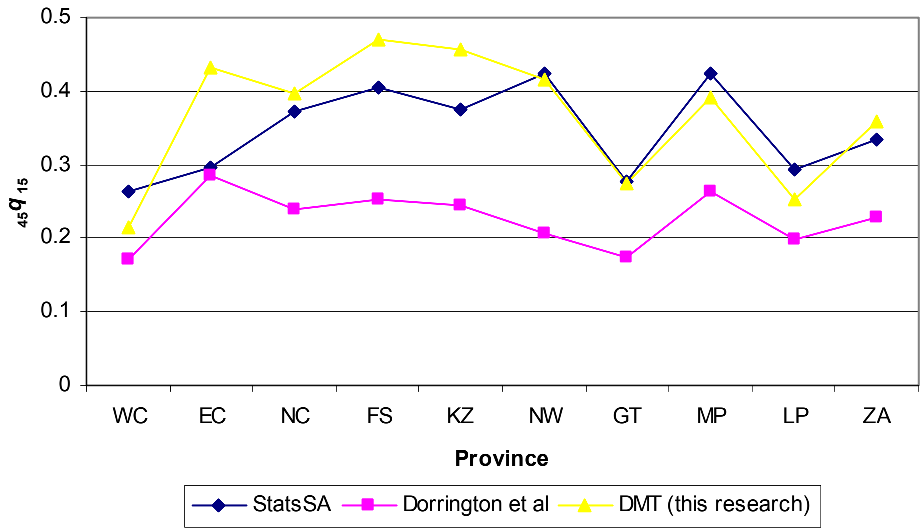


Figure 6.10 Comparison of the probability of dying between the ages of 15 and 60 by province: Females



6.4 Conclusions

With the possible exception of older African men and younger White men our estimates of adult mortality seem quite consistent with an overall trend in mortality which has seen rates falling to the early 1980s, then levelling off for females, but probably increasing for men, particularly African men. There appears to have been a rapid increase in mortality rates of both African men and women since the mid-1990s. This rapid increase is more marked in younger adults where it appears to be more rapid in women than men.

The trend in childhood mortality is also very clear to the early 1990s. However, since we were unable to derive reliable estimates of childhood mortality from the 2001 census data we are unable to answer two very important questions about the trend in rates for the future. The first is whether the up-tick in rates is real, set to continue and is correctly located in time. The second is whether the higher estimate of child mortality for Whites and Indians are correct.

7 Conclusions and discussion

7.1 Editing

One edits census data for two major reasons. The first is to correct 'obvious' errors such as men reporting on the birth of their last child, parents too young to have a child or children too old to be likely to be living with a surviving parent, etc. The second is to 'tidy up' the tables and to produce 'one number' which is unambiguous. While both reasons are understandable and have clear advantages, it should be recognised that not all data should be treated in the same way, and that certainly the data needed to estimate demographic variables (mortality, fertility and migration) are a special case.

These data on vital events are special for a number of reasons. Apart from being needed in their own right (to plan for births, health services, mortuaries and cemeteries, etc), they are crucial for assessing the overall count (by checking consistency between current and past censuses, and births, deaths and migration) and through this potentially impact on all census estimates. In addition, these estimates are crucial to projecting the population forward beyond the date of the census. When it is not possible to derive these estimates in other ways (by, for example, using data on vital registration) and while the quality of data captured in censuses and surveys remains less than perfect, these questions need to analysed by specialists and should not be expected to provide estimates of fertility or mortality directly, without need for further adjustment.

Thus, we recommend that data on questions specifically designed for the purposes of estimating mortality and fertility be released with only minimal edits, in the case of the 2001 census to include age of respondent (or mother in the case of fertility rates), sex, population group of respondent (or the majority of the household, in the case of household death data) and province. Further, we recommend that the data on the number of deaths in households, being both of specialist interest and clearly biased, only be released as part of the 10 per cent sample.

7.2 Mortality rates

One of the most disappointing features of the 2001 census is that one is unable to use these data to produce reliable estimates of the mortality of children. Infant and child mortality are crucial indicators of the well-being of a population and essential health management tools. This need is magnified where a country is suffering an epidemic which has an immediate impact on these indicators. Furthermore, childhood mortality rates significantly determine the estimate of life expectancy at birth, a crucial component of the Human Development Index published by the UNDP. It is of concern that South Africa has now become one of the countries where the estimates upon which such indices are based are no longer up to date.

In the absence of reliable estimates, one might be tempted to resort to extrapolation of past trends to derive estimates of current levels. Unfortunately, in the case of the infant and child mortality this is difficult to do, since the most recent estimates of childhood mortality (Dorrington, Moultrie and Nannan 2002; Nannan, Timæus, Moultrie *et al.* 2001; Udjo 1999; 1999) seem to indicate an up-turn in rates in the late-1990s, possibility for each of the population groups separately, but with too few recent data points from which a reliable trend can be estimated. Besides, such projections could only be relied upon for the very short term, since it would be unwise to hypothesise an indefinitely increasing trend.

As far as the adult mortality rates are concerned, we recommend acceptance of the rates produced by this research largely on the basis of their consistency, both internally (the relative levels by population group, higher mortality for men than women, etc) and externally (similar levels to estimates produced using other data sources). Furthermore, there can be little doubt that estimates derived from the official abridged life tables for 1996 are significantly out of line with our and other estimates. We are most confident of the estimates for the national population and for the African population and least confident of those for the White population. Both the comparison with the orphanhood-based estimates and our evaluation of the registration data against the census suggest that, if our preferred estimate for the latter population group is in error, it is too high rather than too low and that we may be overestimating the death rates of Whites by up to 5 per cent.

The close consistency between our estimates and those projected for this period by the ASSA2000 model lends support to the hypothesis that HIV/AIDS is accounting for increasing numbers of deaths in South Africa (and that these deaths were reported by and large by households in the 2001 census, and are captured by vital registration).

The provincial estimates are more uncertain since in the case of men at least, there is less consistency between the rates, or ranking of the provinces, produced using the household data and those based on the data on survival of parents than was apparent for the estimates by population group. In addition, there is little consistency between rankings based on our estimates and those based on the official abridged life tables for 1996 (Statistics South Africa 2000) or those estimated by Dorrington, Moultrie and Nannan (2002) for men in the same year (although the correlation of our results with these estimates is higher for women). Moreover, the age patterns of mortality in some provinces have some curious features. Nevertheless, the rankings of the rates by province seem reasonable and this and the consistency of the patterns of deaths (overall and those due to external causes) suggest that the deaths in the last 12 months question is a useful one to keep in the census in future.

Having said this there are a few areas that need further investigation, namely:

- The possibility of devising a three parameter fitting procedure which allows for an 'AIDS hump' to be incorporated into the standard table to be used for graduating data, with the extent of the 'hump' being determined by the data.
- Possible explanations for the under- and over-reporting of deaths by households. In particular, the explanation of the rapidly increasing mortality with increasing older ages in the Western Province and Gauteng.
- The relatively low mortality in Gauteng given the relatively high level of HIV prevalence in that province (which could reflect selective return migration to other provinces of those sick with AIDS).
- The differences between the estimates based on the data on survival of parents and those based on vital registration, particularly for White men and women, nationally, and for provincial-level estimates of men's mortality in general.

- Inconsistencies in the estimates of completeness of reporting of deaths (or population group) for Indian men and women.

7.3 Cause of death data

The questionnaire asked two questions about the cause of death. The first asked whether the death was from an accident or through violence. The second asked, in the case of the deceased being a woman under 50, whether the person died while pregnant or within six weeks after delivery. These questions were new to this census and will be of interest to some users and although it is not the purpose of this monograph to analyse these data we offer some observations by way of warning to users of these data.

It will take further analysis to decide how well these questions worked, but at first glance it would appear as if they produced higher estimates of mortality than would be expected.

For example, the proportion of deaths due to external causes from the Statistics South Africa Cause of Death sample for the year 2001 suggests that some 15.7% of male death and 5.5% of female deaths were due to external causes, whereas the census data suggest that some 18.2% of male deaths and 9.0% of female deaths were due to accidents or violence- a difference of some 2.5% to 3.5%. On the other hand the MRC's Burden of Disease report suggests that, including suicides the percentages for 2000 were 18.8% for males and 6.1% for females, suggesting that the excess is only in the data on female deaths, and apparently mainly at the older ages. The distribution of these deaths by age appears to be reasonable; however, it is of some concern to note that well over a third of these deaths have imputed ages.

As far as deaths during pregnancy are concerned, the DHS estimated the maternal mortality ratio to be 150 per 100 000 live births for the period 1989-98. The census estimate, assuming around 1.1 million births is 575 per 100 000 live births for all women 15-49, ranging between 70 for White women to over 650 for African women. Again the age distribution looks reasonable, but again some 38% of these deaths have imputed ages, undoubtedly many on the grounds of the cause of death being limited to women aged 15-49. Although it is not unreasonable to assume cause of death would be more accurately recorded than age, if the cause were wrong in these cases then the maternal mortality rate would be a more reasonable 330 per 100 000 births.

Although an increase since 1989-98 due to HIV/AIDS is not unexpected, it is unlikely that it could be as big as implied by the census data. On the other hand research has found that deaths during or soon after pregnancy account for between 3% and 10% of the total deaths of women aged 15-49 (Boerma 1978). In the case of the census this proportion is 6.4% for all women, ranging from 3.3% in the case of White women to 6.5% in the case of African women. Although this is on the high side it is still within the realms of possibility. There is thus a need to interrogate these data further.

7.4 Life Tables

Without estimates of childhood mortality it is not possible to produce official life tables for the years around the 2001 census. However, should tables be required by Statistics SA for their own use, such as for projecting the population, a number of possible methods exist for estimating childhood mortality. While none of them is obviously satisfactory, if some of them produce apparently reasonable and consistent results, one might use these estimates.

The first approach would be simply to assume that the level of under-reporting in the vital registration of deaths of children is the same as that of adult deaths. This is apparently the assumption made in the production of the abridged life tables for 1996, but undoubtedly will lead to an underestimate of the child mortality relative to adult mortality. The comparison of the registered deaths with those reported by households described in chapter 2 confirms this conclusion.

A second approach would be to use the estimate implied by the graduation of the adult mortality rates corrected for undercount by including this age group in the graduation. The drawback of this method is that it is not only reliant on the appropriateness of the standard chosen for the graduation, which as we indicated earlier is open to question, but also the reasonableness of a fit dominated by adult mortality.

A third approach would be to make use of deaths in this age group reported by households in the census, either directly or adjusted in keeping with the estimated under-reporting in the following age groups. The problem with this is that there are no grounds for asserting that either of these assumptions is correct.

Finally one could simply use the estimates produced by a projection model, such as ASSA2002, which projects mortality rates allowing for the past trend in non-AIDS mortality together with an assumed level of HIV/AIDS mortality.

It might be tempting to some demographers to use the estimates of adult mortality to decide on what level of model life table matches and then to infer the childhood mortality from this table. However, this approach is unlikely to produce anything sensible since most life tables do not reflect the impact of HIV/AIDS and even if they did you would have to be sure that this was to the same extent and that the child-adult relationship implied by the model matched that of South Africa and its sub-populations over this period.

By way of example, Table 7.1 presents various estimates of ${}_5q_0$. Clearly, simply adjusting the recorded deaths for the estimated adult undercount results in estimates that are too low, particularly for the African population but also, to a lesser extent, the Coloured population. Interestingly, the direct estimates derived from the number of deaths reported by households divided by the number of children estimated by the census in this age group (which is considered to be significantly undercounted (Dorrington, Budlender and Moultrie 2003)) seems to produce as good, if not better, estimates for the White and Indian populations as can be derived from either the vital registration data or the ASSA model. Further, if one is to consider using the adjusted household deaths, then it is advisable to use the same adjustment (possibly an average of the male and female adjustment) for both males and females to ensure that the estimate of male mortality remains higher than female mortality. Doing this would produce apparently reasonable estimates for the Coloured, Indian and White populations, but produces estimates for the African (and hence national) population that are clearly too high.

Finally, it is also apparent from these results that it is unlikely that the up-turn in mortality has continued for the Coloured, Indian or White populations.

Table 7.1 Various estimates of childhood mortality ($_5q_0$)

	Applying adult completeness	*Household deaths*	*Adjusted household deaths*	*ASSA2000*
Males				
African	0.059	0.085	0.128	0.099
Coloured	0.031	0.030	0.035	0.037
Indian	0.017	0.021	0.022	0.014
White	0.015	0.014	0.031	0.011
National	0.054	0.075		0.087
Females				
African	0.052	0.076	0.138	0.091
Coloured	0.026	0.025	0.038	0.032
Indian	0.012	0.015	0.020	0.012
White	0.011	0.013	0.016	0.009
National	0.046	0.067		0.079

For illustrative purposes we have produced life tables in Appendix 2 by linking estimates of q_0 and $_5q_0$ derived from the unadjusted deaths in the previous 12 months reported by households in the 2001 census and census estimates of the population aged under five, directly, with the graduated rates produced earlier. Inspection of the rates, however, suggested that q_0 was probably underestimated relative to $_5q_0$ and it was decided to set it to 70% of $_5q_0$.

In addition, to give a sense of the uncertainty in the estimates, we have derived estimates of key indicators based on extreme assumptions of low and high mortality. For adult mortality we assumed that the rates could be 10% lower or higher (i.e. in effect that the estimate of completeness could be wrong by a maximum of 10% either way). As for the child mortality the low estimate was derived assuming the downward trend in rates derived by Dorrington *et al* (In press) from 1986 to 1991 continued through to 2004 (i.e. ignoring the impact of HIV/AIDS). The high estimate was derived by assuming the non-AIDS mortality rate remained constant at the 1993 rates derived by Dorrington *et al* (In press) and adding to this AIDS mortality rates derived using the survival curve suggested by Zaba *et al* (2003) with parameters updated by Zaba (2003) (i.e. ignoring the impact of prevention of mother to child transmission and antiretroviral treatment programmes).

Since in the case of the White population the low estimates of child mortality was significantly higher than that derived from the census data it was decided to set the lower bound on rates equal to the rates derived from the census data.

These indicators are compared to those derived from the illustrative life tables in Table 7.2 to Table 7.4.

There are four useful conclusions that can be drawn from this comparison. The first is that the estimates of child mortality derived from the census data are on the light side (at a rough guess we would say about 10% too light). The second is that the upper and lower estimates provide indicators against which other estimates can be judged. It is unlikely that estimates outside these bands can be plausible. The third is that since the ranking of the life expectancy at birth of the various provinces is very similar to that of the life expectancy at age 15, one can have some degree of confidence in the ranking of provinces by the various indicators. Forth, nationally the difference between female and male life expectancies is very large, nearly 8 years, and this holds for each of the population groups separately and most of the provinces with the (unique) exception of the North West. At the other extreme the difference in life expectancy in Limpopo is 11 years! None of these differences are attributable to differences in child mortality of girls and boys assumed in the illustrative life tables.

7.5 Conclusions

The overall conclusion must be that the data on mortality collected by the 2001 census are disappointing and that the editing of those data does little to improve them. In the case of children the data are so poor as to be useless for deriving reliable estimates of childhood mortality. As far as adults are concerned, the data are more useful. A fairly high degree of consistency exists between the estimates based on data on survival of parents and those produced using the vital registration data at the national level and by population group. However, use of the recent deaths data in the census and the extra assumptions involved in producing provincial level mortality estimates yield results that are only moderately consistent with those from orphanhood data. It remains unclear how much reliance should be placed on these provincial estimates.

Recommendations:

- Statistics South Africa prioritises the capturing and publication of reports on registered deaths.
- ***Statistics South Africa includes in the 2004 general household survey questions on children ever borne and children surviving*** in order to be in a position, as soon as possible, to be able to provide the country with estimates of the level of childhood mortality. In addition everything possible, including enhanced training of fieldworkers, should be done to ensure the success of these questions.

Table 7.2 High, low and indicative mortality indicators: national and population group

National	Male		Low mortality	High mortality	Indian	Male		Low mortality	High mortality
	e0	52.5	54.7	45.4		e0	62.3	64.1	58.6
	e15	42.8	44.5	41.2		e15	49.6	51.0	48.3
	e65	11.6	12.3	10.9		e65	11.1	11.8	10.5
	1000*q0	52.5	43.4	118.9		1000*q0	14.1	13.8	41.8
	1000*5q0	75.1	75.6	189.4		1000*5q0	20.1	24.7	70.8
	45q15	51.1%	47.5%	54.5%		45q15	32.9%	30.1%	35.5%
	Female		Low mortality	High mortality		Female			High mortality
	e0	60.3	63.7	53.6		e0	71.1	73.3	68.5
	e15	50.6	52.2	49.1		e15	58.5	59.6	57.5
	e65	14.7	15.5	14.1		e65	14.5	15.2	13.8
	1000*q0	47.2	27.6	102.2		1000*q0	10.3	8.0	30.6
	1000*5q0	67.4	48.2	161.6		1000*5q0	14.7	14.2	51.2
	45q15	33.6%	30.8%	36.3%		45q15	14.9%	13.5%	16.3%
African	Male		Low mortality	High mortality	White	Male		Low mortality	High mortality
	e0	50.1	52.5	43.2		e0	65.3	67.9	61.6
	e15	40.8	42.6	39.3		e15	52.7	54.1	51.5
	e65	11.2	12.0	10.6		e65	12.8	13.5	12.2
	1000*q0	59.3	47.2	127.0		1000*q0	10.0	10.0	42.0
	1000*5q0	84.7	82.0	201.2		1000*5q0	14.3	14.3	71.2
	45q15	56.3%	52.5%	59.7%		45q15	26.6%	24.3%	28.9%
	Female		Low mortality	High mortality		Female			Low mortality
	e0	58.1	61.8	51.3		e0	72.7	75.1	70.1
	e15	48.8	50.5	47.2		e15	60.1	61.2	59.1
	e65	14.6	15.3	13.9		e65	15.9	16.6	15.3
	1000*q0	53.3	30.2	110.1		1000*q0	8.8	8.8	30.8
	1000*5q0	76.1	52.6	173.3		1000*5q0	12.5	12.5	51.5
	45q15	37.7%	34.6%	40.5%		45q15	13.7%	12.4%	15.0%
Coloured	Male		Low mortality	High mortality					
	e0	57.7	59.2	52.6					
	e15	44.8	46.4	43.4					
	e65	11.0	11.7	10.3					
	1000*q0	21.5	17.7	56.7					
	1000*5q0	30.7	31.7	95.6					
	45q15	45.6%	42.2%	48.8%					
	Female		Low mortality	High mortality					
	e0	65.4	67.4	61.4					
	e15	52.5	53.9	51.3					
	e65	13.3	14.0	12.7					
	1000*q0	17.4	10.2	41.8					
	1000*5q0	24.9	18.2	70.0					
	45q15	28.2%	25.7%	30.5%					

Table 7.3 High, low and indicative mortality indicators: KZ, NC, WC, FS and EC provinces

KZ	Male		Low mortality	High mortality	NC	Male		Low mortality	High mortality
	e0	44.4	47.2	37.3		e0	53.5	55.7	46.7
	e15	35.7	37.5	34.1		e15	42.3	44.0	40.7
	e65	10.7	11.5	10.1		e65	11.5	12.3	10.9
	1000*q0	72.9	54.6	151.8		1000*q0	41.5	29.3	98.3
	1000*5q0	104.2	94.1	237.7		1000*5q0	59.3	52.0	159.6
	45q15	67.3%	63.5%	70.8%		45q15	51.6%	47.9%	55.0%
	Female		Low mortality	High mortality		Female			High mortality
	e0	52.7	57.1	45.3		e0	59.1	62.3	53.1
	e15	44.4	46.3	42.6		e15	47.7	49.4	46.1
	e65	14.1	14.9	13.5		e65	13.8	14.6	13.2
	1000*q0	67.6	36.1	135.9		1000*q0	37.5	16.7	80.7
	1000*5q0	96.6	62.6	211.4		1000*5q0	53.6	29.7	130.3
	45q15	46.0%	42.5%	49.2%		45q15	39.9%	36.7%	42.8%
WC	Male		Low mortality	High mortality	FS	Male		Low mortality	High mortality
	e0	63.5	62.4	55.3		e0	49.2	51.5	41.6
	e15	48.0	49.5	46.6		e15	39.6	41.4	38.0
	e65	12.1	12.8	11.4		e65	11.5	12.3	10.9
	1000*q0	19.9	15.9	59.7		1000*q0	64.7	46.8	135.2
	1000*5q0	39.3	28.4	98.8		1000*5q0	92.5	81.5	213.2
	45q15	37.9%	34.8%	40.8%		45q15	59.1%	55.3%	62.6%
	Female		Low mortality	High mortality		Female			High mortality
	e0	68.2	70.9	64.2		e0	54.7	58.5	47.3
	e15	55.9	57.2	54.8		e15	44.9	46.8	43.1
	e65	14.9	15.6	14.2		e65	15.1	15.8	14.4
	1000*q0	17.6	9.2	47.2		1000*q0	59.6	29.1	118.3
	1000*5q0	25.1	16.3	77.1		1000*5q0	85.1	51.1	185.5
	45q15	21.6%	19.6%	23.5%		45q15	46.8%	43.4%	50.1%
EC	Male		Low mortality	High mortality					
	e0	46.3	47.5	39.1					
	e15	36.5	38.3	34.9					
	e65	11.2	11.9	10.5					
	1000*q0	55.7	62.1	135.7					
	1000*5q0	79.5	106.4	215.2					
	45q15	66.3%	62.4%	69.8%					
	Female		Low mortality	High mortality					
	e0	55.6	58.2	48.1					
	e15	46.1	47.9	44.4					
	e65	14.6	15.3	13.9					
	1000*q0	49.6	41.4	119.3					
	1000*5q0	70.8	71.7	188.2					
	45q15	43.4%	40.1%	46.6%					

Table 7.4 High, low and illustrative mortality indicators: NW, MP, GT and LP provinces

NW	Male		Low mortality	High mortality	MP	Male		Low mortality	High mortality
	e0	56.8	57.0	46.9		e0	51.9	54.4	44.2
	e15	43.9	45.6	42.4		e15	42.6	44.4	41.0
	e65	12.1	12.8	11.5		e65	12.4	13.1	11.7
	1000*q0	53.0	31.5	112.9		1000*q0	58.4	45.5	131.8
	1000*5q0	103.2	55.7	180.1		1000*5q0	83.5	79.3	208.6
	45q15	49.1%	45.5%	52.4%		45q15	52.6%	48.9%	56.0%
	Female		Low mortality	High mortality		Female			High mortality
	e0	57.8	62.0	51.5		e0	58.3	61.8	50.6
	e15	47.5	49.4	45.9		e15	48.6	50.4	46.9
	e65	15.2	15.9	14.5		e65	15.9	16.6	15.2
	1000*q0	46.8	18.8	96.9		1000*q0	51.0	28.5	115.3
	1000*5q0	66.8	33.3	153.5		1000*5q0	72.8	49.9	181.4
	45q15	41.5%	38.3%	44.6%		45q15	39.1%	36.0%	42.0%
GT	Male		Low mortality	High mortality	LP	Male		Low mortality	High mortality
	e0	58.1	60.7	51.3		e0	56.3	57.6	50.4
	e15	47.9	49.5	46.5		e15	45.4	47.1	43.8
	e65	11.8	12.6	11.2		e65	13.7	14.5	13.1
	1000*q0	46.1	30.2	102.3		1000*q0	36.8	40.2	87.3
	1000*5q0	65.9	53.1	161.8		1000*5q0	52.6	69.9	142.4
	45q15	38.1%	35.1%	41.0%		45q15	48.4%	44.9%	51.7%
	Female		Low mortality	High mortality		Female			High mortality
	e0	63.0	66.6	57.0		e0	67.4	69.6	62.0
	e15	52.7	54.2	51.4		e15	56.6	58.0	55.2
	e65	14.0	14.8	13.4		e65	18.0	18.7	17.4
	1000*q0	40.2	19.0	88.5		1000*q0	31.9	26.0	71.1
	1000*5q0	57.4	33.2	138.5		1000*5q0	45.5	45.2	115.0
	45q15	27.9%	25.5%	30.2%		45q15	25.5%	23.3%	27.7%

References

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

Bennett, N G and Shiro Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**(2):217-233.

Boerma, Ties. 1978. "The magnitude of the maternal mortality problem in sub-Saharan Africa", *Soc Sci and Med* **24**:551-558.

Brass, William. 1968. *The Demography of Tropical Africa.* Princeton, New Jersey: Princeton University Press.

Brass, William. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defected Data.* North Carolina: Carolina Population Center.

Brass, William and Kenneth H. Hill. 1973. "Estimating adult mortality from orphanhood," Paper presented at Proceedings of the International Population Conference. Liege, 1973. International Union for the Scientific Study of Population. Vol. 3:111-123.

Buro (sic) of Statistics. 1968. *South African Statistics 1968*. Pretoria: Buro of Statistics, Republic of South Africa.

Buro (sic) of Statistics. 1972. *South African Statistics 1972*. Pretoria: Buro of Statistics, Republic of South Africa.

Carrier, N.H. 1958. "A note on the estimation of mortality and other population characteristics, given death by age", *Population Studies* **12**:149-163.

Central Statistical Service. 1985. *South African Life Tables 1979-81*. Report No. 02-06-03. Pretoria: Government Printer.

Central Statistical Service. 1987. *South African Life Tables 1984-86*. Report No. 02-06-04. Pretoria: Government Printer.

Department of Health. 2002. *South Africa Demographic and Health Survey 1998: Full Report.* Pretoria: Department of Health.

Dorrington, Rob, Debbie Budlender and Tom A. Moultrie. 2003. *Census 2001 Group 3 Report: Are the numbers what we expect?* Pretoria: Statistics Council.

Dorrington, Rob E, Tom A Moultrie, Nadine Nannan and Ian M Timaeus. In press. "Estimates of provincial fertility and mortality in South Africa, 1985-1996", *Southern African Journal of Demography* **9**(2)

Dorrington, Rob E. 1989. "African mortality rates - an initial estimate", *Transactions of the Actuarial Society of South Africa* **7**:169-180.

Dorrington, Rob E. 1998. "Estimates of the Level and Shape of Mortality Rates in South Africa Around 1985 and 1990 Derived by Applying Indirect Demographic Techniques to Reported Deaths." Unpublished MPhil thesis, Cape Town: University of Cape Town.

Dorrington, Rob E., David E Bourne, Debbie Bradshaw, Ria Laubscher *et al.* 2001. *The impact of HIV/AIDS on adult mortality in South Africa.* Tygerberg: South African Medical Research Council.

Dorrington, Rob E., Debbie Bradshaw and Trevor Wegner. 1999. *Estimates of the Level and Shape of Mortaltiy Rates in South Africa Around 1985 and 1990 Derived by Applying Indirect Demographic Techniques to Reported Deaths.* Cape Town: Medical Research Counicl.

Dorrington, Rob E., Tom A. Moultrie and Nadine Nannan. 2002. "Estimates of provincial fertility and mortality 1985-1996," Paper presented at 2002 Demsa Annual Conference. University of the Western Cape, South Africa, 26-27 September 2002.

El-Badry, M.A. 1961. "Failure of enumerators to make entries of zero: errors in recoding childless cases in population censuses", *Journal of the American Statistical Association* **56**(296):909-924.

Hill, Kenneth H. 1987. "Estimating census and death registration completeness", *Asian and Pacific Population Forum* **1**(3):8-13.

Hill, Kenneth H. and James Trussell. 1977. "Further developments in indirect mortality estimation", *Population Studies* **31**(2):313-333.

INDEPTH (International Network for the continuous Demographic Evaluation of Populations and their Health). 2002. *Population and Health in Developing Countries*. Ottawa: International Development Research Centre.
Mostert, W.P., J.L. van Tonder and B.E. Hofmeyr. 1987. *Rekonstruksie van die Sensusouderdomstrukture van die Suid-Afrikaanse Swart Bevolking: 1936 tot 1985 [Reconstruction of the Census Age Structure of the South African Black Population: 1936 to 1985]*. Report S-153 (1987). Pretoria: Human Sciences Research Council.
Moultrie, Tom A. and Rob Dorrington. 2004. *Estimation of fertility from the 2001 South Africa Census data*. Cape Town: Centre for Actuarial Research for Statistics South Africa.
Moultrie, Tom A. and Ian M. Timæus. 2002. *Trends in South African fertility between 1970 and 1998: An analysis of the 1996 Census and the 1998 Demographic and Health Survey*. Cape Town: Medical Research Council. http://www.mrc.ac.za/bod/trends.pdf.
Moultrie, Tom A. and Ian M. Timæus. 2003. "The South African fertility decline: Evidence from two censuses and a Demographic and Health Survey", *Population Studies* **57**(3):265-283.
Murray, Chris J. L., O. B. Ahmad, Alan D. Lopez and Josh A. Salomon. 2000. *WHO System of Model Life Tables*. Global Program on Evidence for Health Policy Discussion Paper Series: No. 8. Geneva: World Health Organization.
Nannan, Nadine, Ian M. Timæus, Tom A. Moultrie and Rob E. Dorrington. 2001. "Census and DHS levels of child mortality in South Africa: Where do they meet?" Paper presented at Annual Conference of DEMSA. University of Pretoria, October 2001.
Office of Population Census and Surveys. 1987. *English Life Tables No. 14*. Series DS no. 7. London: Her Majesty's Stationery Office.
Sadie, Jan L. 1970. "An evaluation of demographic data pertaining to the non-White population of South Africa (Part III)", *South African Journal of Economics* **38**(2):171-191.
Sadie, Jan L. 1970. "An evaluation of demographic data pertaining to the non-White population of South Africa (Parts I and II)", *South African Journal of Economics* **38**(1):1-34.
Sadie, Jan L. 1973. *Projections of the South African Population 1970 - 2020*. Johannesburg: Industrial Development Corporation.
Sadie, Jan L. 1988. *A reconstruction and Projection of Demographic Movements in the RSA and TBVC Countries*. Pretoria: Bureau of Market Research.
Sadie, Jan L. 1993. *A Projection of the South African Population 1991-2011*. Research Report No. 196. Pretoria: Bureau of Market Research.
Statistics South Africa. 2000. *South African life tables, 1985-1994 and 1996*. 02-06-04 (1985-1994 & 1996). Pretoria: Statistics South Africa.
Statistics South Africa. 2002. *Causes of Death in South Africa 1997-2001 Advanced release of Recorded Causes of Death*. P0309.2. Pretoria: Statistics South Africa.
Statistics South Africa. 2003. *Census 2001: Computer Editing Specifications (30 September 2003)*. Pretoria: Statistics South Africa.
Timæus, Ian M. 1992. "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of the United Nations* **33**:47-63.
Timæus, Ian M. 1993. "Adult mortality," in Foote, Karen A., Kenneth H. Hill and Linda G. Martin (eds). *Demographic Change in Sub-Saharan Africa*. Washington, D.C.: National Academy Press, pp. 218-255.
Timæus, Ian M. 2004. "Impact of HIV on mortality in Southern Africa: Evidence from demographic surveillance," Paper presented at Seminar of the IUSSP Committee "Emerging Health Threats" HIV, Resurgent Infections and Population Change in Africa. Ougadougou, 12-14 February.
Timæus, Ian M., Rob Dorrington, Debbie Bradshaw, Nadine Nannan *et al.* 2001. "Adult mortality in South Africa, 1980-2000: From apartheid to AIDS," Paper presented at Population Association of America 2001 Annual Meeting. Washington DC, March 29-31 2001.
Timæus, Ian. M. 1991. "Estimation of mortality from orphanhood in adulthood", *Demography* **28**(2):213-227.
Udjo, Eric O. 1999. "Additional Evidence Regarding Fertility and Mortality Trends in South Africa and Implications for Population Projections," Paper presented at Workshop on Phase 2 of Census 1996 Review. Wanderers Club, Johannesburg, 3-4 December.
Udjo, Eric O. 1999. "A Four-"Race" Model of Mortality in South Africa," Paper presented at Workshop on Phase 2 of Census 1996 Review. Wanderers Club, Johannesburg, 3-4 December.

van Eeden, I.J. and J.L van Tonder. 1975. *Abridged Life Tables for All the Population Groups in South Africa*. Report No. S-34. Pretoria: Human Sciences Research Council.

van Tonder, J.L, W.P Mostert and B.E. Hofmeyer. 1987. *Rekonstruksie van die Sensusouderdomstrukture van die Siud-Afrikaanse Asier- Blanke en Kleuring Bevolkings: 1946 tot 1985[Reconstruction of the Census Age Structure of the South African Asian White and Coloured Populations: 1946 to 1985]*. Report No. S-154. Pretoria: Human Science Research Council.

Zaba, Basia. 2003. "Child mortality: development of a functional representation for net mortality due to HIV," Paper presented at UNAIDS Reference Group on Estimates, Modelling and Projections. Sintra, Portugal, 2-4 December.

Zaba, Basia, Milly Marston and Sian Floyd. 2003. "The Effect of HIV on Child Mortality Trends in sub-Saharan Africa," Paper presented at Training Workshop on HIV/AIDS and Adult Mortality in Developing Countries. Population Division, Department of Economic and Social Affairs, New York, 8-13 September 2003.

Appendix 1: Estimated numbers of reported deaths in the five year intercensal period

National Males	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	14,826	17,129	14,575	15,673	12,744	74,947
1-4	6,177	6,686	5,682	6,716	5,832	31,093
5-9	2,090	2,100	2,039	2,383	2,043	10,654
10-14	1,842	1,974	1,933	2,150	1,727	9,626
15-19	4,037	4,347	4,489	4,776	4,215	21,865
20-24	8,528	8,898	9,103	9,298	8,606	44,432
25-29	12,201	12,936	14,449	15,544	16,352	71,481
30-34	12,832	14,247	16,579	18,710	20,573	82,942
35-39	12,816	14,170	16,710	18,948	20,627	83,270
40-44	12,220	13,476	15,366	17,207	18,793	77,063
45-49	12,592	13,587	15,205	16,714	17,823	75,921
50-54	11,660	12,726	14,091	15,683	16,876	71,036
55-59	13,425	13,601	14,499	14,624	14,792	70,941
60-64	11,044	11,942	13,038	14,412	15,332	65,769
65-69	13,220	13,044	13,330	13,019	13,245	65,859
70-74	12,341	12,535	13,327	13,659	14,301	66,163
75-79	11,184	10,913	11,047	10,813	11,121	55,078
80-84	6,878	7,487	7,924	8,615	9,307	40,211
85+	6,864	6,557	6,852	7,346	7,526	35,145
Total	186,778	198,354	210,238	226,290	231,835	1,053,495

African Males	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	10,256	10,316	9,957	12,034	9,697	52,260
1-4	4,449	4,506	4,241	5,235	4,507	22,938
5-9	1,500	1,536	1,455	1,815	1,601	7,908
10-14	1,247	1,300	1,349	1,638	1,312	6,848
15-19	2,739	2,815	2,983	3,174	2,844	14,555
20-24	5,623	5,838	6,353	6,684	6,017	30,515
25-29	9,233	9,934	10,624	11,622	12,341	53,754
30-34	10,230	11,396	12,691	14,150	15,751	64,217
35-39	10,107	11,155	12,208	13,790	14,939	62,199
40-44	8,966	9,881	10,734	12,306	13,141	55,027
45-49	8,871	9,545	10,181	11,222	11,846	51,664
50-54	6,798	7,650	8,481	9,947	10,828	43,705
55-59	7,846	8,068	8,504	8,657	8,514	41,590
60-64	6,044	6,629	7,209	8,089	8,710	36,681
65-69	7,043	6,980	6,941	6,699	6,888	34,551
70-74	6,388	6,662	6,983	7,299	7,478	34,810
75-79	5,726	5,618	5,583	5,422	5,233	27,582
80-84	3,214	3,554	3,970	4,397	4,706	19,840
85+	3,266	3,329	3,316	3,569	3,502	16,982
Total	119,546	126,710	133,764	147,750	149,855	677,625

White Males	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	195	189	191	164	175	913
1-4	112	109	85	127	100	532
5-9	41	40	49	33	34	197
10-14	54	54	40	74	54	275
15-19	247	239	237	272	181	1,176
20-24	406	395	407	379	345	1,932
25-29	521	507	541	466	443	2,478
30-34	463	465	475	462	465	2,329
35-39	623	617	607	561	633	3,042
40-44	767	776	747	806	824	3,919
45-49	963	962	976	1,041	891	4,834
50-54	1,196	1,214	1,397	1,180	1,189	6,176
55-59	1,360	1,407	1,452	1,512	1,557	7,288
60-64	1,336	1,449	1,575	1,699	1,850	7,909
65-69	2,054	2,054	2,164	2,156	2,020	10,448
70-74	2,239	2,279	2,415	2,328	2,346	11,607
75-79	2,372	2,411	2,528	2,309	2,598	12,218
80-84	1,728	1,781	1,750	1,871	2,033	9,163
85+	1,828	1,844	1,838	1,850	1,929	9,289
Total	18,503	18,792	19,474	19,289	19,668	95,726

Indian Males	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	72	70	49	70	77	337
1-4	22	22	16	30	19	110
5-9	17	17	24	13	13	83
10-14	22	22	32	30	11	117
15-19	70	70	79	52	76	347
20-24	120	123	94	117	167	621
25-29	143	146	134	187	147	757
30-34	124	126	118	136	137	640
35-39	149	153	162	156	167	787
40-44	167	171	170	159	200	867
45-49	303	309	293	328	339	1,571
50-54	161	204	267	307	420	1,359
55-59	290	307	282	357	384	1,621
60-64	253	270	310	292	320	1,445
65-69	293	304	276	346	359	1,579
70-74	261	272	240	234	392	1,398
75-79	196	203	185	209	251	1,043
80-84	101	109	95	118	160	583
85+	83	84	96	84	76	423
Total	2,847	2,981	2,921	3,224	3,714	15,688

Coloured Males	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	655	658	731	772	541	3,357
1-4	200	200	227	210	176	1,013
5-9	110	112	155	109	87	573
10-14	105	109	150	110	97	571
15-19	446	448	420	491	450	2,255
20-24	636	640	630	679	639	3,224
25-29	862	867	904	908	834	4,376
30-34	885	898	921	930	924	4,559
35-39	979	1,013	1,091	994	1,155	5,232
40-44	882	927	919	1,073	1,082	4,883
45-49	1,018	1,069	1,200	1,111	1,226	5,624
50-54	923	982	1,073	1,180	1,110	5,268
55-59	1,068	1,101	1,226	1,073	1,205	5,673
60-64	863	978	1,086	1,302	1,410	5,639
65-69	1,068	1,106	1,260	1,198	1,122	5,753
70-74	788	828	830	841	1,040	4,327
75-79	580	599	656	606	652	3,092
80-84	362	376	386	368	454	1,946
85+	386	382	362	419	346	1,896
Total	12,816	13,293	14,229	14,374	14,548	69,260

National Females	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	12,873	14,020	12,446	13,403	11,977	64,719
1-4	5,871	6,740	5,603	6,042	5,306	29,562
5-9	1,658	1,689	1,745	1,827	1,559	8,479
10-14	1,526	1,726	1,424	1,764	1,483	7,923
15-19	3,139	3,111	3,403	3,631	3,543	16,826
20-24	6,504	7,028	8,193	9,561	9,788	41,073
25-29	7,858	9,537	12,142	15,402	18,103	63,042
30-34	7,617	9,235	11,922	15,414	17,827	62,015
35-39	7,055	8,439	10,532	13,367	15,299	54,693
40-44	6,194	7,343	8,767	10,870	12,446	45,620
45-49	6,708	7,410	8,603	9,668	10,574	42,964
50-54	6,018	6,748	7,822	9,041	9,922	39,551
55-59	8,317	8,490	8,837	8,979	9,210	43,832
60-64	10,245	9,814	10,365	11,499	12,198	54,120
65-69	12,256	12,094	12,736	12,550	13,035	62,671
70-74	11,428	11,695	12,615	14,228	15,312	65,277
75-79	12,278	12,010	12,028	11,854	12,320	60,489
80-84	9,840	10,552	11,611	12,779	14,015	58,797
85+	13,361	13,139	13,787	15,083	15,975	71,345
Total	150,745	160,821	174,579	196,960	209,892	892,998

African Females	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	9,278	9,323	8,943	10,208	9,240	46,993
1-4	4,111	4,167	4,102	4,697	4,146	21,223
5-9	1,145	1,170	1,230	1,349	1,132	6,025
10-14	986	1,024	976	1,301	1,078	5,365
15-19	2,555	2,607	2,640	2,888	2,662	13,352
20-24	6,987	7,129	6,789	7,673	7,774	36,352
25-29	6,677	8,105	9,646	12,373	14,357	51,157
30-34	6,519	7,899	9,357	12,074	13,896	49,745
35-39	5,542	6,658	7,786	10,082	11,395	41,463
40-44	4,238	5,119	6,113	7,622	8,971	32,063
45-49	5,080	5,517	5,904	6,679	7,021	30,202
50-54	3,697	4,225	4,741	5,702	6,255	24,620
55-59	4,944	5,068	5,189	5,336	5,456	25,994
60-64	6,206	6,393	6,196	6,830	7,266	32,891
65-69	7,206	7,284	7,447	7,074	7,716	36,727
70-74	6,180	6,737	6,926	8,131	8,921	36,895
75-79	6,414	6,318	6,121	6,114	6,133	31,101
80-84	4,552	5,160	5,995	6,613	7,329	29,649
85+	5,737	6,017	6,075	6,763	7,019	31,611
Total	98,052	105,922	112,175	129,511	137,767	583,427

White Females	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	152	147	168	138	113	718
1-4	75	73	62	86	63	360
5-9	41	40	47	30	37	195
10-14	38	38	48	28	39	191
15-19	115	110	117	85	101	528
20-24	157	151	152	156	117	734
25-29	160	156	164	134	143	756
30-34	213	214	217	223	214	1,081
35-39	299	298	317	283	293	1,490
40-44	343	349	353	348	382	1,775
45-49	499	503	511	539	491	2,543
50-54	641	655	692	626	715	3,329
55-59	817	839	965	834	862	4,317
60-64	1,080	1,111	1,158	1,127	1,231	5,707
65-69	1,445	1,447	1,557	1,503	1,327	7,278
70-74	1,801	1,823	1,901	1,902	1,818	9,246
75-79	2,372	2,389	2,537	2,331	2,401	12,028
80-84	2,622	2,650	2,713	2,611	2,780	13,376
85+	4,464	4,498	4,438	4,371	4,817	22,588
Total	17,335	17,492	18,116	17,354	17,943	88,240

Indian Females	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	47	45	35	36	54	217
1-4	11	11	8	8	14	52
5-9	20	20	21	17	19	97
10-14	22	23	30	32	13	120
15-19	41	40	33	52	36	202
20-24	56	56	62	46	55	276
25-29	52	53	58	52	53	269
30-34	53	53	58	59	51	274
35-39	73	75	63	83	89	382
40-44	83	85	85	111	79	444
45-49	114	118	101	160	127	620
50-54	155	162	191	194	159	861
55-59	197	207	201	267	226	1,098
60-64	218	232	192	243	334	1,219
65-69	263	281	285	327	353	1,509
70-74	271	290	229	356	399	1,545
75-79	207	219	205	229	293	1,154
80-84	195	207	197	214	276	1,089
85+	110	115	97	105	165	592
Total	2,187	2,293	2,152	2,593	2,795	12,020

Coloured Females	1996/97	1997/98	1998/99	1999/00	2000/01	1996-2001
0	517	520	569	615	430	2,651
1-4	195	196	181	199	211	983
5-9	71	71	87	83	55	366
10-14	69	71	75	92	68	374
15-19	159	159	123	160	191	793
20-24	291	293	288	346	267	1,484
25-29	415	418	507	410	372	2,122
30-34	553	562	520	591	622	2,848
35-39	600	624	616	711	704	3,256
40-44	566	598	648	712	656	3,181
45-49	714	723	733	740	751	3,661
50-54	599	648	708	747	825	3,527
55-59	664	691	722	778	755	3,610
60-64	913	941	980	1,073	964	4,871
65-69	792	834	910	973	912	4,420
70-74	857	905	1,045	1,007	999	4,813
75-79	687	717	709	753	869	3,736
80-84	559	579	574	635	648	2,996
85+	679	694	725	686	755	3,539
Total	9,900	10,245	10,721	11,312	11,053	53,232

Appendix 2: Illustrative life tables

	National			
	Male		*Female*	
Age	l_x	$_nL_x$	l_x	$_nL_x$
0	100000	104068	100000	104079
1	94745	307209	95282	310729
5	92493	461187	93260	465353
10	91982	458793	92882	463580
15	91535	454908	92551	460622
20	90430	446075	91700	452446
25	88011	428765	89290	437246
30	83534	403204	85634	418377
35	77815	374102	81747	399627
40	71904	344461	78131	382221
45	65966	312702	74782	363880
50	59235	278518	70806	343451
55	52315	242087	66617	319942
60	44718	203531	61430	291200
65	36942	163486	55162	256861
70	28791	121406	47760	215100
75	20269	79195	38605	164347
80	12107	42194	27732	106922
85	5599	16850	16088	53391
90	1830	4574	6643	16607

* This life table is illustrative and should not be taken as implying accuracy of childhood mortality rates

	African				Coloured			
	Male		*Female*		*Male*		*Female*	
Age	l_x	$_nL_x$	l_x	$_nL_x$	l_x	$_nL_x$	l_x	$_nL_x$
0	100000	103829	100000	103701	100000	96516	100000	99364
1	94072	304292	94670	309021	97853	396388	98256	368117
5	91531	456280	92385	460911	96933	483719	97509	486928
10	90981	453721	91980	459014	96555	481850	97263	485706
15	90507	449817	91626	455811	96185	476998	97020	483723
20	89421	440799	90700	446690	94619	466955	96470	479567
25	86910	422200	87990	429359	92174	452236	95359	472706
30	82017	393398	83788	407445	88742	434500	93728	463131
35	75433	360230	79232	385765	85083	414171	91533	450882
40	68761	326908	75110	366046	80625	390973	88834	436398
45	62114	291932	71340	345716	75814	362067	85744	417849
50	54810	255200	66992	323719	69116	326504	81433	393923
55	47447	217003	62547	298946	61628	284337	76194	364498
60	39591	178209	57114	269961	52358	236224	69701	325436
65	31964	140360	50986	236885	42476	185162	60681	278300
70	24510	102255	43942	197195	32075	133531	50924	224263
75	16867	65221	35254	149318	21973	84098	39284	160281
80	9848	34193	25053	96619	12543	42696	25774	93320
85	4516	13626	14543	48782	5494	16147	13000	40094
90	1485	3713	6157	15392	1693	4234	4509	11272

* These life tables are illustrative and should not be taken as implying accuracy of childhood mortality rates

	Indian				*White*			
	Male		*Female*		*Male*		*Female*	
Age	l_x	$_nL_x$	l_x	$_nL_x$	l_x	$_nL_x$	l_x	$_nL_x$
0	100000	101284	100000	117759	100000	132040	100000	128616
1	98593	328221	98968	269789	98999	235168	99123	242571
5	97991	489281	98525	492001	98570	492426	98747	493301
10	97722	487764	98276	490694	98400	491486	98574	492507
15	97384	484481	98002	488892	98194	488844	98429	491173
20	96410	477718	97555	486234	97344	482957	98041	488776
25	94682	467806	96939	483136	95842	474546	97470	485966
30	92449	457256	96316	479912	93982	465656	96917	482685
35	90461	445728	95650	475805	92285	455970	96158	478185
40	87844	431521	94674	470277	90111	443456	95118	472413
45	84783	411447	93440	462337	87286	426987	93850	464257
50	79845	384377	91502	449613	83536	405053	91860	452215
55	73980	347981	88362	429198	78536	376194	89041	434899
60	65389	301907	83366	399828	72035	339403	84951	409680
65	55636	244439	76659	356802	63889	290809	78993	373942
70	42708	177068	66312	295836	52787	230461	70735	326139
75	29001	111972	52555	223819	39989	161172	60014	262236
80	16876	57478	37783	143839	25554	90644	45545	178223
85	7402	21606	21315	68755	12304	37002	27295	89618
90	2243	5608	8231	20577	4014	10034	10991	27476

* These life tables are illustrative and should not be taken as implying accuracy of childhood mortality rates