

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

The estimation and interpretation of adult mortality rates of
African South Africans using Census 2001 data

Mark-Ross B. Dobbie

Centre for Actuarial Research (CARE)

University of Cape Town

This research was supported by the National Research Foundation (NRF), the Andrew Mellon Foundation and the UCT Postgraduate Funding Department.

ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation (NRF), the Andrew Mellon Foundation and the UCT Postgraduate Funding Department. Special thanks must go to my supervisor, Prof. Rob Dorrington.

University of Cape Town

ABSTRACT

This research develops estimates of mortality rates for adult Africans in South Africa for the twelve months preceding the census night, 9/10 October 2001, using Census 2001 10% sample data. The approach used to estimate these rates follow the work done by Dorrington, Moultrie and Timæus (2004) working with the full dataset, which is not publicly available, and demonstrate that the 10% sample can be used to produce similar results to the full database.

The approach makes use of indirect estimation techniques for estimating the completeness of reporting of deaths in the vital registration system at a national level, namely the combination of Generalized Growth Balance method (GGB) and the Synthetic Extinct Generations (SEG) method adapted to allow for net immigration over the inter-censal period.

Comparison of the number of deaths by sex and age from the vital registration system corrected for under-reporting with those reported by households in the census allows one to estimate adjustment factors that can be used to adjust deaths reported by households in various sub-groups (e.g. by province), on the assumption that reporting by households does not differ by these sub-groups, to produce more reliable estimates of mortality than if the deaths reported by households had been used directly to produce mortality rates.

Thus sex and age-specific mortality rates were produced by the following divisions: province; various household characteristics such as income level of the head of the household, employment status of the head of the household, and educational status of the head of the household; access to services; main power source used for cooking; toilet facilities; asset ownership of the household and access to piped water. These rates are analysed both univariately and multivariately, using negative binomial regression, to identify the socio-economic and other factors that are associated with differences in mortality.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	II
ABSTRACT	III
TABLE OF CONTENTS	IV
LIST OF TABLES	VI
LIST OF FIGURES	VII
1 INTRODUCTION	1
2 A REVIEW OF ADULT MORTALITY ESTIMATION TECHNIQUES AND RESEARCH	3
2.1 Socio-economic impact of HIV/AIDS and working-age mortality	6
2.2 Mortality rates for different socio-economic groupings	7
2.3 South African-specific adult mortality research	8
2.4 Indirect estimation techniques	11
3 CENSUS POPULATIONS, HOUSEHOLD DEATHS AND NET IN-MIGRATION OVER THE INTER-CENSAL PERIOD	15
3.1 South Africa population census 1996	16
3.2 South Africa population census 2001	18
3.3 Edited and non-edited Census 2001 10% sample data	18
3.4 Estimation of net in-migration over the inter-censal period.....	21
3.5 Household deaths in the 12 months prior to census 2001.....	22
4 METHODS USED TO ESTIMATE ADULT MORTALITY RATES	26
4.1 Generalized Growth Balance Approach.....	26
4.2 Allowing for migration	27
4.3 Allowing for differential census coverage and incomplete reporting of deaths	28
4.4 Bennett-Horiuchi (SEG) Method	29
4.5 Estimation of inter-censal deaths.....	30
4.6 Estimation of deaths of adult Africans for the calendar years 1996-2001	32
4.7 Completeness of deaths estimates.....	34

4.8	Generating adjustment factors for deaths in the year prior to census 2001	35
5	RESULTS	36
5.1	Estimated adjustment factors	36
5.2	Comparison of mortality rates with those of Dorrington, Moultrie and Timæus	38
5.3	Provincial mortality rates	39
5.4	Univariate analyses	41
5.5	Conclusion	60
6	MULTIVARIATE ANALYSES	62
6.1	Using multivariate techniques	62
6.2	Description of the approach	62
6.3	Findings	63
7	DISCUSSION	70
8	REFERENCES	74
9	APPENDIX 1: DERIVED ADULT AFRICAN MORTALITY RATES	80
10	APPENDIX 2: REPORTED DEATHS BY YEAR OF DEATH, 1997– 2001	82

LIST OF TABLES

Table 3.1	Edited and non-edited data for the population group and sex variables, Africans, Census 2001 10% sample	20
Table 4.1	Calculated completeness of death registration, adult Africans by sex, 9/10 October 1996 to 9/10 October 2001	34
Table 5.1	Correlation matrix for variables to be used in asset index	57
Table 5.2	Component loadings and summary statistics for variables used in developing the first principal component, census 2001 10% sample	58
Table 6.1	Negative binomial regression output for adult Africans using adjusted deaths, by sex, age group and enumeration area type, Census 2001 10% sample data set.....	65
Table 6.2	Negative binomial regression output for adult Africans using adjusted deaths, by sex, age group, income level and employment status of the household head, Census 2001 10% sample data set	66
Table 6.3	Negative binomial regression output for adult Africans using adjusted deaths, by sex, age, employment status and income level of the head and access to piped water, using Census 2001 10% sample data set	67
Table 6.4	Negative binomial regression output for adult Africans using adjusted deaths, by sex, age, employment status and income level of the head and access to piped water, using Census 2001 10% sample data set	69

LIST OF FIGURES

Figure 3.1	Population estimates by province, Census 1996 Community Profiles (descriptive) data set: African males and females combined	17
Figure 3.2	Age distribution by province, Census 1996 Community Profiles (descriptive) data set: African males and females combined	17
Figure 3.3	Edited and non-edited data for the age variable, Census 2001 10% sample data set: Males and Females combined.....	19
Figure 3.4	Edited and total entries distributions for the age variable, Census 2001 10% sample data set: Males and Females combined	20
Figure 3.5	Net in-migration estimates, African males, Census 1996 and 2001 Community Profiles and 10% sample data sets	22
Figure 3.6	Net in-migration estimates, African females, Census 1996 and 2001 Community Profiles and 10% sample data sets	22
Figure 3.7	Household deaths in the 12 months prior to 9/10 October 2001, African males by province, Census 2001 10% sample data set.....	23
Figure 3.8	Household deaths in the 12 months prior to 9/10 October 2001, African females by province, Census 2001 10% sample data set.....	23
Figure 3.9	Household deaths by edit type applied to age of the deceased, Census 2001 10% sample data set: African males and females combined	24
Figure 4.1	Distribution of reported deaths by age and year of death, African males, Statistics SA (personal correspondence).....	31
Figure 4.2	Distribution of reported deaths by age and year of death, African females, Statistics SA (personal correspondence).....	31
Figure 4.3	Estimated deaths by age and year of death, African males, 1996-2001	33
Figure 4.4	Estimated deaths by age and year of death, African females, 1996-2001	33
Figure 5.1	Plot of $n(x) - r(x+) + nm(x+)$ against $d(x+)$, African males.....	36
Figure 5.2	Plot of estimated ${}_5N_x$ to observed ${}_5N_x$, adult Africans.....	37
Figure 5.3	Extent of completeness of household deaths in the 12 months prior to 9/10 October 2001, adult Africans.....	38
Figure 5.4	Comparison of the calculated mortality rate estimates using the 10% sample and those rates using the full census data for adult African males and females (log scale)	39
Figure 5.5	Estimated mortality rates for adult African males by province (log scale)	40
Figure 5.6	Estimated mortality rates for adult African females by province (log scale)	40
Figure 5.7	Ratio of urban to rural mortality rates.....	43
Figure 5.8	Standardised distribution of female-headed households by age of household head by area type.....	44
Figure 5.9	Ratio of mortality rates for households using electricity as main power source for lighting to those using some other power source	45
Figure 5.10	Ratio of mortality rates for households using electricity as main power source for cooking to those using some other source	46
Figure 5.11	Ratio of mortality rates of those with employed household head to that of others	48
Figure 5.12	Ratio of mortality rates by completed education level of household head to those with incomplete primary: males.....	49
Figure 5.13	Ratio of mortality rates by completed education level of household head to those with incomplete primary: females.....	50
Figure 5.14	Ratio of mortality rates of those living in households where the head earns more than R9 600 per annum to those of other households	51
Figure 5.15	Ratio of mortality rates of those in households earning above R3 000 p.a. to those of those in other households: males	53
Figure 5.16	Ratio of mortality rates of those in households earning above R3 000 p.a. to those of those in other households: females	53
Figure 5.17	Ratio of mortality rates of those with access to piped water to those of those without access: males	55

Figure 5.18	Ratio of mortality rates of those with access to piped water to those of those without access: females	55
Figure 5.19	Ratio of mortality rates of those in houses with a modern toilet to those of those without	56
Figure 5.20	Ratio of mortality rates of the wealthiest households to those of those in the rest of the households	59
Figure 5.21	Ratio of mortality rates of those with refuse collection to those of those without	60
Figure 9.1	Estimated adult African male central mortality rates for the 12 months prior to 9/10 October 2001, by age and province	80
Figure 9.2	Estimated adult African female central mortality rates for the 12 months prior to 9/10 October 2001, by age and province	81
Figure 10.1	Adult African male deaths 1997-2001, Statistics SA (personal communication)	82
Figure 10.2	Adult African female deaths 1997-2001, Statistics SA (personal communication)	83

University of Cape Town

1 INTRODUCTION

The calculation of mortality rates using census data on numbers of people by age and sex (and other classifications) and the numbers of deaths reported by households by those classifications directly, without any adjustment, is often of limited use. Census data are often inaccurate due to both mis- and under-reporting by respondents (Bogue, 1965). Respondents might choose not to disclose certain information while enumerators might not allow respondents sufficient time to answer all the census questions (*ibid*). So, in order to derive more reliable mortality estimates, we need to adjust household deaths for mis-reporting.

This research seeks to derive robust mortality estimates for adult Black/Africans for the 12 months prior to the Census 2001 night, 9/10 October 2001, using publicly available census data and death data by population group and by year made available by Statistics SA (personal communication). In particular, we are interested in whether adult mortality estimates similar to those developed by Dorrington, Moultrie and Timæus (2004) using the full data set may be derived using the 10% sample data set. We do this in order to investigate the usefulness of the 10% sample as a tool for estimating mortality.

Sex and age-specific mortality rates are to be developed for the following divisions: province; various household head characteristics; access to services; main power source used for cooking; toilet facilities; asset level of the household and access to piped water. This is done in an attempt to identify the socio-economic and other factors that most explain differences in mortality rates between groups.

Estimates of adult mortality rates are derived using the approach followed by Dorrington, Moultrie and Timæus (2004). This approach makes use of a combination of the Generalized Growth Balance method (GGB) (Hill, 1987; 2001) and the Bennett-Horiuchi method (Bennett and Horiuchi, 1981; 1984), also known as the Synthetic Extinct Generations approach (SEG) with an allowance for net in-migration over the inter-censal period to adjust census data for relative under-enumeration and registered deaths in the inter-censal period for incompleteness of registration.

Dorrington, Moultrie and Timæus (2004) used these estimates to develop adjustment factors to correct household deaths for misreporting. This research uses the same approach to derive adjustment factors using the 10% sample data set and other publicly available information such as the vital registration data. Adult mortality rates are derived for different sub-groups using these calculated adjustment factors on the assumption that the level of completeness of reporting

of deaths by household for each sub-group was the same, which we use in the absence of anything better. The reasonableness of this assumption is discussed in Chapter 7.

Chapter 2 reviews various adult mortality demographic estimation techniques and other mortality research. Chapter 3 investigates census data using the 10% sample data set. Where data were missing or entries were clearly incorrect, Statistics SA applied various imputation techniques. The results of this imputation is considered in detail for those variables needed to develop our adjustment factors, in order to check that no biases were introduced to the variables used. Chapter 4 gives a description of the methods followed and includes descriptions of both the Generalized Growth Balance and the Synthetic Extinct Generations methods.

Chapter 5 sets out the estimated adult mortality rates for South Africa nationally and for the following sub-groups: provinces; various household head characteristics; access to services; main power source used for cooking; toilet facilities; asset level of the household and access to piped water. These sub-groups were chosen based on data availability and on the assumption that mortality differs by these sub-groups. These rates are derived in order to investigate whether the mortality of these sub-groups differs significantly. In Chapter 6, multivariate analyses of the socio-economic factors that explain mortality differences between the groups are undertaken using negative binomial regression. Chapter 7 discusses the research findings and draws conclusions about the implications of these findings. It also sets out possible avenues for future research.

2 A REVIEW OF ADULT MORTALITY ESTIMATION TECHNIQUES AND RESEARCH

In order to derive reliable mortality estimates, demographers require both suitable deaths and exposure (to the risk of dying) data. Mortality estimates may be derived directly by dividing the number of deaths by person-years of exposure. Of course, direct estimation will only produce reliable estimates if both the population used to estimate exposure and the number of deaths reported for that population are accurate.

An accurate and complete vital registration system will provide an accurate estimate of the number of deaths. Unfortunately, many vital registration systems are not complete: according to Gakidou, Hogan and Lopez (2004: 712) only 74 of the then 191 World Health Organisation (WHO) member countries had “complete and reliable vital registration systems in 2000”. As pointed out by Bah (1998b: 2) the collection of vital registration is determined by “a country’s political history, governance and the civil society” and, South Africa’s apartheid history has contributed in no small measure to the incompleteness of its vital registration system. In South Africa, deaths were not reported fully over the period 1996-2001 (Statistics SA, 2000b; Statistics SA, 2005), with Statistics SA estimating that only 67 per cent of deaths were reported in 1996. This is much lower than the 85 per cent estimate of Dorrington *et al* (2001), which can be attributed to errors in application of methods by Statistics South Africa. In addition, cause of death data that are collected in South Africa are also incomplete (Bradshaw *et al*, 2003).

More generally, Timæus and Jasseh (2004) note that adult mortality data are often limited due to infrequent data collection and, where data are collected, incompleteness of the data collected. Both poor and infrequent data collection hinder our understanding of mortality. Studies of adult mortality require large population exposures that yield a large number of deaths in order to derive estimates where any random variation has little impact. Such data might be collected by either surveying a large group of lives or by choosing a long period of investigation (Timæus, 1991) and, in addition, the answers to mortality questions should be provided by a reliable respondent if the collected data are to be accurate.

Estimates of adult mortality by country have been produced by various international organisations (e.g. Lopez *et al*, 2002; United Nations, 2003). However, these rates should be used with caution. For example, Timæus and Jasseh (2004) point out that the UN Population Division does not use adult mortality data for many African countries when developing its mortality estimates. Instead it assumes that the Princeton North model life table applies and calibrates the model table using estimates of under-five child mortality. So, the published results are highly

dependent both on whether the relationship between adult and child mortality assumed in the North model life tables holds for these countries and whether the shape of the model life table matches closely that of actual mortality. For example, in an HIV/AIDS environment, the actual mortality shape may differ markedly to that of the model life table. Accurate and complete mortality data are critical to demographers. Bourne (1995) gives a full listing of official South African mortality data for the period 1910-1992. These data refer to numbers of deaths by age of deceased and cause of death. The article does not, however, list the official life tables for this period. Bourne (*ibid*) also outlines some of the key pieces of legislation that governed the vital registration system in South Africa over this period. For example, the Births, Deaths and Marriages Registration Act (No. 17) of 1923 made it compulsory to register deaths in urban areas for all population groups, though the registration of deaths of Black/Africans in rural areas was voluntary. Since 1924, mortality data have also been reported by place of residence of the deceased (*ibid*).

But, while the death notification form includes place of residence, this is often mis-reported as place where the death was registered (Dorrington, Moultrie and Timæus, 2004). This is problematic should one want to derive mortality estimates by province or by district, for example.

With the Population Registration Act Repeal Act of 1991, population group was dropped from the civil registration system (Khalfani *et al*, 2005). Currently, the collection and reporting of vital events are governed by the Births and Deaths Registration Act of 1992 (Act 51 of 1992) (*ibid*). A new death notification form, which asks for the population group of the deceased on a separate page, was developed with the revision of the Act in 1998 and, from 1999 onwards, information on population group was recorded, and currently 75 per cent of all forms have population group recorded (Statistics SA, 2005).

At present, the vital registration system in South Africa is run by the Department of Home Affairs (DHA). The DHA collects deaths notification forms and then passes these registration forms to Statistics SA. Statistics SA then captures the cause of death on all death certificates and not only those with identification numbers and hence on the population register (e.g. Statistics SA, 2002; Statistics SA, 2005). The population register does not contain data on non-South African citizens living temporarily in South Africa nor of those lives that died before their births were reported (Khalfani *et al*, 2005).

Both the under-reporting of deaths and incomplete notification forms might in part be due to inattention of medical practitioners tasked with the responsibility of completing the death notification form (Statistics SA, 2005). In addition, particularly where the death occurs outside of

a healthcare facility, the under-reporting of deaths might be due to there being insufficient incentives on the part of potential respondents to report a death. So, while it is a legal requirement to register a death, it appears that this is not necessarily done in practice.

However, despite these problems, death notifications in South Africa have improved over the period 1989-2000 (Dorrington *et al*, 2001). Of course, derived mortality rates will be understated where deaths are under-reported, all other things being equal, and where no efforts are made to correct these reported deaths for under-reporting.

In addition to vital registration data, censuses and surveys can provide demographers with key mortality data. Gakidou, Hogan and Lopez (2004: 715) argue that “household surveys provide the only way, at present, to get timely information on the impact of new health problems”. Udjo (2005) argues that surveys and censuses provide important mortality data, particularly where vital registration data are incomplete. Timæus (1991) expands on the use of surveys and comments that retrospective survey questionnaires, both once-off and repeat visit surveys, may be key sources of mortality data.

Mortality research at both the Agincourt and Africa Centre demographic surveillance sites (e.g. Kahn *et al*, 1999; Kahn *et al*, 2007; Hosegood, Vanneste and Timæus, 2004; Hosegood *et al*, 2004) use repeat visit surveys to collect cohort mortality data over a period of time. These South African specific examples are considered in further detail later in this chapter.

Surveys might ask if there have been any deaths in the household in a fixed period prior to the survey date. For example, in Census 2001, households (i.e. respondents) are asked “Has any member of this household died in the past 12 months, i.e. between 10 October 2000 and 10 October 2001?” (2001 census questionnaire, Statistics SA). Of course, for household deaths to be reported, it is necessary that the household in which a death occurred continues to exist following that death. If not, the death will most probably go unreported (Timæus, 1991). We require both that households continue to exist following a death in the household and that survey and census questionnaires to be answered accurately, which might not be the case where respondents either do not know or choose not to report the correct answers (Timæus, 1991).

Timæus (1991) suggests that the under-reporting of deaths in census and survey data might also be due, in part, to enumerator negligence, whereby enumerators might choose to exclude questions on household deaths, on the basis that a death in a single household in, say, a 12 month period, is a relatively uncommon event. So, household deaths are typically under-reported in census and survey data. Where it is believed that deaths have been mis-reported, the statistical organisation that collected the data might choose not to make these data available (Timæus and Jasseh, 2004).

Repeat visit surveys are often able to provide high quality mortality data where there is little under-reporting of household deaths, as enumerators are able to question the reasons for the absence of any household members at the follow-up visit that were present at the initial survey date (Timæus, 1991). But, these surveys are costly and, as such, only relatively small populations are typically surveyed in this way making it difficult to extend any mortality findings arising from such surveys to other populations. As such, these surveys typically have small sample sizes and yield results that are not generalisable. In the next sub-section, we consider selected research on the socio-economic impact of HIV/AIDS and working-age mortality.

2.1 Socio-economic impact of HIV/AIDS and working-age mortality

A key research focus has been the socio-economic impact of HIV/AIDS (see for example Smith, 2002; Booyesen, 2003; Oni, Obi, Okorie, Thabede and Jordan, 2002; Monasch and Boerma, 2004). Booyesen (2003) finds that the severity of poverty for affected households, defined as those households with at least one member that is HIV-positive or where the household has suffered one or more HIV/AIDS related deaths in the past six months prior to the survey, is worse than for unaffected households, all other things being equal.

Similarly, Smith (2002) considers the household economic impacts of a HIV-positive male household head. These include reduced household income and increased medical costs, which often result in the household having to sell existing household assets and take on additional debt. Further, there are often opportunity costs with household members that would otherwise have been formally employed needing to care for sick members and children more likely to leave formal education early (*ibid*). The findings of Oni *et al* (2002) in their study of rural households in Limpopo Province are very similar to those of Smith (2002), namely that households affected by HIV/AIDS experience reductions in household income and increased medical and funeral costs.

Yamano, Jayne and McNeil (2002) investigated the impact of prime or working-age mortality on rural households in Kenya and found that household farming activities decreased following the death of a household member of working age. They note that the magnitude of the impact depended on both the sex of the deceased and position of the deceased within the household. Also, as a short-term coping mechanism, households would, following a death, typically sell small livestock, which complements the findings of Smith (2002).

Both Urassa *et al* (2001) and Hosegood *et al* (2004) argue that households are more likely to dissolve in the event of the death of the household head. Indeed, Urassa *et al* (2001) found that, in the context of their study, for the 44 per cent of households where the head of the household died, the remaining household members all left the household.

Also, there might exist a selection effect whereby sick lives will move to a “preferred place of dying” (Urasa *et al*, 2001: 2022). The authors believe there is evidence that suggests that urban dwellers in Tanzania that become terminally ill might choose to return to the rural areas as it is there that they wish to die. Should this be true, urban mortality might be under-stated relative to rural mortality (*ibid*), as sick urban people remove themselves from the urban population only to be reported as a rural death upon death.

2.2 Mortality rates for different socio-economic groupings

Timæus (1991) argues that it is often inappropriate to use any socio-economic factors that apply to a respondent as a proxy for the socio-economic factors of the deceased as such factors might not apply to the deceased. Further, any mortality investigations for different socio-economic groups will involve the sub-division of existing mortality data. While we might expect to increase homogeneity given appropriate sub-divisions, we might also reduce the credibility of any derived estimates due to the reduced exposures for each sub-group. Ideally, we would want exposure data to be sufficiently large to negate the impact of any random variation on the derived estimates.

Few studies focussing on adult mortality for different socio-economic factors and other groups have been undertaken. This is not altogether surprising given the data requirements that such studies would require. A description of selected studies that derive adult mortality estimates for different socio-economic groups follows. These studies consider mortality levels in the United Kingdom, Europe and Finland. While these studies are of interest to South African demographers, direct application, without adjustment, of the methodologies outlined in these studies in a South African context could be inappropriate.

In the *Health Inequalities Decennial Supplement* (Drever and Whitehead (ed.), 1997) adult mortality by “social class” is investigated, through the categorisation “social class” determined by the Registrar General using occupation data. The findings are based on analyses of the ONS Longitudinal Study data set, which is a one per cent representative sample of census and vital registration data, incorporating the 1971, 1981 and 1991 censuses linked together. The data set includes approximately 500 000 people.

Occupation type is collected at both the census date and at registration of death (*ibid*). Importantly, there may be reporting inconsistencies due to the different respondents at the census date and at the reporting of death. The social class field consists of six sub-groups ranging from “Professionals” through to “Unskilled” (*ibid*). The approach assumes that occupation type does not change over the exposure period, this might be problematic where there are substantial changes over the period. Individuals are assigned a social class level on entry based on census

occupation data, while individuals with missing occupation field are assigned a social level based on information on either the individual's spouse or parent, where available (*ibid*).

Mortality rates derived from these data were then used to estimate life expectancies for the different social levels. In addition to social class, the report also considers mortality by comparing life expectancies by employment status, access to a car and housing tenure. Of course, the selection of these factors is influenced by available data.

Gjonca (2003) also makes use of the ONS Longitudinal Study and investigates mortality differences by both education level and social class.

Similar to both Drever and Whitehead (ed.) (1997) and Gjonca (2003), Huisman *et al* (2005) analysed longitudinal mortality study data sets for eight European countries, whereby both vital registration and census records are linked. These countries included England and Wales, Norway, Finland, Austria, Switzerland and Belgium. Huisman *et al* (2005) also attempt to develop different social-economic groups and use completed education level from census data to do so. Specifically, the study considers only two education levels, namely High and Low.

Similar work was undertaken by Martikainen, Valkonen and Martelin (2001) for Finland, in which they investigated mortality differentials by socio-economic levels. These levels, Manual and Non-manual, are derived using an approach similar to that used in the *Health Inequalities Decennial Supplement* in that social class is determined using occupation data. For people where occupation data were not available, the authors used the derived social class level of the head of the household as a proxy for the social class level of those people in the household. Only lives aged 35 years and older were considered in order to reduce any possible unfavourable biases that might have been introduced when allocating people to social groups where occupation data were not available, i.e. we would not expect children to have permanent occupation. From the study data, the authors found that though life expectancies for both the Manual and Non-manual groups improved, mortality differences between the groups, for both males and females, increased over the study period.

Importantly, these studies make use of very accurate and complete enumeration and death data, i.e. it would be inappropriate to apply the same approaches as above to the South African data as we would need to correct for both census under-enumeration and incomplete deaths data.

2.3 South African-specific adult mortality research

Estimates of both the level and shape of mortality in South Africa in 1985 and 1990 were derived by Dorrington, Bradshaw and Wegner (1999), using reported deaths and indirect mortality estimation techniques to correct for incomplete death data.

Bradshaw *et al* (2004) comment that over the period 1998-2003 the reported number of adult deaths on the population register increased by 68 per cent. The authors point out that this represents a real increase of more than 40 per cent after considering both improvements in reporting and population growth.

Igumbor, Bradshaw and Laubscher (2003) investigated causes of death in Limpopo for the period 1997-2001. They used a random sample of death records for the province, drawn by Statistics SA. Similar work on the causes of death in Cape Town over the 2001-2004 period was undertaken by Groenewald *et al* (2007), using deaths data made available by the Cape Town City Health Department and population estimates developed by the Centre for Actuarial Research at the University of Cape Town (CARE). Mortality rate estimates were derived for each district, while districts were described in terms of various socio-economic factors; for example, by the proportion of households within a district without access to piped water or electricity. But, while mortality rates by district were estimated, the authors did not derive mortality estimates for different levels of socio-economic factors. The authors were unable to link the socio-economic factors to the death records due to the nature of the data available.

Udjo (1997) examined mortality levels and trends using the 1995 October Household Survey (OHS) data set. The OHS was conducted by the Central Statistical Services (currently Statistics South Africa) (*ibid*). Udjo developed life tables using child mortality estimates and adult mortality estimates, derived using the orphanhood method (*ibid*). Udjo (1999) then compared the mortality estimates derived using the OHS data set to rates derived using the Census 1996 data set and found that the 1995 OHS over-estimated mortality when compared to the 1996 census.

Anderson and Phillips (2006) estimate adult mortality for the period 1997-2004, using death notification data adjusted for under-reporting. Here, the authors estimate the extent of under-reporting of deaths by comparing reported deaths to estimates of deaths developed by Statistics SA using the population modelling software Spectrum, which is used to produce the mid-year estimates. Deaths corrected for under-reporting and mid-year population estimates were then used to derive adult mortality rates. This, of course, assumes the population model used by Statistics SA to be correct, and is a significant weakness of this research.

2.3.1 South African life tables

Van Eeden and van Tonder (1975) constructed abridged life tables from previously published comprehensive official life tables developed by the Department of Statistics (an early forerunner of Statistics SA) using census data over the period 1921-1970 for Whites, Coloureds and Asians. Life tables for Africans were developed using survival probability estimates produced by Sadie (1973). The authors did not apply any adjustments to the comprehensive tables to correct for any

under-reporting of deaths. Life tables published by Old Mutual, based on the observed mortality of its insured population, were also included. Of course, one might expect mortality of the insured population to be lighter than that of the aggregate population due to the fact that those insured might be considered to be in a higher socio-economic group than those that are not insured.

The abridged life tables were developed by extracting those l_x values from the comprehensive tables for those values of x where x was the lowest limit for each age group in the abridged table (*ibid*). Bah (1998a; 2005) pointed out that the official South African life tables published before 1985 were constructed on the assumption that reported deaths were complete. As such, one might expect the constructed life tables to understate mortality due to the fact that nothing was done to adjust for the completeness of reporting of deaths. Life tables produced by the national statistics agency for years 1921, 1926, 1936, 1946, 1951, 1960, 1970, 1980 and 1985 were constructed using census data and registered deaths data (Bah, 2005). These tables were published by sex and for population groups Whites, Coloureds and Indian/Asian. Thus, little is known about African mortality experience over this period.

Abridged Life tables for the period 1985 to 1994 and for 1996, for each population group, were published by Statistics SA (2000a). Tables for 1996 were constructed using registered deaths and census population estimates (Bah, 2005). Importantly, in developing its life tables for 1996, Statistics SA acknowledged that reported deaths were incomplete and attempted to correct for the under-reporting of deaths using the Growth Balance method (but applied to the whole age range). As one might expect the level of reporting of deaths of children to differ to that of adults, the application of the GGB to produce a single estimate of completeness of death reporting for all ages is a weakness of the approach used by Statistics SA.

2.3.2 Mortality estimates from demographic surveillance data

The Africa Centre study site is based in a rural area in KwaZulu-Natal of approximately 89 000 people as at 1 January 2001 (Hosegood, Vanneste and Timæus, 2004a) while the Agincourt study site, which was established in 1992, covers part of Bushbuckridge in the north-east of South Africa and consists of approximately 63 000 people (Tollman *et al*, 1999).

These surveillance sites collect demographic data using repeat visit retrospective surveys. The authors argue that the mortality data are complete (e.g. Hosegood *et al*, 2004), so mortality can be estimated directly without further adjustment. The completeness of mortality data is due to the fact that the surveys undertaken are repeat visit, which allows enumerators to collect better quality data than that from a one-off survey or census, all other things being equal. Mortality data were collected using verbal autopsy interviews of the primary caregiver of the deceased so as to

determine the cause of death (Hosegood, Vanneste and Timæus, 2004). The authors also derived retrospective adult mortality estimates using the orphanhood method and found HIV/AIDS to be the leading cause of death in adulthood for the study site in 2000.

Kahn *et al* (1999) and Tollman *et al* (1999) also make use verbal autopsies in order to collect cause of death data for the Agincourt field site. Tollman *et al* (1999) derived mortality rates by means of both direct estimation and indirect techniques, namely the orphanhood method. Kahn *et al* (2007), building on their earlier work, then developed age-specific mortality rates for the Agincourt site for the period 1992-2003. The authors found that mortality for both males and females increased over the study period.

Importantly, the derived mortality estimates apply to these study sites only. It is difficult to infer mortality rates to other geographical areas. This is a shortcoming of these studies as the results are not generalisable. Also, selective migration (mentioned earlier) might impact on the estimates of mortality, i.e. where sick lives move to the rural areas to die.

2.4 Indirect estimation techniques

In order to correct for incomplete information, various indirect estimation techniques have been developed. These methods typically require certain assumptions (e.g. that the population is stable, that the level of under-reporting constant across all ages) and, so, do not make up fully for an incomplete vital registration system (Timæus, 1991).

A number of demographers have reviewed both direct and indirect adult mortality estimation techniques (e.g. Timæus, 1991; Hill, 2003; Gakidou, Hogan and Lopez, 2004) and the following is a summary of some of the key methods and techniques that have been developed.

One group of indirect techniques derives adult mortality estimates using data on survival provided by relatives of the deceased. These methods were developed on the assumption that relatives of the deceased are most likely to correctly know of these deaths. These methods include the orphanhood, sibling survivorship and widowhood methods (e.g. Brass and Hill, 1973; United Nations, 1983; United Nations, 2002). Mortality estimates developed using these methods are estimates of average mortality derived from conditional survival over extended periods up to the survey (United Nations, 1983). These methods typically make use of simple questions that do not require respondents to provide information on either the date of death or the age of the deceased at death.

The orphanhood method (Brass and Hill, 1973), described in detail in United Nations (1983), uses information provided by respondents on the survival of their biological parents to derive adult conditional survival probabilities. These survival probabilities are conditional probabilities as the parents would need to be alive at conception, and birth in the case of

mothers. The method refers to biological parents and, where respondents report on the survival of their adoptive parents, mortality estimates will be under-stated. Hill and Trussell (1977) refer to this as the 'adoption effect'. The method derives mortality estimates for adults that are, or were, parents, and so, does not estimate the mortality of adults without children and to the extent that non-parents have higher mortality will lead to an underestimate of mortality for the population as a whole.

The sibling survivorship method was first developed by Hill and Trussell (1977) and is also described in detail in United Nations (1983). This approach also assumes that relatives of the deceased are most likely to provide the most accurate information on the deceased. Here, respondents are asked about whether their brothers and sisters, if any, are currently alive or not. Answers provided by respondents might not always be correct, particularly where respondents are not be aware of deaths of siblings that occurred either before the respondents' births or when they were very young (Timæus, 1991). An example of these methods being used is the research by Timæus and Jasseh (2004) where they use both orphanhood and sibling survivorship data from the Demographic and Health Surveys (DHS) to generate estimates of adult mortality for 23 sub-Saharan African countries.

The widowhood method (Hill, 1977; United Nations, 1983) also develops conditional survival probabilities, but makes use of information on the survival of a respondent's first spouse since marriage. This method derives mortality estimates for the ever-married population only and since the mortality of those not married is likely to be higher, the method underestimates the mortality of adults in the population as a whole. Importantly, this method assumes that the mortality experience of the deceased is independent of the spouse's mortality, and where this assumption does not hold, mortality estimates derived using this method will be biased (Timæus, 1991). This method might not be applicable where a high proportion of the population marry late, where marriage is a vague concept, such as South Africa, or where marriages often end in separation or divorce.

As has been mentioned, a suitable respondent must be alive in order for mortality information to be collected. Further, for these methods, there exists a time reference problem – we do not know exactly when a death, reported by a respondent, occurred (Gakidou, Hogan and Lopez, 2004; United Nations, 2002). This death might have occurred at any point since the event (birth, marriage, etc) from which survival is being measured. Brass and Bamgboye (1981) developed an approach to address this time reference problem, while Chackiel and Orellana (1985) argue that the problem might be avoided by asking further questions on the date of death. Alternatively, Gakidou, Hogan and Lopez (2004) suggest that one limit the period of analysis of

deaths to a fixed number of years prior to the survey date. Of course, while this provides more certainty on the timing of deaths, it also reduces the period of exposure.

A second group of adult mortality indirect estimation techniques uses information on the distribution of deaths by age. These methods include the Brass Growth Balance method (Brass, 1975), the Generalized Growth Balance method (Hill, 1987), the Preston and Coale method (Preston *et al*, 1980) and the Synthetic Extinct Generations method (Bennett and Horiuchi, 1981; 1984), which are also described in, among others, United Nations (1983), United Nations (2002), Timæus (1991) and Hill (2003). These methods compare the number of deaths by age to estimates of deaths from the population to obtain a level of under-reporting of deaths (United Nations, 1983).

Both the Brass (1975) and Preston and Coale (Preston *et al*, 1980) methods assume that the population is stable and closed to migration (United Nations, 1983). The Preston and Coale method requires information on deaths for a given period, by age for a given sex, and an estimate of population by age for that sex. The method estimates the level of under-reporting of deaths by comparing the population to an estimate of the number of future deaths arising from that population (*ibid*). By comparing reported deaths to the estimated number of deaths, the method yields a single adjustment factor for all ages, i.e. the method assumes that the completeness of reporting of deaths is the same for all ages. However, the method might yield unreliable results where age misreporting occurs. Indeed, for all methods that use an age distribution of deaths, any estimates will be inaccurate where significant age misreporting occurs (Hill, Choi and Timæus, 2005).

Like the Preston and Coale method, the Brass Growth Balance method (1975) also seeks to estimate completeness of adult death registration, but based on the following equation:

$$\frac{N(x)}{N(x+)} = r + \frac{D^*(x+)}{N(x+)} \quad (\text{United Nations, 1983})$$

The above equation is true for a stable, closed population (United Nations, 1983) and may be read as follows: the rate of entry into a population aged x years and older equals the stable population growth rate plus the rate of exit from the population aged x years and older. Here, $D^*(x+)$ refers to the true number of deaths, which can be re-stated as reported deaths divided by the estimate of completeness, assumed to be constant for all ages. For a practical example of these methods being used, see Dorrington, Bradshaw and Wegner (1999) whereby the authors makes use of a number of indirect estimation techniques, including the Preston and Coale and Brass Growth Balance methods, to derive estimates of mortality for African South Africans in 1985.

The Brass Growth Balance method was extended by Hill (1987) by relaxing the assumption of a stable population. The Generalized Growth Balance method (GGB), which requires population data from two censuses and the reported number of deaths in the inter-censal period, estimates both the completeness of reporting of the inter-censal deaths relative to the census populations and the relative under-enumeration of one census to the other (Hill, 1987; United Nations, 2002).

Bennett and Horiuchi (1981; 1984) suggested an alternative method, also known as the Synthetic Extinct Generations (SEG) method, which also relaxes the stable population assumption. The SEG method is an extension of the work done by Preston, Coale, Trussell and Weinstein (1980) and, using the age-specific growth rates of the population together with reported deaths in the inter-censal period, derives an estimate of the number of lives aged between x and $x+5$ (United Nations, 2002). The chief idea underlying this method is that, for a cohort of lives aged x now will equal the sum of all future deaths arising from this cohort until the cohort becomes extinct (Dorrington, Moultrie and Timæus, 2004). An estimate of the completeness of reported deaths might then be derived by comparing the expected number of people aged between x and $x+5$ to the corresponding group from the observed population.

These methods are described in more detail in Chapter 4 where they are used to estimate adjustment factors to correct for the mis-reporting of household deaths in the South African 2001 Census.

Inter-censal survival methods constitute a third group of indirect mortality estimation techniques. These methods are subject to significant bias arising from age misreporting and where relative under-enumeration between censuses exists (Hill, 2003) and are probably only used where there are insufficient data to allow other methods to be used. These methods typically assume a population closed to migration and accurate census data and use information from two censuses, though they can be modified to allow for international migration (Hill, 2003). These methods are based on the fact that, for an inter-censal period of n years, those aged $x+n$ at the second census will reflect those aged x at the first census, for a closed population, i.e. these methods track a cohort of lives and derive mortality experience for that cohort.

3 CENSUS POPULATIONS, HOUSEHOLD DEATHS AND NET IN-MIGRATION OVER THE INTER-CENSAL PERIOD

The estimation of adult African mortality rates by sex, age group and province for the 12 months prior to 9/10 October 2001 requires certain data. These include population estimates at both the Census 1996 and Census 2001 nights, deaths that were registered in the five year period between the two censuses and deaths reported by households in the 12 months prior to 9/10 October 2001.

Data sources used in this research include the Census 1996 and Census 2001 Community Profiles, together with the Census 1996 and 2001 10% sample data sets. The Community Profiles datasets are pre-tabulated output from the full censuses, but do not include data on deaths. The 10% samples are the unit records, which include data on deaths, migration and enumeration area type (urban/rural). The 2001 census 10% sample is drawn from the 2001 census as a 10% sample of households and collective living quarters and the homeless (Statistics SA, 2003a) and includes an indicator of type of edit, if any. The 1996 census was only minimally edited, while the 2001 census was edited extensively. Death notifications data were used to estimate deaths in the inter-censal period. These differed slightly to those used by Dorrington, Moultrie and Timæus (2004) who used data directly from the population register because Statistics SA had fallen behind processing the deaths. Thus the difference is the deaths with missing IDs.

Both the South African population censuses were *de facto* censuses (Statistics SA, 2003a). The 10% sample data set consists of a 10% sample of households and a 10% sample of collective living quarters and the homeless (Statistics SA, 2003b). An analysis of the Census 2001 10% sample data is necessary in order to investigate whether any unfavourable biases were introduced due to the editing of census data.

Edits, if any were used, were not identified in the Census 1996 10% sample of records, with unspecified responses indicated as such, while the 2001 census 10% sample data were used to evaluate the edits undertaken in 2001 as the Community Profiles data sets do not allow for edits to be identified. Statistics SA edited the 2001 census extensively and data entries that were edited were flagged in the 10% sample data set.

The 1996 and 2001 censuses provide initial best estimates of the population as at 9/10 October 1996 and 9/10 October 2001 respectively, provided that the edits used did not introduce inappropriate biases. So, it is important to establish whether possible biases were introduced to the data. This could not be done for the 1996 census as edits, if any, were not

flagged. The statistical software package *STATA 10* (StataCorp, 2007) was used to access both the census 1996 and census 2001 10% sample data sets.

This research is confined to data on the Black/African population group, which we refer to as “African” for brevity. The reason is that it is the majority population group and large enough to produce useful estimates of mortality when sub-divided.

3.1 South Africa population census 1996

Missing or not-stated entries exist and these were included in the Census 1996 10% sample data set. But, the Generalized Growth Balance (GGB) method, to be applied, corrects for the relative under-enumeration of census data. Missing entries are ignored and hence considered to be part of the undercount of the 1996 census relative to that of 2001.

3.1.1 1996 population estimates

Estimates of the African population as at 9/10 October 1996 by sex, age group and province of residence are required. A question relating to province of usual residence was included in the household questionnaire. But, this variable was subject to some editing, while province of enumeration was not. Since in aggregate the numbers are not very different it was decided to use data by province of enumeration.

Statistics SA (1998) created weights to correct for under-enumeration identified in a post enumeration survey (PES) carried out after the census. Thus the weights in the 10% sample data set reflect both the adjustment for undercount, Statistics SA (1998) estimates that 10.7 per cent of people were not enumerated in the 1996 census, as well as the adjustment required to scale the 10% sample up to the aggregate population.

People with age “not stated” or aged 120 years or older were excluded from the population estimates extracted from the Community Profiles data set as not reflecting the true age. By excluding these lives, we assume, in the absence of anything better, that the true age distribution of these people is that of the distribution of lives with age correctly stated. African population estimates at the 1996 census night by age and province for both males and females are given in Figure 3.1 and 3.2 respectively.

Figure 3.1 Population estimates by province, Census 1996 Community Profiles (descriptive) data set: African males and females combined

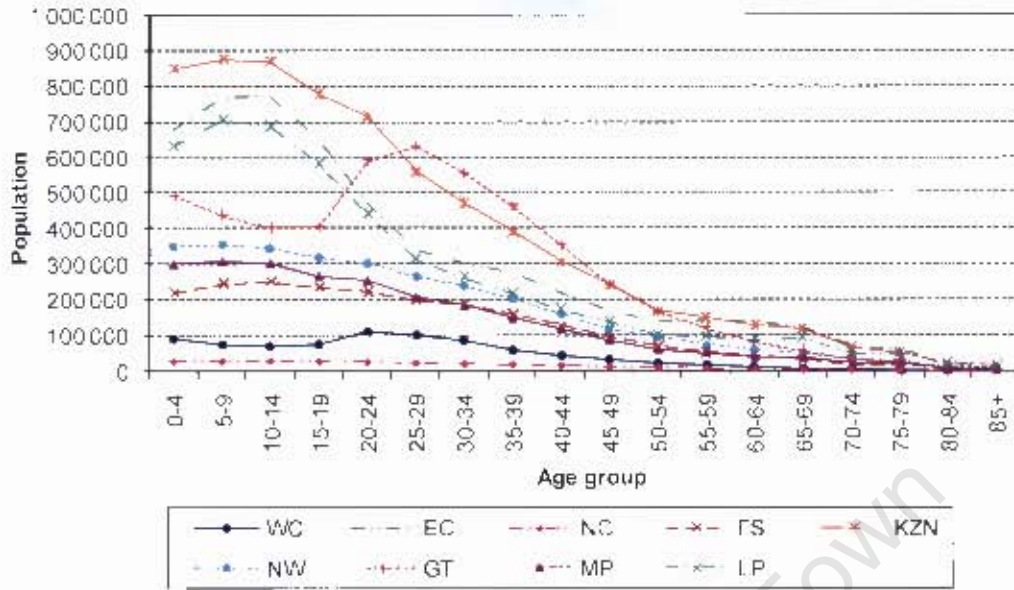
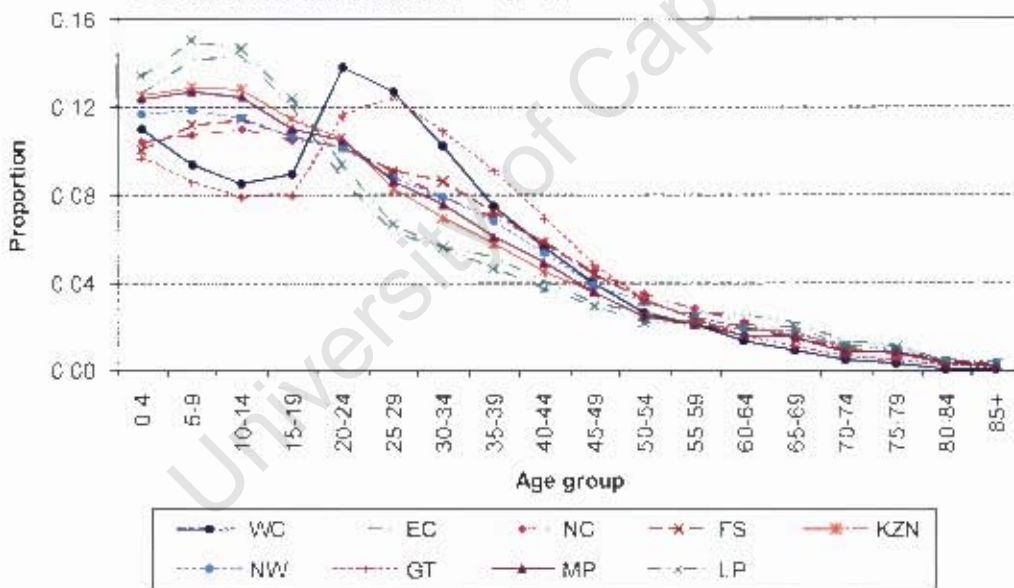


Figure 3.2 Age distribution by province, Census 1996 Community Profiles (descriptive) data set: African males and females combined



Abbreviations for the provinces, as used above and throughout, are: Western Cape (WC), Eastern Cape (EC), Northern Cape (NC), Free State (FS), KwaZulu-Natal (KZN), North West (NW), Gauteng (GT), Mpumalanga (MP) and Limpopo (LP).

The distributions of males and females are very similar. There were an estimated 414 315 African males in WC, while KZN had the largest African male population with 3 155 837. NC had the smallest number with 136 639 African males. Both WC and GT have a large proportion

of people in the economically active age range 20-34 years. This reflects the economic opportunities in the metropolitan areas and the resultant flow of migrants into these provinces. These migrants consist of both migrants from other provinces and migrants from outside the national borders.

3.2 South Africa population census 2001

Similar to Census 1996, Statistics SA made use of a post enumeration survey (PES) in order to correct for any coverage errors and weights were created to correct for the undercount (Statistics SA, 2003b). The weights in the 10% sample data set reflect, in part, these PES weights and in main part the scaling up of the sample.

Unlike the 1996 census the data missing and erroneous data were imputed in the 2001 census. Ideally the edits applied should not introduce any biases. So, it is necessary to investigate whether any such biases exist because of these edits.

In analysing possible bias, person weights were used when extracting census data, subdivided by edit type. Specifically, population group, age last birthday as at the census date and sex were all considered.

3.3 Edited and non-edited Census 2001 10% sample data

Many variables in the Census 2001 10% sample data set were edited, although the extent of editing differed by variable. Editing methods comprised four categories: logical imputation from blank; logical imputation from non-blank; hot-deck from blank and hot-deck from non-missing (Statistics SA, 2003b). Edits from blank were performed where cell entries were missing, while edits from non-blank were applied where there was a response to the question but was considered to be obviously incorrect or inconsistent with answers to other questions. The impact of editing on each of the key variables is discussed below.

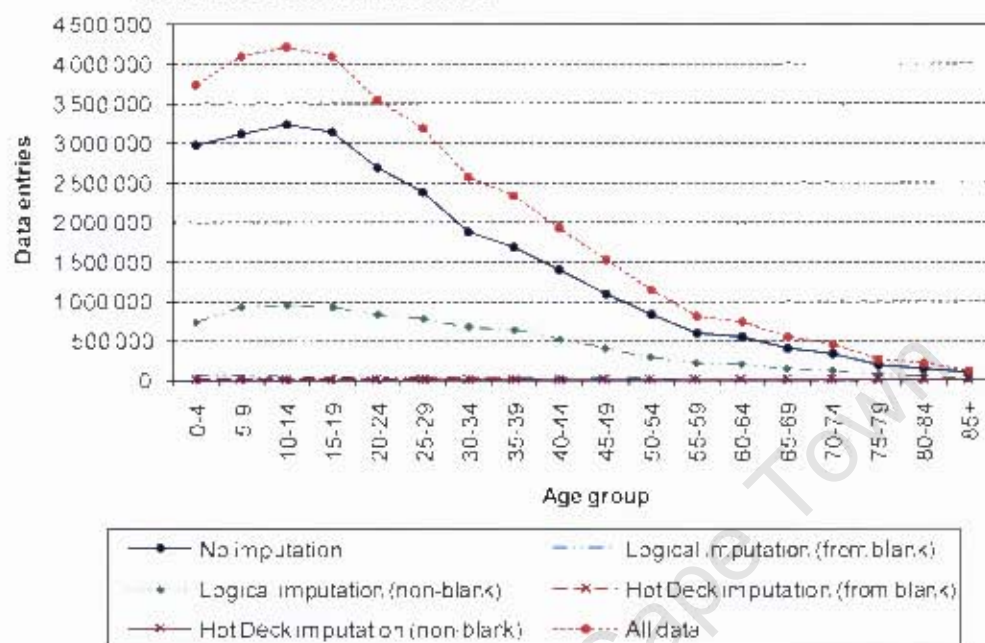
3.3.1 Age edits

The age variable for Africans was extracted by edit type. Entries with no imputation were also extracted, while the person weights were used to scale up the 10% sample data. In order to determine whether any unfavourable biases had been introduced by the editing process, distributions of edited data, data with no imputations and all data (both edited and non-edited) were derived.

Distributions with marked differences for a given variable would indicate the introduction of biases unless there was a good reason for the differences. Importantly, there may be a reason why the edited data should be different. Further investigations would then be needed

to decide whether the differences are appropriate. The age distributions by edit-type are given in Figure 3.3.

Figure 3.3 Edited and non-edited data for the age variable, Census 2001 10% sample data set: Males and Females combined

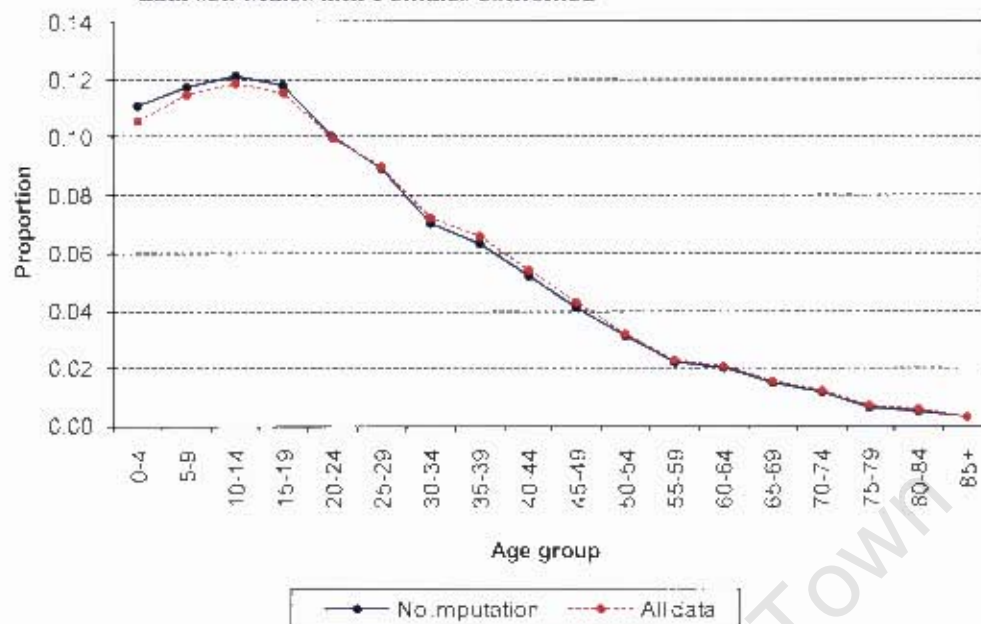


The patterns for males and females are very similar with edited entries accounted 24.7 per cent and 25.0 per cent of total entries (weighted observations) for males and females respectively. For males, no imputation accounted for 75.3 per cent of records; logical imputation from blank, 1.1 per cent; logical imputation non-blank, 23.4 per cent; hot deck imputation from blank, 0.2 per cent and hot deck imputation non-blank, 0.02 per cent. Similarly, for females, no imputation accounted for 75.03 per cent; logical imputation from blank, 0.97 per cent; logical imputation non-blank, 23.80 per cent; hot deck imputation from blank, 0.18 per cent and hot deck imputation non-blank, 0.02 per cent.

The edit types applied to males and females are very similar, while logical imputation edit types were most commonly applied to the age variable.

A comparison of the distributions by age of the data not requiring editing and the total records for males and females combined is shown in Figure 3.4.

Figure 3.4 Edited and total entries distributions for the age variable, Census 2001 10% sample data set: Males and Females combined



The distributions for those records with no imputation and all records are very similar for males and females combined. There is no evidence to suggest that significant bias has been introduced to the age variable by the editing process. So, the age variable in the Census 2001 10% sample data set can be used without further adjustment.

3.3.2 Population group and sex edits

Edit types applied to the population group and sex variables, as per the 10% sample, are summarised in Table 3.1. Africans, identified by population group, were extracted by sex and edit type. For the extraction, person weights were used, as before, to both scale up the 10% sample and to correct for undercount.

Table 3.1 Edited and non-edited data for the population group and sex variables, Africans, Census 2001 10% sample

Edit type	Population group		Sex	
	Males	Females	Males	Females
No imputation	16 675 387	18 291 181	16 712 103	18 269 213
Logical imputation (from blank)	154 719	166 129	103 035	158 719
Logical imputation (non-blank)	11 994	11 228	4 724	17 442
Hot Deck imputation (blank)	16 821	16 982	47 163	46 089
Hot Deck imputation (non-blank)	8 102	5 943	-	-

Source: Census 2001 10% sample data set

From the table above, approximately 1.1 per cent of entries classified as African were imputed, suggesting that, due to the small proportion of imputed values to total entries, no significant bias could have been introduced. The use of edits also does not alter proportion of African males significantly, with 47.8 per cent of Africans being male and 52.2 per cent being female according to the unedited data, while for all data, the proportions are 47.7 per cent and 52.3 per cent respectively. The edits applied do not introduce biases and the population estimates are used in the GGB and SEG methods to follow without adjustment.

3.4 Estimation of net in-migration over the inter-censal period

The Generalized Growth Balance method will underestimate the number of inter-censal deaths where net in-migration to South Africa over the inter-censal period is assumed to be negligible but is, in reality, positive. However, an adjustment to the method may be made to allow for net in-migrants where these are assumed to be non-trivial (Dorrington, Moultrie and Timæus, 2004). This correction requires that an accurate estimate of net in-migrants over the inter-censal period be available.

Three approaches were considered in an effort to estimate the net number of in-migrants over the inter-censal period. The first considered the specific migration question asked in the 2001 census, in which people were asked whether they had moved in the past five years and if so, where they had moved from. This, however, produced very low estimates of net in-migration. The second considered the differences between foreign-born in each census (Dorrington, Moultrie and Timæus, 2004).

The third approach was very similar to the second, although it considers citizenship other than South African instead of place of birth. Of course, citizenship might be problematic as it can change over time. Also, individuals might be more likely to deliberately misreport their true citizenship than place of birth, making this approach less likely to give accurate migration estimates.

The net in-migration estimates for each of the approaches above are plotted in Figure 3.5 and Figure 3.6. The curve labelled “Census 2001” gives those migration estimates derived using the specific migration question used in the 2001 census. All three approaches yield distributions of number of migrants by age with peaks at the economically active age groups. Migration estimates derived by considering the migration question appeared to be understated, and so it was decided to use a simple difference of the foreign-born, without further adjustment, to obtain an estimate of the number of in-migrants over the inter-censal period. Secondly, in doing so, we also make an assumption, in the absence of anything better, that out-migration of African South Africans over the period was negligible (Dorrington, Moultrie and Timæus, 2004).

Figure 3.5 Net in-migration estimates, African males, Census 1996 and 2001 Community Profiles and 10% sample data sets

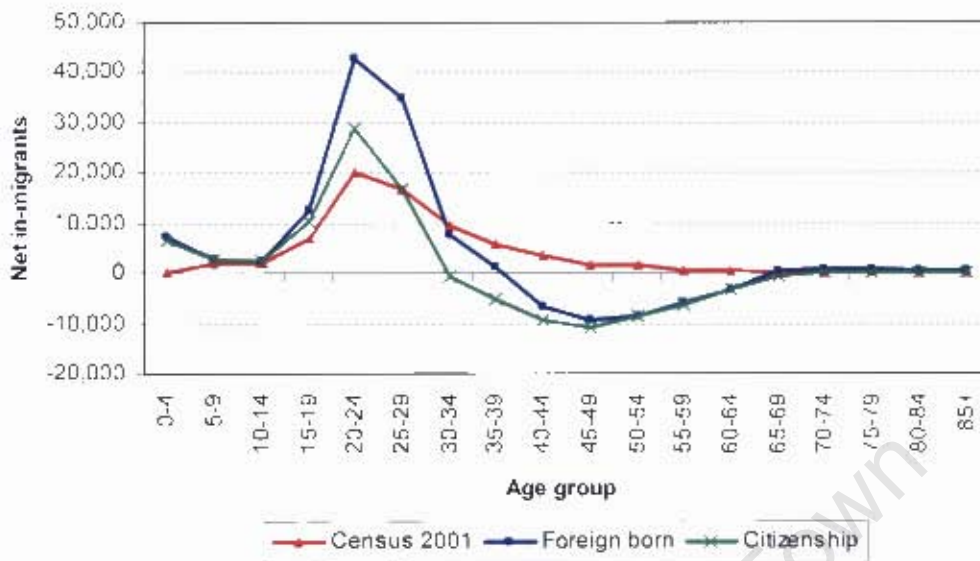
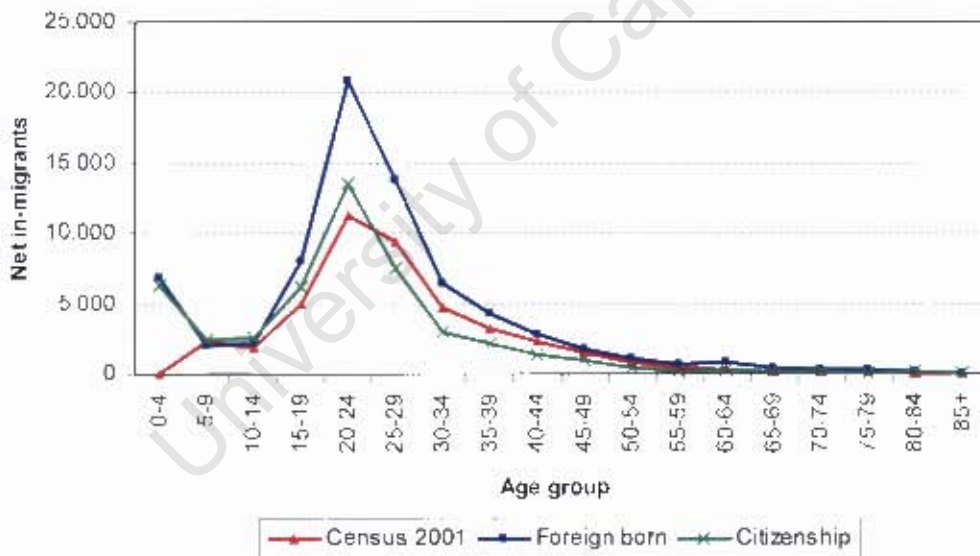


Figure 3.6 Net in-migration estimates, African females, Census 1996 and 2001 Community Profiles and 10% sample data sets



3.5 Household deaths in the 12 months prior to census 2001

Household deaths in the 12 months prior to the 2001 census night are not included in the Community Profiles data set and are extracted from the Census 2001 10% sample data set. These deaths are scaled up using the household weights, in the absence of anything better. Household weights are used to scale up households in the 10% sample to yield households at a national level, but these might not give an entirely accurate estimate of total household deaths as the household

weights might not correct for the under-reporting of household deaths, but there is no alternative. In addition, death estimates become more sensitive to adjustment at lower subdivisions. Population group of the deceased was not captured during enumeration. Instead, household deaths were assigned the population group of the majority population group within the household, on the assumption households with mixed population groups are fairly rare.

Household deaths for Africans, by sex, age group and province, are plotted in Figure 3.7 and Figure 3.8.

Figure 3.7 Household deaths in the 12 months prior to 9/10 October 2001, African males by province, Census 2001 10% sample data set

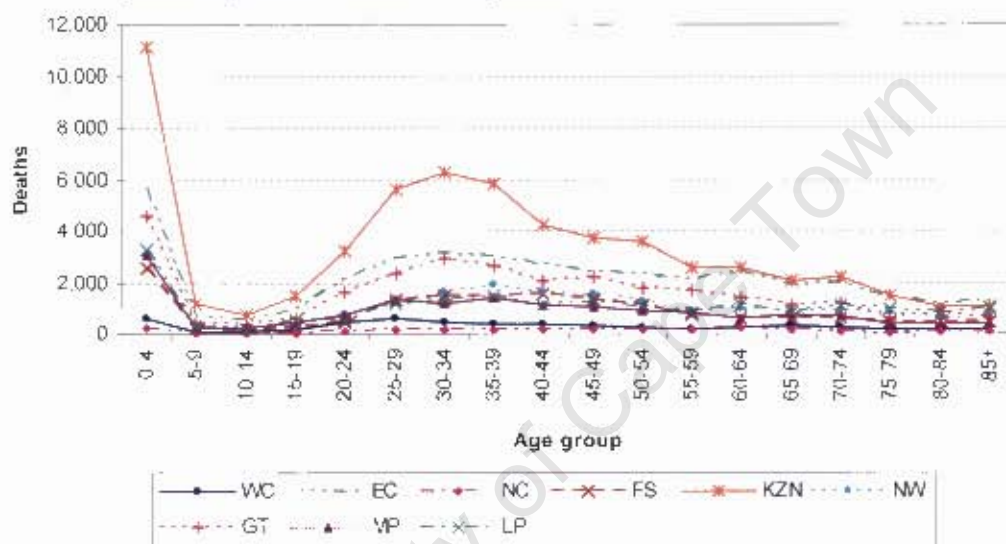
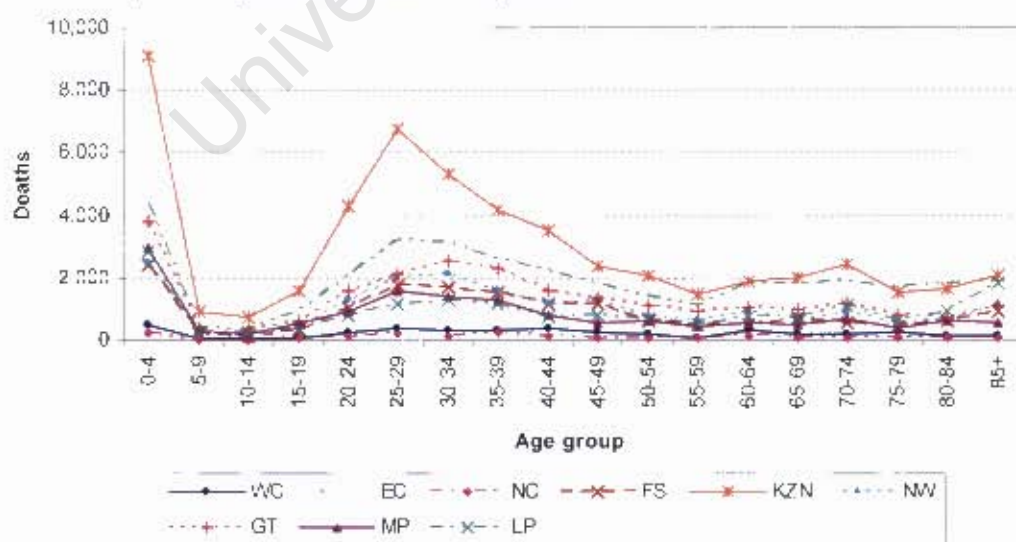


Figure 3.8 Household deaths in the 12 months prior to 9/10 October 2001, African females by province, Census 2001 10% sample data set

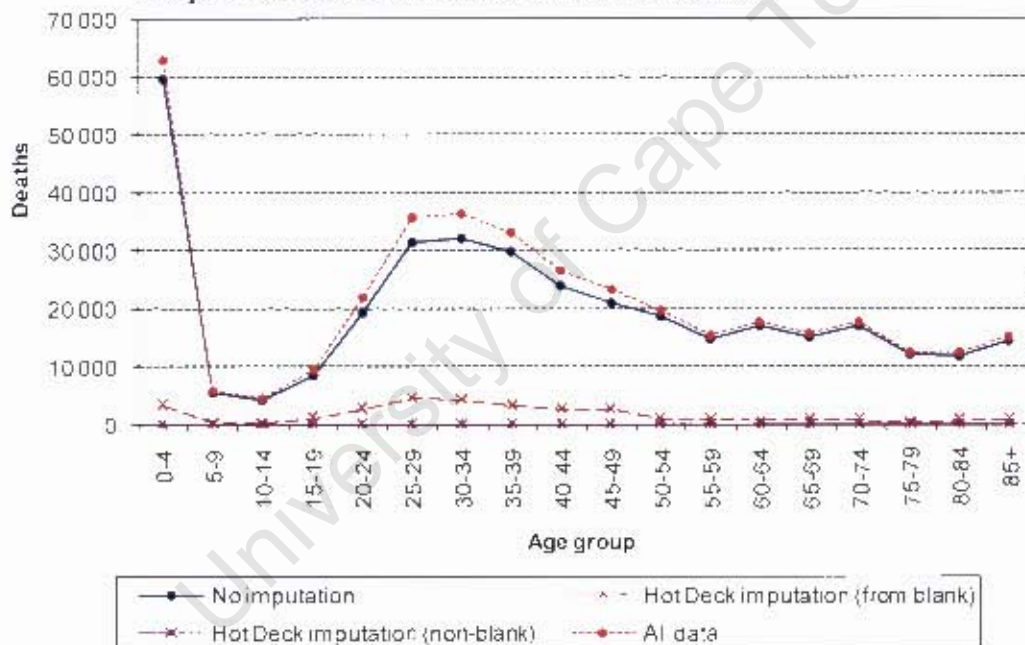


The plotted deaths distributions by age for males are very similar to those of females. Deaths are typically greatest at the youngest age group 0-4 years. KZN has the highest number of deaths of the nine provinces. From the above, we can see a pattern of deaths by age emerge. Deaths peak at the youngest age group but decrease rapidly and then rise again to form a local maximum over 25-39 years, the economically active age range that indicates the impact of HIV/AIDS mortality. This peak is perhaps most pronounced for KZN.

3.5.1 Edits of the death data

Edits were applied to both the age and sex variables of the deceased in the census data. In order to be satisfied that data are of high quality, and that no biases have been introduced to the deceased data set, an analysis of the edits carried out on these variables was conducted. Deaths distributions by edit type for both African males and females are plotted in Figure 3.9.

Figure 3.9 Household deaths by edit type applied to age of the deceased, Census 2001 10% sample data set: African males and females combined



From the figures above, it is clear that the edits applied have introduced bias as the deaths distribution derived using all data and that derived using only non-edited data are noticeably different. The edits applied have resulted in there being a greater proportion of deaths occurring at the economically active ages. However, this distortion was accepted and the reported household deaths used without further adjustment because both the overall effect is small and the numbers of deaths are going to be subjected to further adjustment to correct for under (or over) reporting.

The use of edits for the sex of the deceased amounted to 1.9 per cent of entries for deceased with household majority population group of Africans. The sex distribution for data with no imputation is 52.8 per cent for males and 47.2 per cent for females. When taking into account all data including edited entries, males accounted for 52.4 per cent and females 47.6 per cent.

University of Cape Town

4 METHODS USED TO ESTIMATE ADULT MORTALITY RATES

Adult mortality rates for Africans for the 12 months prior to the census night 9/10 October 2001 were derived using the same methods as used by Dorrington, Moultrie and Timæus (2004).

Both the GGB and SEG are used to estimate both the coverage of the censuses as well as completeness of death reporting. The Generalized Growth Balance (GGB) method (Hill, 1987; 2001) is used to estimate the relative under-enumeration of census estimates. The Bennett-Horiuchi (1981; 1984) (or Synthetic Extinct Generation (SEG)) method for calculating completeness of death estimates was then applied to the data, after adjusting the censuses for relative under-enumeration (Hill, Choi and Timæus, 2005).

The completeness of reporting of deaths for African males and females in the inter-censal period were estimated. These estimates of completeness are then used to produce estimates of completeness for individual years on the assumption, following Dorrington, Moultrie and Timæus (2004), that mortality rates at the older ages for each year showed no particular trend.

A comparison of household deaths in the 12 months prior to the census, extracted from the 10% sample data and deaths over the same 12 month period from death registration records after adjusting for completeness, allows for the calculation of age-specific adjustment factors by sex (Dorrington, Moultrie and Timæus, 2004) for correcting the census data for misreporting.

4.1 Generalized Growth Balance Approach

A detailed description of this method is given in the United Nations manual on estimating adult mortality (United Nations, 2002). The GGB method, proposed by Hill (1987; 2001) is an extension of the original Growth Balance method developed by Brass (1975) and described in United Nations manual on indirect estimation (1983). The explanation of the GGB method below is a summary of the description in Hill (1987; 2001) and United Nations (2002).

The method requires estimates of the population by age group at two census dates. Here, we use population estimates in five year age groups with open age interval at age 85 years. These estimates for the first and second censuses are referred to as $P_1(x)$ and $P_2(x)$ respectively. We also require deaths $D(x,5)$ that occurred between the two census dates, grouped by age at death into the same five year age groups with open interval 85+ years.

The GGB method does not explicitly make an allowance for migration over the inter-censal period, instead assuming it to be negligible (Dorrington, Moultrie and Timæus, 2004). However, as Dorrington, Moultrie and Timæus (2004) and others point out, the method may be adapted to allow for populations open to migration. So, estimates of net in-migration $NM(x,5)$ in

five-year age groups over the inter-censal period are required. These are estimated from the census 10% samples by differencing the numbers foreign-born people at the two census dates.

Estimates for people aged x and over for each population at each of the census dates, $P_1(x+)$ and $P_2(x+)$ respectively, were derived using the population estimates. Similarly, deaths aged x and older in the inter-censal period were calculated and are denoted by $D(x+)$. Average person-years lived for people aged x and above, $PYL(x+)$, were calculated using the formula $PYL(x+) = t[P_1(x+).P_2(x+)]^{0.5}$, where t is the length of the inter-censal period.

The number of lives that reach age x during the inter-censal period, $N(x)$, is estimated by taking one fifth of the geometric mean of those that reach age x in the five years following the first census and those that reach age x in the five years following the second census (United Nations, 2002): $N(x) = 0.2t[P_1(x-5,5).P_2(x,5)]^{0.5}$

Ignoring migration, the growth balance equation, on which the GGB method is based, is:

$$D(x+) = P_1(x+) + N(x) - P_2(x+)$$

Re-arranging this equation gives $N(x) - [P_2(x+) - P_1(x+)] = D(x+)$ and dividing through by the estimated person years lived $PYL(x+)$ gives:

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

where $n(x) = N(x) / PYL(x+)$, $r(x+) = [P_2(x+) - P_1(x+)] / PYL(x+)$ and $d(x+) = D(x+) / PYL(x+)$.

4.2 Allowing for migration

Migration is allowed for by adding $NM(x+)$ to the growth balance equation as follows:

$$D(x+) = P_1(x+) + N(x) - P_2(x+) + NM(x+)$$

where $NM(x+)$ is the net number of in-migrants over the inter-censal period. Re-arranging this equation yields $N(x) - [P_2(x+) - P_1(x+)] + NM(x+) = D(x+)$ and dividing this as before by $PYL(x+)$ gives $n(x+) - r(x+) + nm(x+) = d(x+)$.

4.3 Allowing for differential census coverage and incomplete reporting of deaths

Both the growth balance equation and the growth balance equation allowing for migration require the true values to hold. So, the symbols used above refer to the true values and assume that there is no under-enumeration or misreporting.

In practice, we would expect under-reporting and the equations that follow were developed to deal precisely with this situation. We can rewrite the equations to allow for this by letting k_1 and k_2 represent the completeness of each of the first and second census respectively and c be the completeness of the reporting of deaths.

People aged x and older at the first census, $P_1^*(x+)$ will equal $k_1.P_1(x+)$ where $P_1(x+)$ refers to the true values. Similarly, estimated people aged $x+$ years at the second census, $P_2^*(x+)$ will equal $k_2.P_2(x+)$, while $D^*(x+)$ will equal $c.D(x+)$. Substitution and manipulation of the balance equation with an allowance for migration over inter-censal period yields:

$$n^*(x) - r^*(x+) + nm(x+) = a + b.d^*(x+)$$

where $a = [\ln(k_1/k_2)]/t$ and $b = (k_1k_2)^{0.5}/c$. No completeness factors were applied to the estimated net in-migrants $NM^*(x+)$ as the approach assumes these to be correct. If this assumption was not made, we would also require a completeness factor for net in-migrants.

$n^*(x) - r^*(x+) + nm(x+)$ was then plotted against $d^*(x+)$ and a straight line was fitted to generate estimates for both intercept, a and the slope, b , as suggested by United Nations (2002) by setting the maximum of k_1 and k_2 equal to one. The smaller of these values then becomes a measure of relative completeness of the less complete census to the more complete census.

As mentioned, the method only allows an adjustment for relative and not absolute census undercount, though this is not a problem when estimating mortality rates as the completeness of deaths is also not absolute. Instead, completeness of deaths is estimated relative to the census populations. Completeness of reporting of deaths c can also be estimated using the estimates for k_1 and k_2 and an estimate of slope b .

Hill, Choi and Timæus (2005), however, suggest that best estimates of adult mortality rates are to be generated using an approach that combines both the GGB and SEG methods. Here, they suggest that one apply the GGB method first to estimate the relative under-enumeration of the censuses and then to use these adjusted population estimates in the SEG method to estimate the completeness of reporting of deaths.

A description of the SEG method follows.

4.4 Bennett-Horiuchi (SEG) Method

The SEG method is based on the idea that a number of lives at some point in time, t , will be equal to the sum of all subsequent deaths arising from this cohort up to when the cohort becomes extinct (Dorrington, Moultrie and Timæus, 2004). Rather than waiting until the whole cohort has died, the deaths and the population growth rates by age are used to estimate future deaths arising from the cohort (United Nations, 2002).

The SEG method was used to estimate completeness of recorded deaths. Population estimates as at the respective census nights $P_1(x)$ and $P_2(x)$ together with inter-censal deaths estimates $D(x,5)$ were required. These were used to determine age-specific growth rates. The formula used to calculate rates $r(x,5)$ but allowing for net in-migration is (United Nations, 2002):

$$r(x,5) = [\ln(P_2^{adj} / P_1^{adj})] / t - (NM(x,5) / [P_1^{adj}(x) \cdot P_2^{adj}(x)]^{0.5}) / t$$

The SEG approach requires that estimates of $N(x)$ be estimated from the estimates of the number of future deaths expected from the cohort age x now where $N(x)$ is the number of lives that reach age x as before. The following equations as suggested by United Nations (2002) together with the calculated age-specific growth rates $r(x,5)$ were used to estimate $N(x)$ as follows Bennett and Horiuchi (1981):

$N(x-5) = N(x) \cdot \exp[5 \cdot r(x,5)] + D(x-5,5) \cdot \exp[2.5 \cdot r(x,5)]$ for age groups other than the open interval; and

$N(x) = D(x+) \cdot \{ \exp[r(x+) \cdot e(x)] - ([r(x+) \cdot e(x)]^2 / 6) \}$ for the open age interval $x+$ where $e(x)$ is the future expectation of life estimate for a person aged x years exactly.

The equation for estimating the numbers at the age at which the open interval starts requires an estimate of expectation of life estimate for lives aged at that age. Since reliable estimates of $e(x)$ are not available, an alternative approach, as described in UN *Manual X* (1983), was used to estimate $N(x)$ for the open interval. Here, the number of people aged x was calculated using $N(x) = D(x+) \exp[r(x+) \cdot z(x)]$ where $N(x)$, $D(x+)$ and $r(x+)$ are defined as before and $z(x) = a(x) + b(x) + c(x) \cdot \exp[D(45+) / D(10+)]$.

Although the coefficients in *Manual X* were derived pre-HIV/AIDS and may not be wholly appropriate when estimating mortality in an HIV/AIDS environment, it is assumed that the impact of the epidemic on mortality and the fact that a relatively high age (85) is to be used for the open interval will limit the impact of any errors in the estimate of life expectancy that might arise. The West regional family coefficients were chosen in the absence of anything better and at such a high age for the open interval the choice does not make much difference.

Once $N(x)$ have been estimated for all age groups, these estimates are used to produce estimates of ${}_5N_x$ where ${}_5N_x$ represents the number of people in the age group x to $x+5$. The equation, ${}_5N_x = 2.5[N(x) + N(x+5)]$, as suggested by Bennett and Horiuchi (1981), is used to calculate ${}_5N_x$ estimates, denoted ${}_5N_x^{em}$. Observed ${}_5N_x$ values, denoted ${}_5N_x^{obs}$, were then generated using the equation ${}_5N_x = 5.(P_2^{adj} - P_1^{adj}) / \ln(P_2^{adj} / P_1^{adj})$.

These estimates of ${}_5N_x$ values, ${}_5N_x^{em}$, were then divided by the corresponding observed ${}_5N_x$ values, ${}_5N_x^{obs}$, i.e. ${}_5N_x^{em} / {}_5N_x^{obs}$ (Bennett and Horiuchi, 1981). Completeness of death registration was estimated to be some average of these ratios. Bennett and Horiuchi (1981) suggest that the median value be used to estimate completeness.

4.5 Estimation of inter-censal deaths

A report on registered deaths for the period 1997-2001 was published by Statistics SA (2005). This report was based on information obtained from the death notification forms received by Statistics SA from the Department of Home Affairs (DHA).

Deaths are often recorded at the place where the death is registered and not the place of residence of the deceased (Dorrington, Moultrie and Timæus, 2004). So, it was necessary to use household deaths from the census when calculating provincial mortality rates. Inter-censal deaths for adult Africans by sex and five-year age group with open age interval 85+ were required.

Deaths by population group were not reported fully for each of the years 1997-2001 (Statistics SA, 2005). The death certificate ceased to ask population group of the deceased in 1991 when the Population Registration Act was repealed (Khalfani *et al*, 2005), however, a new death notification form, which asked for population group of the deceased was developed and implemented in 1998 (Statistics SA, 2005). Old forms were received for recorded deaths in both 1997 and 1998 resulting, along with a reluctance by some to record population group, in an incomplete record of population group. According to Statistics SA (2005), from 1999 onwards, information on population group was recorded on approximately 75 per cent of all forms received.

Under-reporting of African deaths in the years 1997 and 1998 was suggested by the fact that reported deaths in years 1999, 2000 and 2001 being significantly larger than those in 1997 and 1998. The under-reporting of African deaths may have been due to both deaths not being recorded and to the under-reporting of African deaths where population group was not included.

The approach used to calculate inter-censal deaths for Africans was that used by Dorrington, Moultrie and Timanus (2004) who fitted an exponential trend to the recorded deaths for the years 1999, 2000 and 2001 to project backwards to estimate the number of adult African deaths by sex and age in the calendar years 1996, 1997 and 1998. The data on recorded deaths were made available by Elsie Mentz of Statistics SA (personal communication). These estimates of the number of deaths extracted from deaths notification data collected by Statistics SA are included in Appendix 2.

Figure 4.1 Distribution of reported deaths by age and year of death, African males, Statistics SA (personal correspondence)

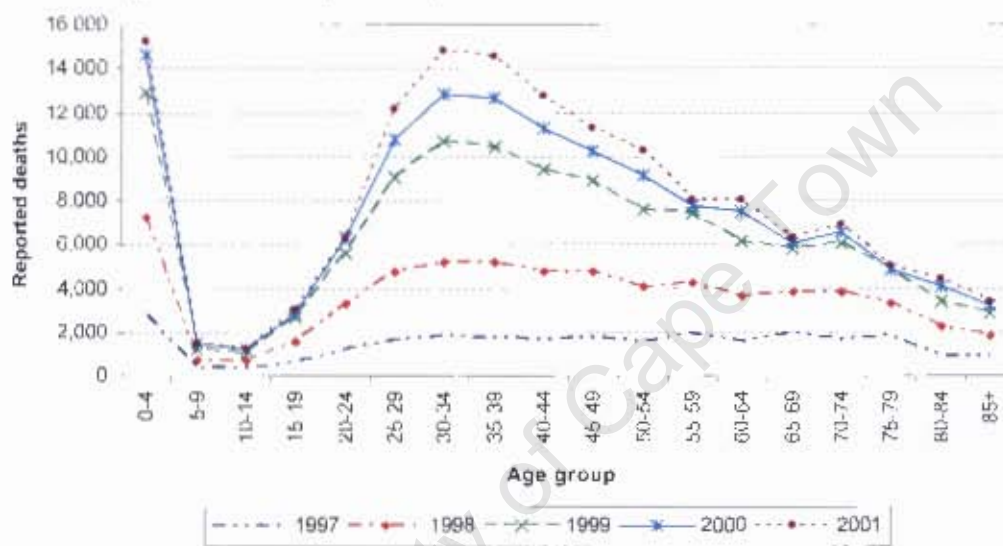
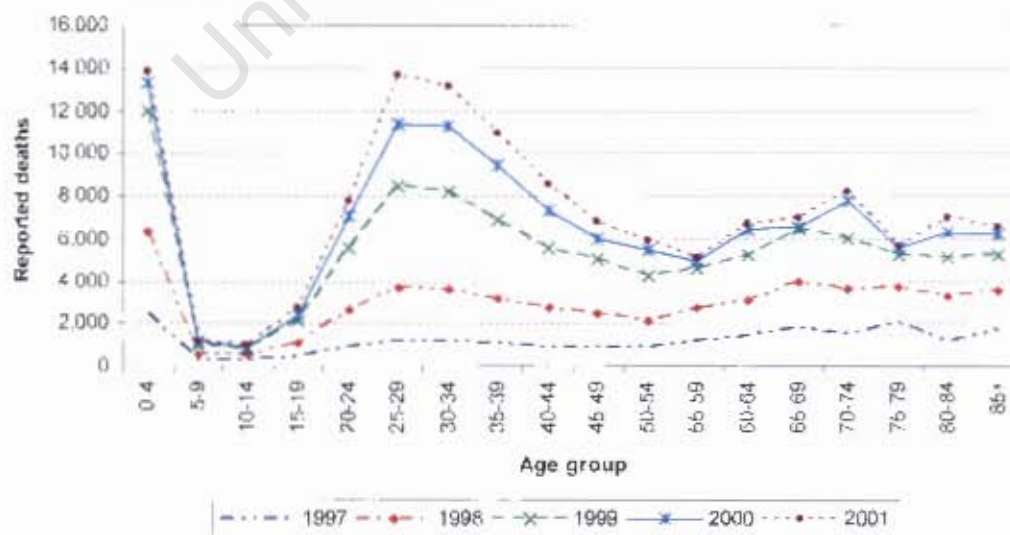


Figure 4.2 Distribution of reported deaths by age and year of death, African females, Statistics SA (personal correspondence)



It was necessary to correct for the under-estimation of deaths in years 1997-1998 where such under-reporting was due to the absence of population group.

Estimates of the number of inter-censal deaths, if taken directly from death notification data, as in Figure 4.1 and Figure 4.2 would be understated greatly. The approach used to estimate African deaths for calendar years 1996 through 2001 inclusive, using available death records, is described next. The application of the GGB and SEG methods yields a single completeness factor to correct for the under-reporting of deaths over the inter-censal period. However, this research requires only an adjustment factor for those deaths in the year prior to the 2001 census.

4.6 Estimation of deaths of adult Africans for the calendar years 1996-2001

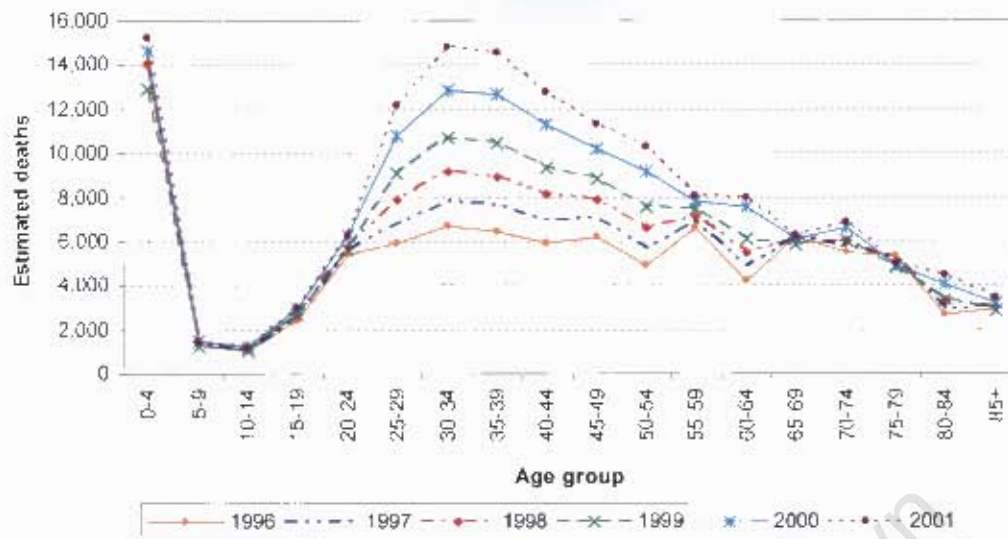
Recorded deaths data for years 1999, 2000 and 2001 were used to generate estimates of the number of deaths for calendar years 1996, 1997 and 1998. The estimation approach also follows the work of Dorrington, Moultrie and Timæus (2004) where deaths by sex and age group for the years 1996, 1997 and 1998 were estimated by back-projection techniques.

Age groups were distinguished on the basis of the number of deaths and the ages where HIV/AIDS could be expected to have an impact. Average deaths were assumed to apply to calendar year 2000 (*ibid*). Population estimates at each of the census dates were extracted using both census 1996 and census 2001 Community Profiles data sets. This approach was used where mortality rates were assumed to be largely constant over time. It would be inappropriate to use this estimation approach where mortality rates were increasing.

Knowledge of the impact of the HIV/AIDS epidemic suggests that mortality rates were rising for ages 25-64 for males and 20-59 for females. To allow for increases in mortality rates over time, exponential trends for each age group over these age ranges were fitted to the registered deaths data by sex and age group for years 1999-2001 (Dorrington, Moultrie and Timæus, 2004). The fitted trends were accepted only where the trend was convincing (with an R^2 of at least 85 per cent) so as not to force an exponential trend on the rates (*ibid*).

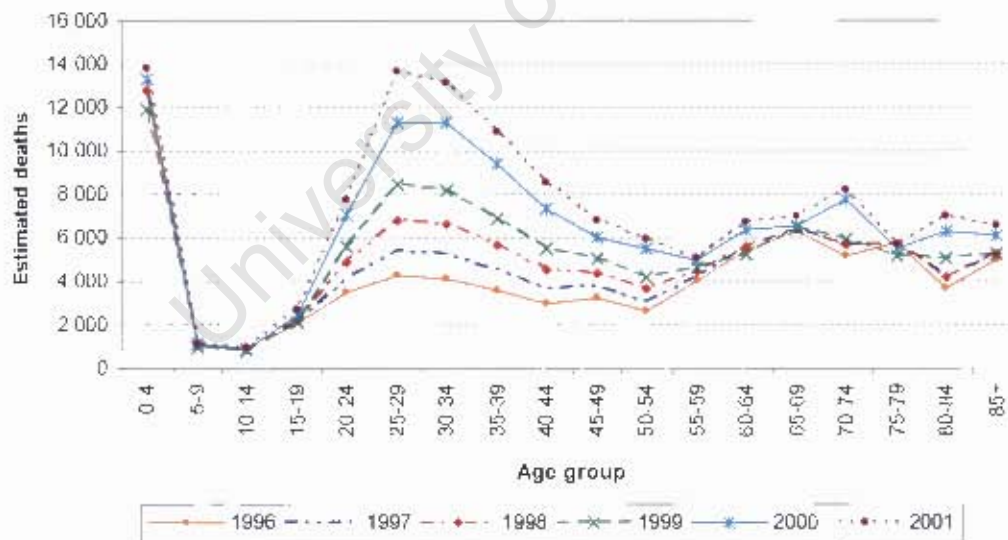
Figure 4.3 and Figure 4.4 plot the estimated numbers of recorded deaths of Africans by sex, age group and year of death using this approach.

Figure 4.3 Estimated deaths by age and year of death, African males, 1996-2001



Estimated deaths for ages 0-24 years for African males were largely relatively unchanged over the calendar years 1996-2001 in line with the population growth rates used to back project average deaths for years 1999-2001 together with assumption of constant mortality over the five year period. Similarly, the number of deaths over ages 65+ for males differed little by calendar year. But, deaths for ages 25-64 for males increased with calendar year of death.

Figure 4.4 Estimated deaths by age and year of death, African females, 1996-2001



Similarly, deaths estimated for ages 0-19 years and 60+ years for African women were largely similar over the calendar years 1996-2001. African female deaths increased by year of death for ages 20-59 years due to increasing HIV/AIDS mortality over the five year period. Where mortality was assumed to be constant, average deaths in years 1999-2001 were back-projected using estimated population growth rates to estimate deaths in 1996, 1997 and 1998.

while exponential trends were fitted where the age-specific mortality rates were assumed to be increasing.

The deaths were then converted from calendar years to estimates for each of the five inter-censal years with the first year starting 10 October 1996. These deaths, in turn, were summed to estimate deaths over the inter-censal period.

4.7 Completeness of deaths estimates

The SEG approach adjusts reported deaths for completeness and this adjustment for completeness applies to all recorded deaths during the inter-censal period. An assumption was made that the adjustment required to correct for the under-recording of deaths need not be constant over the five year inter-censal period (Dorrington, Moultrie and Timæus, 2004). This is in line with recorded deaths needing to be corrected both for under-count and for not recording population group of deceased.

To allow for differential under-reporting of deaths by year, it was assumed (following Dorrington, Moultrie and Timæus, 2004) that mortality of those in the group aged 65 and over shows no trend over time. Completeness by year was estimated (following Dorrington, Moultrie and Timæus, 2004) by assuming that the completeness changed linearly over the five year inter-censal period in a way that ensured no trend in those 65 and older. The estimate of completeness applicable to the full inter-censal period was, thus, applicable midway through the interval. A trend line was fitted to the 65+ mortality rates and the completeness was adjusted to yield the best fit horizontal trend line through the rates, this was suggested by Dorrington (personal communication).

The estimates of completeness for individual years are given in Table 4.1. To calculate mortality rates for the 12 months prior to 9/10 October 2001, only the completeness factor that applies to the year 2000/2001 was required. This factor was used to adjust the registered deaths to produce a best estimate of national adult African deaths during the year.

Table 4.1 Calculated completeness of death registration, adult Africans by sex, 9/10 October 1996 to 9/10 October 2001

	Completeness of death registration				
	1996/1997	1997/1998	1998/1999	1999/2000	2000/2001
Males	56.6%	58.1%	59.6%	61.1%	62.5%
Females	58.2%	60.0%	61.8%	63.6%	65.4%

These estimates are similar to those by Dorrington, Moultrie and Timæus (2004).

4.8 Generating adjustment factors for deaths in the year prior to census 2001

The adjustment factors were age- and sex- specific and were calculated by comparing registered deaths for the year prior to the census corrected for incompleteness of registration to the household deaths from census 2001 10% sample using the household weights to gross up the deaths from the 10% sample. Thus the adjustment factors are applied to the household deaths to yield adjusted deaths at a national level. These factors can then used to calculate estimates of adult deaths for each of the provinces and for other sub-groups on the assumption that whether or not a household reported a death as having occurred is independent of the sub-group into which the household falls.

Only adult mortality may be estimated using these methods as the GGB and SEG methods both assume a constant level of under-reporting. It is to be expected that the level of completeness of child deaths would be different from that of adult deaths.

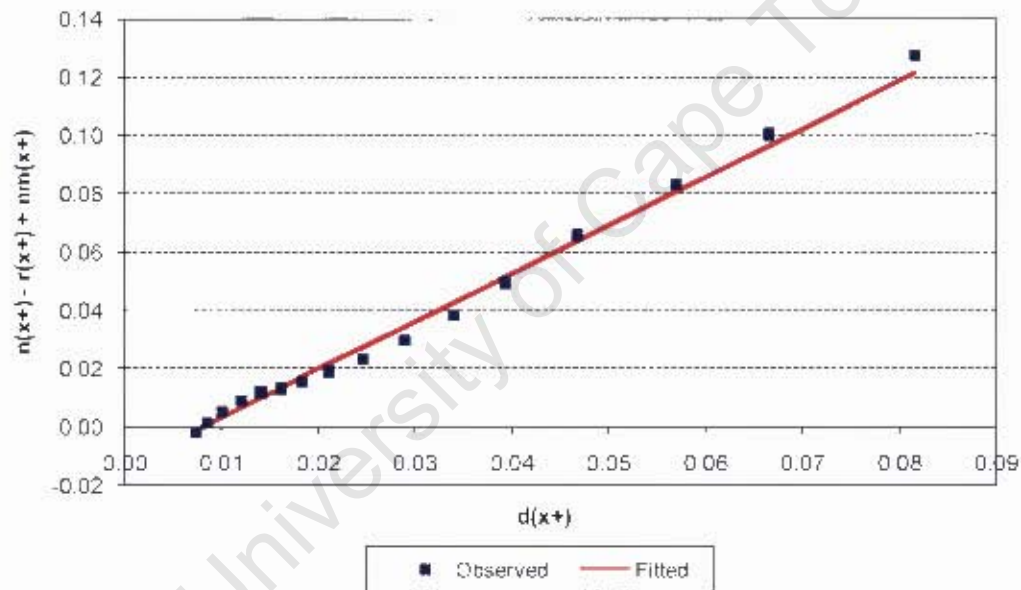
The estimated mortality rates for each province are given in Chapter 5.

University of Cape Town

5.1 Estimated adjustment factors

Application of the GGB method estimated the 1996 census to be 93 per cent complete relative to that of the 2001 census in the case of both African males and females, while application of the SEG method produced estimates of completeness of 60 per cent and 62 per cent for male and female adult registered deaths over the inter censal period respectively. In order to estimate the relative under-enumeration of the 1996 to the 2001 census, $n^*(x) - r^*(x-) + nm(x+)$ was plotted against $d^*(x+)$. From this, a straight line is fitted and the intercept of this line is used to relative under-enumeration. While this was done for both males and females, only a plot for males is included below, as the plot for females is very similar.

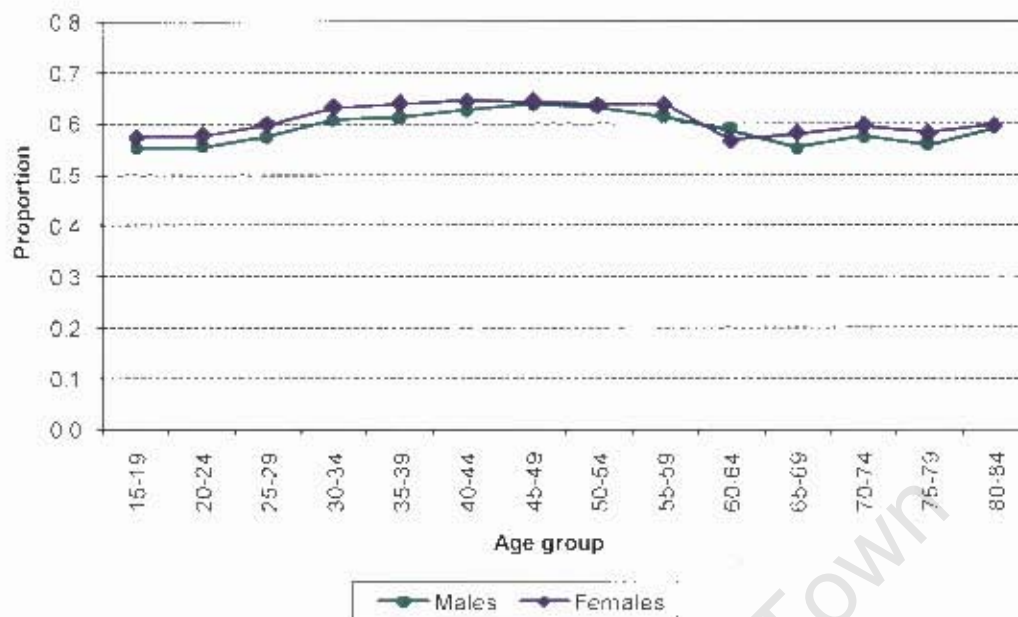
Figure 5.1 Plot of $n(x) - r(x+) + nm(x+)$ against $d(x+)$, African males



There is curvature in the points to which the straight line is fitted, so the points are not ideally linear and, as such, the estimates are approximate.

The GGB method was used to correct for the relative under-enumeration. Then the SEG method was applied using the adjusted census populations and the estimated ${}_5N_x$ values were divided by the corresponding observed ${}_5N_x$ values in order to estimate the completeness of death registration, by taking the median of these ratios over the age range 15-64 years (Bennett and Horiuchi, 1981). These ratios for African males and females are given in Figure 5.2 below.

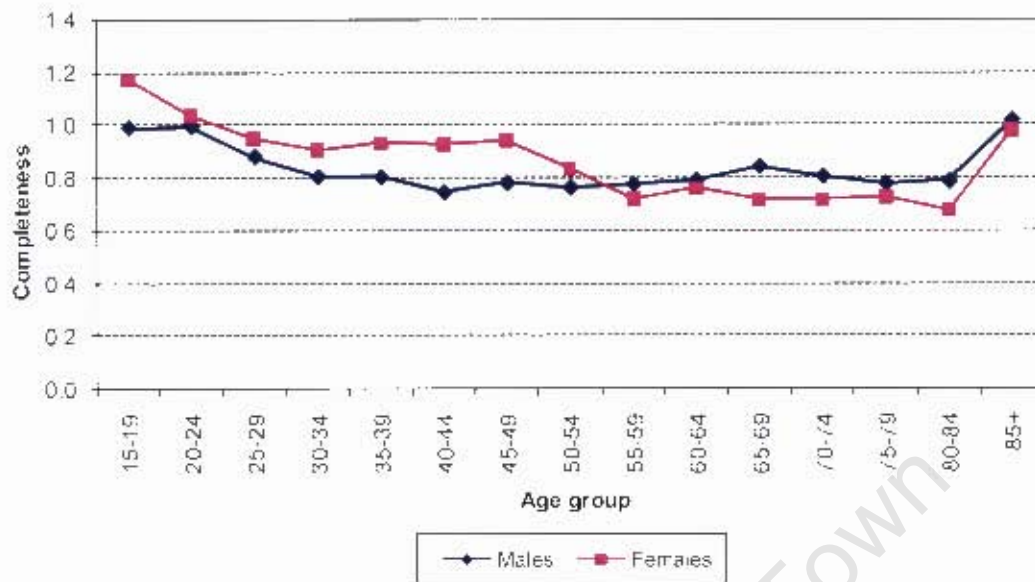
Figure 5.2 Plot of estimated sN_x to observed sN_x , adult Africans



These estimates of completeness of reporting of deaths are very low, but reflect both the under-reporting of deaths and the under reporting of population group of deaths that were reported. These were the average rates of completeness for the five year inter-censal period. It was also found that the recording of reported deaths improved over the five year period. The correction factor allowed for the estimation of the correct number of deaths for the whole period, but correction factors were only required for the 12 months prior to the 2001 census. The calculation of these estimates of completeness applicable to deaths in the year prior to the census is described in Chapter 4. These estimates of completeness were 63 per cent and 65 per cent for males and females respectively.

These adjustment factors were then used to adjust registered deaths and these deaths were compared to household deaths from the census (Dorrington, Moultrie and Timæus, 2004). By comparing the registered deaths corrected for completeness of death registration to the household deaths from the census, adjustment factors for household deaths were calculated. It was then assumed that these adjustment factors could be applied to the provincial household deaths as the level of reporting of deaths by household by province was assumed to be the same. The approach sought to adjust reported household deaths from the census to yield estimates of all deaths over the period. So, the calculated adjustment factors should correct for both the level of mis-reporting of household deaths and any institutional deaths not recorded in census. The extent of completeness of household deaths can be seen in Figure 5.3.

Figure 5.3 Extent of completeness of household deaths in the 12 months prior to 9/10 October 2001, adult Africans



Dividing household deaths by the level of completeness will yield best estimates of the number of household deaths by sub-group. From the above, we can see that the relative completeness for young adults is high. The high level of reporting at the younger ages might be due to these deaths being reported by multiple households until ages at which they are considered to be living in their own homes. The completeness estimates are derived using the assumption that the level of reporting of inter-censal deaths is constant over the adult ages and this might not hold here at the lowest age group.

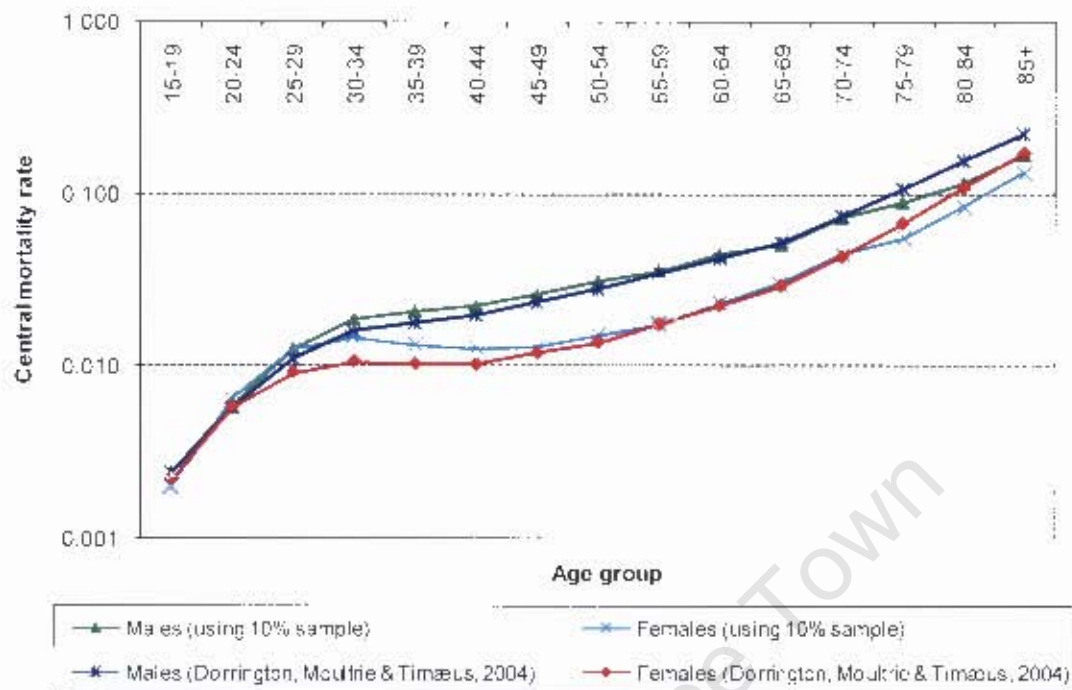
5.2 Comparison of mortality rates with those of Dorrington, Moultrie and Timæus

The adult African mortality rates derived by this research were compared to those produced by Dorrington, Moultrie and Timæus (2004) in order to determine whether it is possible to produce estimates which approximated those from the full census data using the sample data sets.

Mortality estimates derived by this research were almost identical to those developed using the full census data set.

Calculated mortality rates for adult African males and females using both the 10% sample and the full data are given in Figure 5.4.

Figure 5.4 Comparison of the calculated mortality rate estimates using the 10% sample and those rates using the full census data for adult African males and females (log scale)



Comparison of the calculated mortality rates indicates that mortality rates, calculated by this research, give higher rates at working ages 20-54 years, although the mortality at the older ages is higher for those calculated by Dorrington, Moultrie and Timæus (2004). Importantly, this research has used more recent deaths notifications data and could be seen as providing a set of updated mortality estimates as more complete data were used. The higher mortality shown by Dorrington, Moultrie and Timæus (2004) at the older ages is due to the fact that the authors graduated these rates to produce higher rates to compensate for possible age exaggeration. Work by Dorrington (personal communication) replicating Dorrington, Moultrie and Timæus (2004) but using the Statistics SA death data reproduced, almost exactly, the estimates of this research. Thus the higher mortality of the young adults, particularly females aged 25 to 44 years, is the result of better death notification data and not the fact that the 10% sample was used, while the results at the older ages are due to the fact that rates calculated are ungraduated.

5.3 Provincial mortality rates

As reasonable mortality rate estimates can be derived using the 10% sample, mortality rates by province are estimated by applying the adjustment factors developed earlier to adult deaths by province and by dividing these corrected deaths by average person years lived.

Estimated mortality rates by province, sex and age, were derived and are given in Figure 5.5 and Figure 5.6

Figure 5.5 Estimated mortality rates for adult African males by province (log scale)

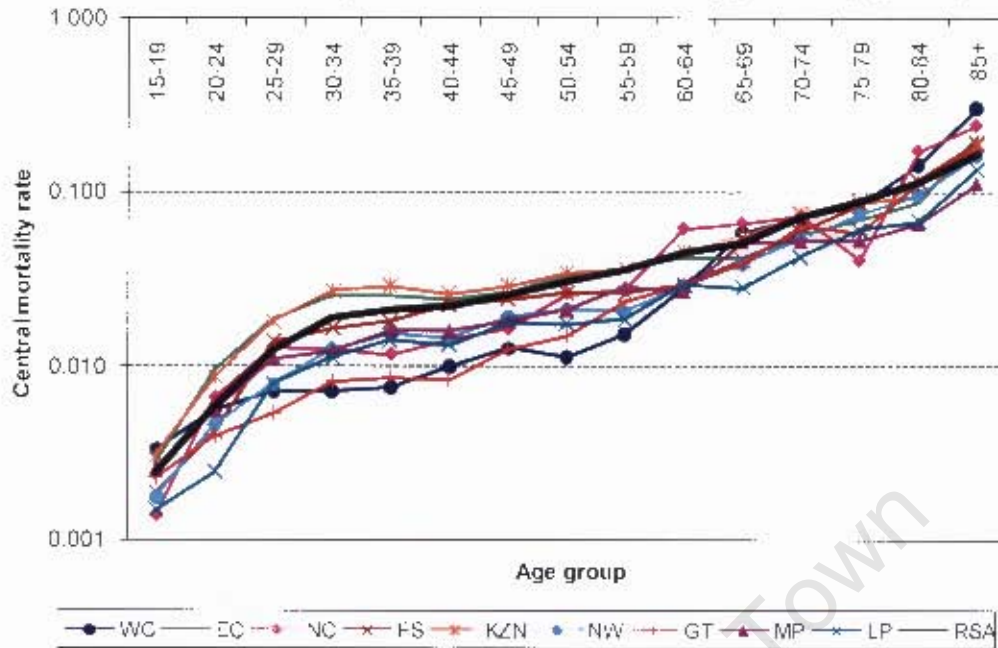
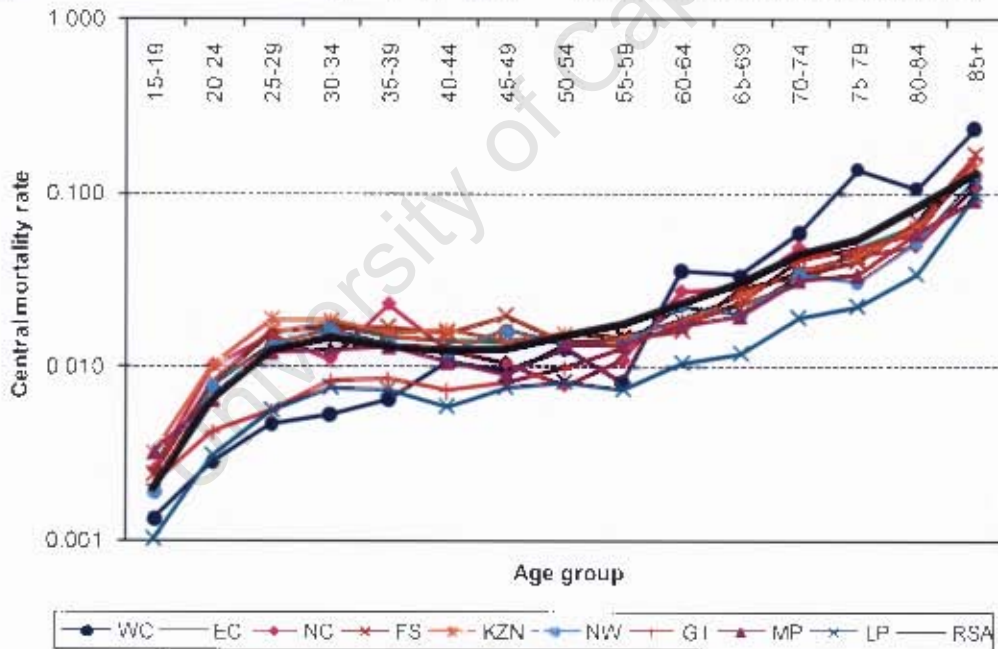


Figure 5.6 Estimated mortality rates for adult African females by province (log scale)



The uneven pattern or volatility of mortality rate estimates particularly at the older ages is due, in part, to the small number of deaths at those ages. This is most noticeable for the WC and NC where, for the latter, mortality at age group 70-74 is greater than that for 75-79 while that of the open interval 85+ is less than that for age group 80-84 years.

The WC, GT and LP lie below the national estimates while NC, FS, NW and MP have similar mortality to the national estimates. FS experiences slightly higher mortality relative to the national average over the younger ages and below national rates of mortality over the older ages. For African males, KZN and EC have high mortality rates in excess of the national rates. WC and GT have the lowest mortality rates for the 12 month period over 15-64 years. But, WC has very high mortality, in excess of the national rates, for ages 65+, suggesting that reporting is a function of province and age. If so, it would be inappropriate to apply the same adjustment factors to each sub-group.

High mortality rates for KZN relative to the national average, particularly over the working ages, might indicate the impact of HIV/AIDS on mortality. EC has high mortality relative to the national rates over 15-54 years while experiencing similar rates to those of the national estimates at the older ages 55-85+ years. NC has similar mortality to the national rates over the age range 15-54 years but relatively high mortality at the older ages. KZN has the highest rates of all nine provinces. We see that LP, particularly for females, has low mortality when compared to the country as a whole. Dorrington, Moultrie and Timæus (2004) comment on the relatively low mortality in LP and suggest that the mortality rates for LP are understated.

Adult African females, while experiencing lower absolute rates, experience similar relative patterns of mortality by province. The WC, GT and LP lie below the national estimates over the age range 15-59 years while NC, KZN, NW and EC lie above the national estimates. MP experiences very similar rates to the national average over the same age range. There is less variation about the national estimates relative to males and we see that the WC, FS and KZN experience high mortality rates compared to the country as a whole at the older ages.

5.4 Univariate analyses

We now investigate mortality rates for different sub-groups (other than by province) to see whether mortality differences exist for different socio-economic groups. This requires that we assume that the adjustment factors that apply at a national level also apply to different sub-groups of the national population, i.e. that the reporting of deaths by households is independent of the socio-economic sub-groups. Although this is quite a strong assumption, there is no evidence to guide us on this and this seems reasonable as the default assumption.

For example, the approach assumes that when considering mortality by the education level of the household head, the same adjustment factor to correct for the mis-reporting of deaths should be applied to households where the household head has incomplete primary education to those where the household has completed secondary education. We also assume

that the socio-economic categories that apply at the census night applied in the 12 months preceding the census.

The mortality file in the 10% sample gives very little information on the deceased, relevant information to this research includes only age and sex of the deceased. However, household variables can be linked to individual death records by the common variable *serialno*, which is a unique number assigned to each household in the Census 2001 data set. So, this allows us to link a variable such as type of toilet in the household to the specific death record from that household where a death occurred.

Socio-economic groups were determined by the variables in the 10% sample. Variables of interest included: whether a household was situated in an urban or rural area; various household characteristics such as income level of the head of the household, employment status of the head of the household, and educational status of the head of the household; and access to services; main power source used for cooking; toilet facilities; and access to piped water. These socio-economic variables are chosen in line with both evidence that mortality differ by these categories and the availability of data.

Mortality rates were developed by extracting household deaths for each socio-economic group by age and sex. These deaths were corrected for under-reporting using the adjustment factors calculated earlier and were then divided by an estimate of population of the respective socio-economic group as the mid-point of the period of 12 months prior to the census night.

The mid-year population estimates were derived using the 1996 census, corrected for relative under-enumeration, and 2001 census 10% samples where variables across censuses were consistent. We assume that the adjustment for census coverage and completeness of deaths applies equally for each socio-economic sub-group. Age-specific growth rates over the intercensal period were used to project the socio-economic population estimates as at 9/10 October 2001 exactly six months prior to the census night.

The population group of the deceased is not included in the mortality data set and so was estimated using the household variable *popgroup*, which is the majority population group within a household (Statistics SA, 2003b). Person weights and household weights were used to scale up the respective population and deaths accordingly.

Mortality rates by specific socio-economic groups were estimated for adults aged 15 years and older, though age ranges for which estimates were calculated were truncated where the exposures were very low and the results appeared implausible. Mortality rates were estimated for those levels of variables where it was thought that the level of the variable would have a significant impact on adult mortality. Mortality rate ratios and approximate 95% confidence

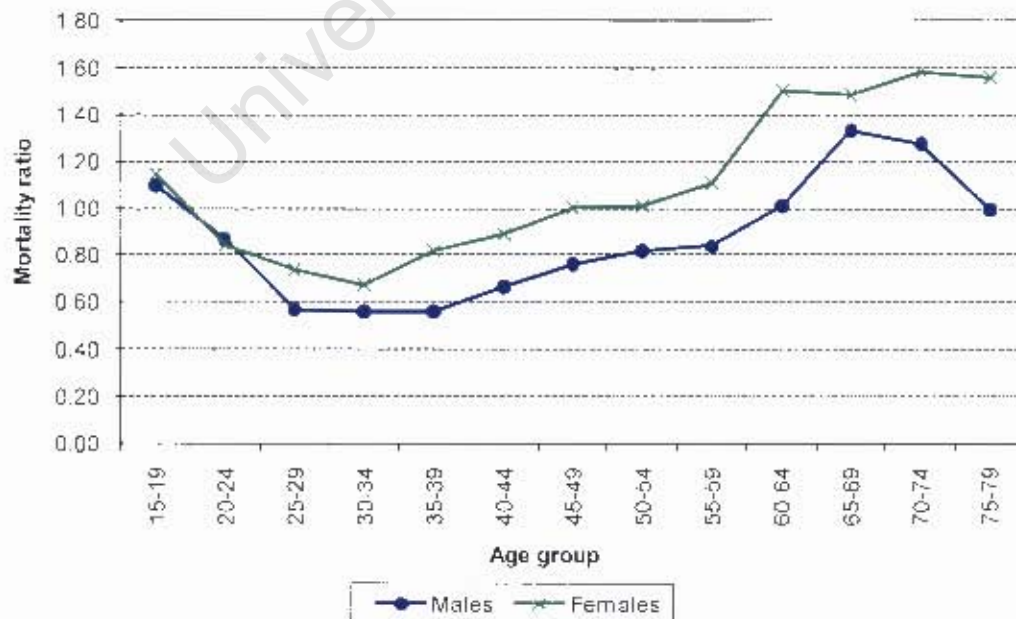
intervals for these ratios were developed using the mortality rates. The confidence intervals were derived using the assumption that deaths have a Poisson distribution, i.e. deaths have equal mean and variance, and that this distribution is approximately Normal. Then, random estimates were derived for both the numerator and denominator, which were then used to derive random estimates of the ratio. The confidence intervals were then derived using simulation techniques. For ages 65 to 69 years, the estimated confidence intervals are approximately of width 4 to 6 per cent, approximately 4 to 7 per cent for ages 70 to 74 years and approximately 5 to 8 per cent for ages 75 to 79 years. As such, if a ratio equal to one lies in this interval, then the ratio is not significantly different to one. Of course, the consideration of the socio-economic factors that follow is limited to those variables included in the 10% sample and so cannot be seen to be a complete listing of those factors that impact mortality levels.

5.4.1 Enumeration area (urban/rural)

Mortality rates might be expected to differ between those living in urban areas and those living in rural areas since socio-economic living standards, in particular access to health care facilities, waste disposal initiatives and sewerage, may differ significantly. To investigate whether a mortality differential does exist by whether an area is an urban area or not, mortality rates for both urban and rural areas were derived using the approach described earlier. Areas were coded as urban based on whether the household is in an area classified as urban in 1996 (Statistics SA, 2003).

Age- and sex- specific mortality rates for adult Africans by urban-rural residence were estimated and the ratios of urban to rural rates are plotted in Figure 5.7.

Figure 5.7 Ratio of urban to rural mortality rates



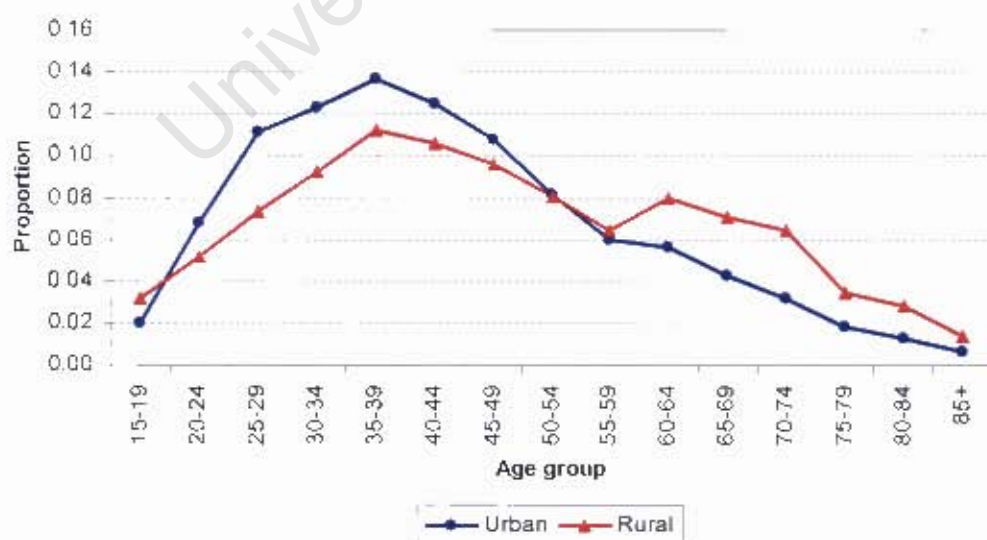
These ratios indicate that the mortality rates for those living in urban areas are lighter than their rural counterparts for much of the working age range. This difference appears to be less marked for females. Rates over age 60 appear to be higher in the urban areas, particularly for females, although for males the rates are subject to random variation due to small numbers. This result seems implausible. We would expect rural females to be at a disadvantage to their urban counterparts in terms of income and access to health facilities and safe drinking water.

However, it is possible that the same adjustment factors should not be applied to urban and rural female deaths for the older age groups, in particular it is possible that some rural people move to urban areas for health services.

A closer look at the distribution of female headed households by urban/rural split might explain this anomaly. The adjustment factors applied might be inappropriate should rural households be worse at reporting female deaths in the household.

Figure 5.8 plots the proportion of female-headed households by age of household head for both urban and rural areas. It can be seen that there is a greater proportion of female headed households at the older age groups, specifically 55-74 years, in rural areas than in urban areas. For a household to report a death in the 12 months prior to the census night, the household must remain intact following that death up to the census night. In the event that a female household head dies, we might expect the subsequent dissolution of that household with remaining members joining other households. In turn, that death will most probably go unreported, resulting in a possible undercount of female deaths at age the older ages in the rural areas.

Figure 5.8 Standardised distribution of female-headed households by age of household head by area type



While the mortality rates for men in the younger working ages is noticeably lower for those using electricity for lighting, there is no noticeable difference for women or for older men.

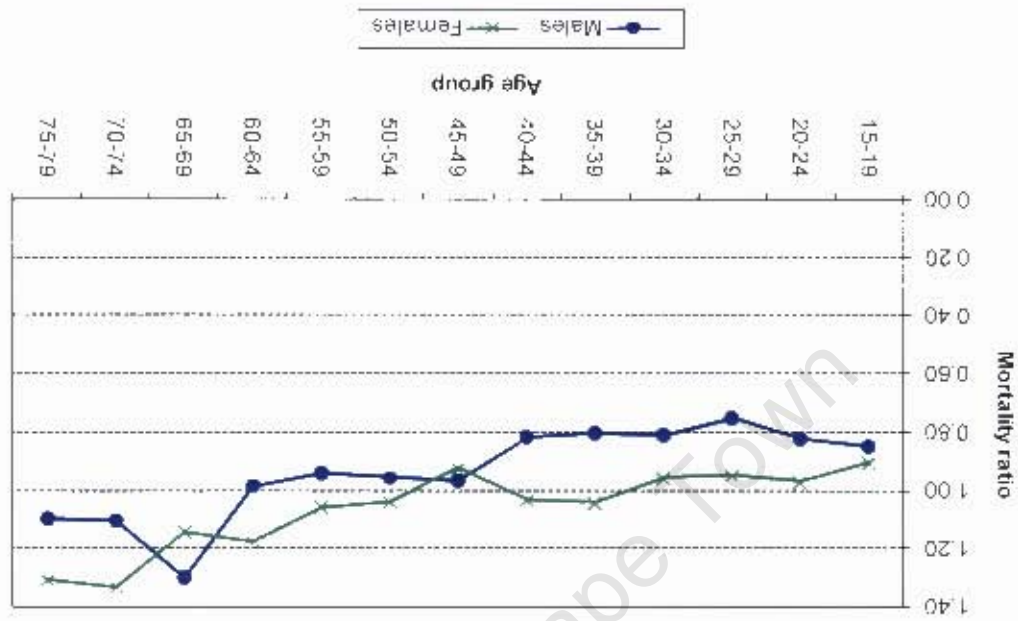


Figure 5.9 Ratio of mortality rates for households using electricity as main power source for lighting to those using some other power source for lighting

5.4.2.1 *Main power source used for lighting*
Mortality rates, by sex and age, were derived for households where electricity was the chief power source used for lighting and those where it was not. The ratio of the rates is shown in Figure 5.9.

Electricity is a safe and clean source of power and its use implies ownership of certain electrical appliances and connection to the electrical grid indicates some level of access to public services. So, one might expect higher mortality rates for households that use electricity for heating and lighting purposes.

5.4.2 Main power source for lighting and cooking

working ages, particularly for men.
Thus urban mortality rates appear to be somewhat lower than rural mortality rates in the comparisons above age 59 should be treated with caution.

of adult rural female deaths appears to carry through in the results that follow and for this reason adjustment factors applied to rural deaths might not state fully these deaths. The under reporting are proportionally more rural female deaths at the older ages to go unreported and so the not to have an adult left behind to report the death of the head of household. Given that there Female headed households at the older ages are more likely than male headed households

Importantly, the variable is likely to provide an appropriate measure of the extent of electrification, which is perhaps more a reflection of public sector service delivery and policy than any measure of socio-economic status.

5.4.2.2 Main power source used for cooking

Electricity when used for lighting purposes is a highly economical power source and the variable main power source for lighting might be used as a proxy for the electrification of households. However, electricity becomes expensive relative to other power sources when used for heating purposes. In addition, electricity as the primary power source for cooking also implies ownership of certain electrical goods (e.g. stove, oven). These are both an indication of higher income or wealth which might be expected to be associated with lighter mortality. The ratios of mortality rates are given in Figure 5.10.

Figure 5.10 Ratio of mortality rates for households using electricity as main power source for cooking to those using some other source



The ratios suggest that mortality rates for adults living in households where the main power source used for cooking purposes is electricity are lower than those that live in households that do not use electricity for cooking up to age 59. Once more, we can see a degree of random variation at the older age groups due to the small numbers at these age groups.

5.4.3 Household head variables

For our purposes, information on the level of income, employment status and educational attainment of the deceased prior to death would have been of interest, but unfortunately very little information was collected on deaths in the household.

As such information is not available, we consider instead variables for the household head, defined to be the chief decision-maker or main source of income in a particular household (Statistics SA, 2003b), as identified by respondents in the 10% sample, assuming that these variables provide some indication of the socio-economic class of the deceased. This is not a perfect solution, the very concept of household headship, as discussed by Budlender (2003), has significant shortcomings. This assumption is made in the absence of being able to make a better assumption.

5.4.3.1 Employment status of the household head

Employment status is an important socio-economic determinant with mortality differentials, in particular, between those living in households where the head is employed and those where he/she is not employed.

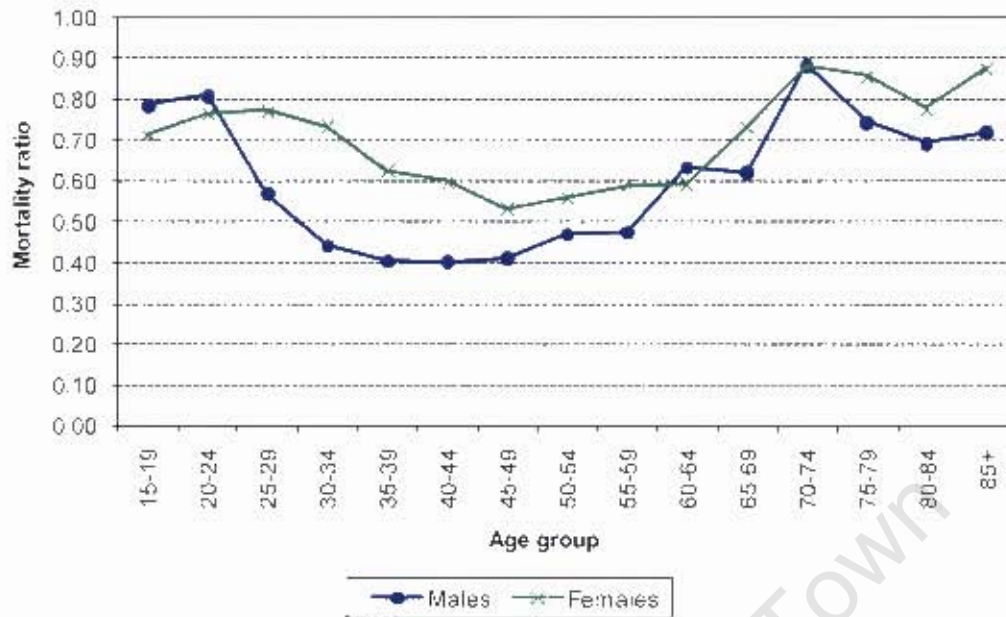
Mortality rates were derived both for lives with employed household head, as per the expanded definition, and for lives with household head not employed as at the census night. The expanded definition does not require that an individual has taken active steps to find employment (Statistics SA, 2003b).

As the household head is expected to be the main source of income for the household (Statistics SA, 2003b) we would expect the household head in almost all circumstances be a member of the economically active population, i.e. employed or, if not employed, willing to work.

For the purposes of this research, the reference category of lives with household head not employed (as opposed to unemployed) includes all those lives where the household head does not have formal employment either because that person is unemployed or is younger than 15 years or older than 65 year or is not formally employed for some other reason. Here, the argument is not that the deceased would have had the same employment status as the household head, but that living standards and hence mortality might differ by the employment status of the household head. Unfortunately this approach includes those in households where the household head is a retiree as being not employed and it is also possible that households where the household head is not employed may still have relatively high income levels. For one, those that have reached retirement age are eligible for a state old age pension, but there could also be other sources of non-employment income.

The ratio of the mortality rates appear in Figure 5.11.

Figure 5.11 Ratio of mortality rates of those with employed household head to that of others



The ratios indicate that the mortality of those living in households where the head is employed is significantly lighter (particularly over the working ages) than that of those living in households where the head is unemployed.

5.4.3.2 Completed education level of household head

The level of completed education might be associated with mortality as we expect both a household's earnings potential and standard of living to increase, all other things being equal, as years of completed formal education of the household head increases. As such, education might be a proxy of wealth and socio-economic status. Secondly, education might better enable people to access health information and services (Cutler, Deaton and Lleras-Muney, 2006).

Clearly the educational status of the household head, particularly in a South African context, has its limitations as a measure of educational attainment of all household members. It is certainly possible, for example, in a household where the household head has incomplete secondary education for there to be household members with tertiary qualifications. This is particularly relevant in South Africa given the Bantu education and might certainly lead to the variable not being significant. While educational attainment of the household head might not serve as a proxy for that of other household members, it might be a mortality risk factor for household members, due to possible wealth effects, and should be investigated. So, again an educational level is not assigned to household members based on the educational attainment of the household head. Here, we seek to use the educational attainment of the household head to determine different socio-economic categories.

The completed education variable in the 10% sample data set is very detailed. From this, we divided the households into three distinct sub-groups, namely those where the head of the household had “incomplete primary”, which consists of no formal schooling through to Grade 6 or its equivalent, “incomplete secondary” which is Grade 7 to Grade 11 or its equivalents and “completed secondary” or Grade 12 or equivalent and higher. Certificates and diplomas attained with less than Grade 12 were grouped in incomplete secondary. For example, those with highest educational attainment of Grade 7 would be classified as incomplete secondary.

The education variable refers to completed level of education and not current level of education, which could be different where the person is still enrolled in formal education. However, this is not problematic as it is likely that the household head is no longer enrolled in formal education and due to the grouping of years of education. Figure 5.12 and Figure 5.13 show mortality rate ratios for both males and females by completed level of education of the household head.

Figure 5.12 Ratio of mortality rates by completed education level of household head to those with incomplete primary: males

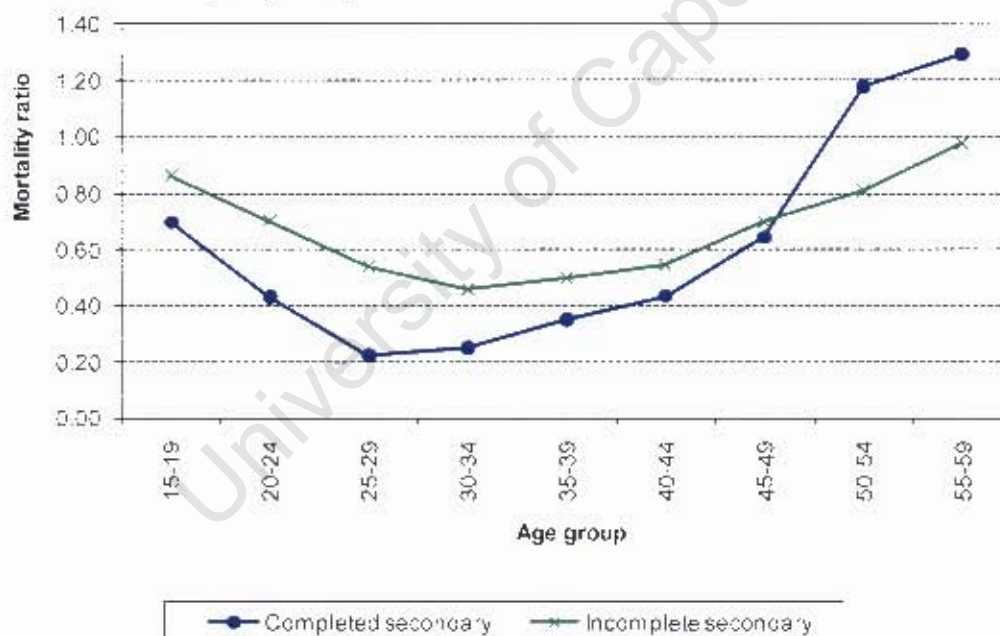
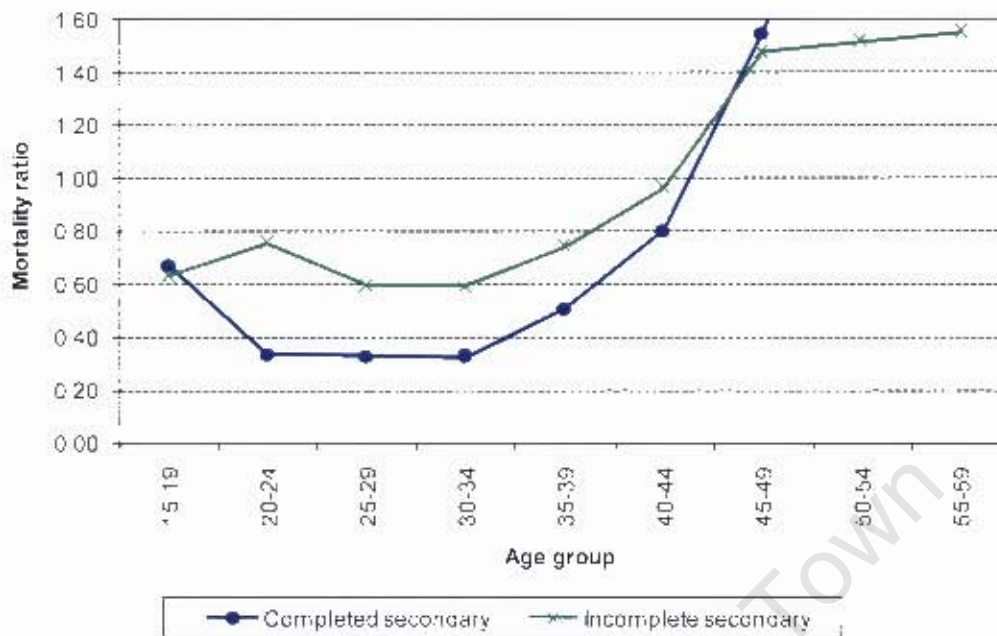


Figure 5.13 Ratio of mortality rates by completed education level of household head to those with incomplete primary: females



The mortality ratios for both males and females were truncated at age 60 years since the numbers become too small to produce useful results. Nevertheless, the figures above indicate that mortality rates for males and females are significantly different by completed education level of the household head, though the mortality differential is greatest for males and that completed secondary mortality is lighter than incomplete secondary.

5.4.3.3 Income level of household head

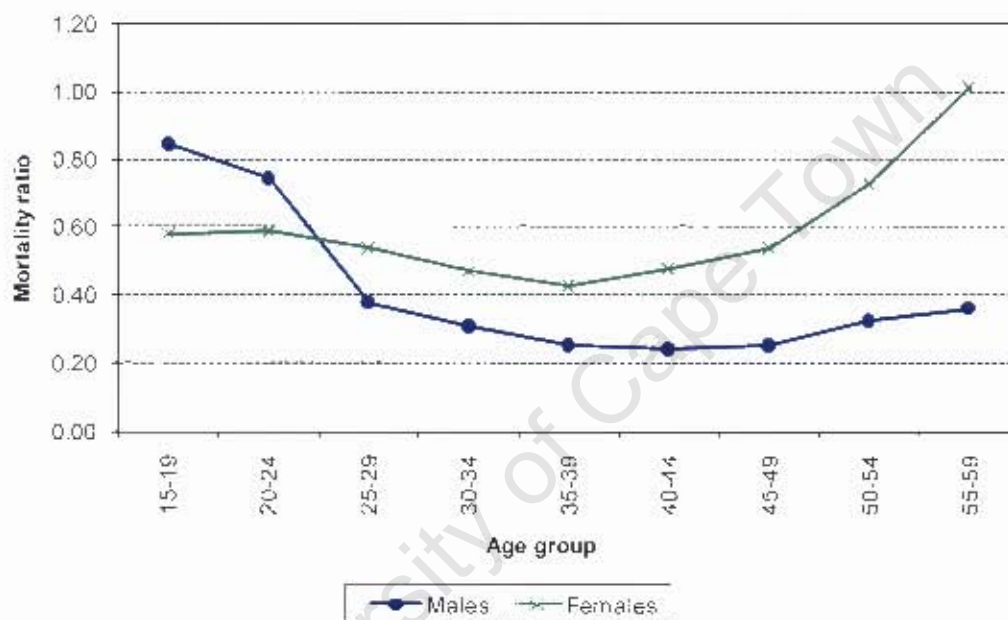
Mortality rates can be expected to differ by income levels. Socio-economic factors such as an increased standard of living are typically thought to have led to lighter mortality in the early 1900s (Gaspari and Woolf, 1985). Income levels are a good indicator of standard of living (*Ibid*).

However, the income variable in the 10% sample data set is not without its drawbacks. The reference person is required to provide information on the monthly or annual income for each household member by selecting the salary band, as determined by Statistics SA, that is most appropriate for each household member. Of course, the respondent might not know the precise Rand amounts or might be unwilling to divulge such information. In most cases, we expect the information provided by the respondent to be an estimate. These estimates might differ substantially from the actual incomes. As such, an incorrect income band might be selected by the respondent. Statistics SA provides a note cautioning users of the 10% sample when analysing the income variable (Statistics SA, 2003b) and thus the variable, at best, should only be used as a rough indicator of standard of living.

For the purposes of this analysis, we sub-divided adult Africans into two groups, namely those with household head with income of zero to R9 600 per annum and those with household head with income in excess of R9 600 per annum. It is reasonable to expect that in most cases the individual with the largest income in a household to be the household head, in line with the definition of the household head (Statistics SA, 2003b).

Mortality ratios for both males and females, by income level of the household head, are given in Figure 5.14.

Figure 5.14 Ratio of mortality rates of those living in households where the head earns more than R9 600 per annum to those of other households



Again, mortality rates are truncated due to the low numbers at the older ages. Nevertheless, it is clear that mortality rates are significantly lighter for both males and females over 15-59 years living in households where the head earns in excess of R9 600 per annum than those where the household head earns less. The mortality differential between income levels is greatest for males, with the greatest estimated difference for males aged 40-44 years with a mortality ratio of only 24 per cent. This analysis suggests that the income level of the household head is a significant risk factor with respect to adult mortality.

5.4.4 Average household member income

An investigation of mortality rates for groups of lives developed using average household member income may be seen as an extension of the analysis based on household head income described above.

While there exists strong evidence that the level of mortality is associated with income of the household head, i.e. that mortality rates are lighter with higher income, analysis of income of the head of household income alone does not capture total household income nor the number of household members. All things being equal and ignoring possible economies of scale, we expect large households to require larger household incomes than smaller households so as for all household members from each household to be equally well off. Using income level of the household head alone does not reflect total household incomes fully.

In order to derive estimates of average income per household member this research follows the approach suggested by Statistics SA (2003b). As individual incomes are given in terms of income bands in the 10% sample, one needs to apply an appropriate income point estimate to each band. Fixed income values per band are provided by Statistics SA (*ibid*). These values are considered to be reasonable and are used to develop estimates of income earned by each person. Individual incomes are then summed across all members within a household and the total is then divided by the number of household members. This is done for each household. This approach and the subsequent analysis implicitly assumes that total household income is distributed equally across all household members, in the absence of anything better. In reality, one would not expect household income to be shared equally amongst all household members. Indeed, income distribution among household members might be associated with the relative age of each member. For example, one might expect a larger share of total household income to go to both very young and very old members where health costs are relatively high. By assuming that each member gets an equal share, this approach does not take into account the age composition of the household, but nevertheless gives a rough measure of relative socio-economic status.

Households are then divided into three groups depending on the level of the average household income per member. These are: households with average income per member of less than R3 000 per annum, households with average income per member of R3 000 to R6 999 per annum and households with average income per member in excess of R7 000 per annum. Mortality rate ratios, with reference to mortality in households with average income per member of less than R3 000 per annum, for both males and females are given in Figure 5.15 and Figure 5.16.

Figure 5.15 Ratio of mortality rates of those in households earning above R3 000 p.a. to those of those in other households: males

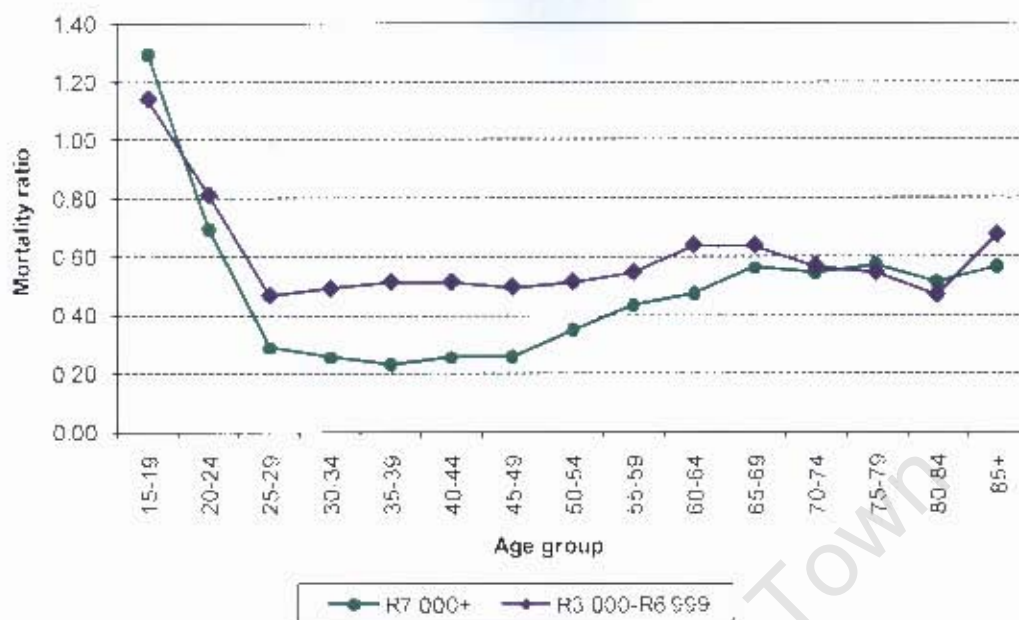
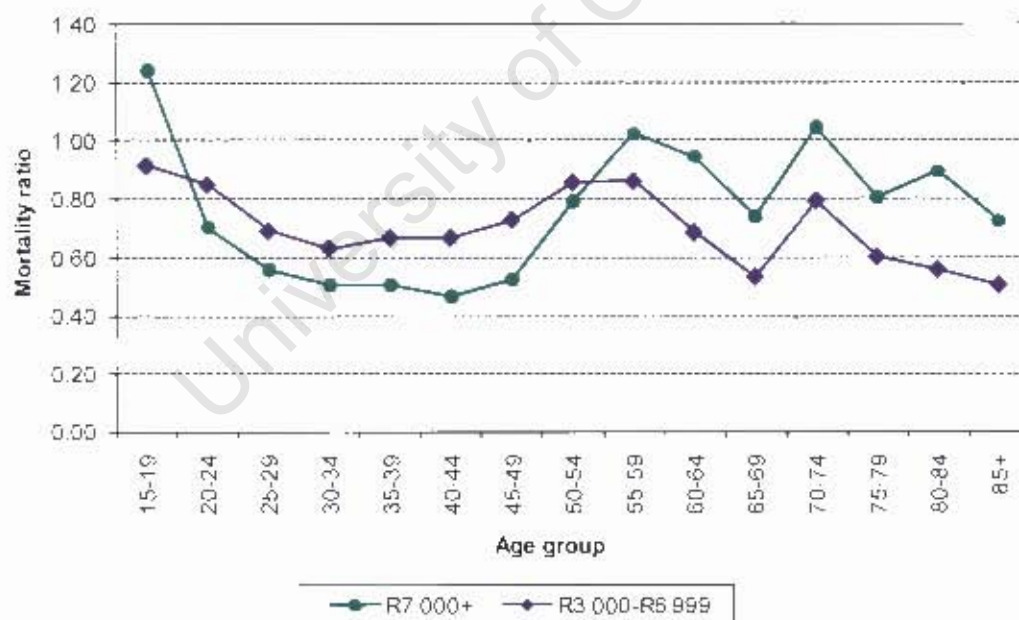


Figure 5.16 Ratio of mortality rates of those in households earning above R3 000 p.a. to those of those in other households: females



Interestingly, the rates for those in the higher income group are higher for age group 15-19 years. This might be due to either random fluctuation due to the few deaths at these ages or to the higher risk of accidental death for the higher income groups. Large mortality differences exist for the different average member income groups for both males and females. Again, mortality differences are greatest for males. From the above, mortality rates for males aged 20 years and

older with average household member income in excess of R7 000 per annum are between 20 per cent and 60 per cent of mortality rates for those with average member income of less than R3 000 per annum. For females, there appears to be significant random variation at the older ages, so the derived rates for ages in excess of 54 years while included but are uncertain.

From the above, both the level of income of the head of household and average household income per capita yield similar results. In Chapter 6, we investigate whether both income level of the household head and average income explain mortality differences when considered together.

5.4.5 Piped water and household toilet type

Mortality might be expected to be linked to access to safe drinking water and a working sanitation infrastructure. Cutler, Deaton and Lleras-Muney (2006) argue that mortality improves in line with improvements in the water supply as exposure to water-borne diseases falls. Piped water implies some level of water quality, i.e. the variable is an indicator of the quality of drinking water, while for household toilet type, a flush toilet implies that the household has access to a working sewerage infrastructure. An analysis of adult mortality by both access to piped water and by household toilet type follows.

5.4.5.1 Access to piped water

Piped water in South Africa is safe to drink and is accessible where appropriate infrastructure exists.

Using the access to piped water variable in the 10⁰% sample, we derived three groups. These are: households with access to piped water by tap either in the household structure or in the yard, households with access by means of a community stand, and households without access to piped water. Mortality rate ratios for these groups with reference to households with no access to piped water were derived for both males and females and are plotted in Figure 5.17 and Figure 5.18.

Figure 5.17 Ratio of mortality rates of those with access to piped water to those of those without access: males

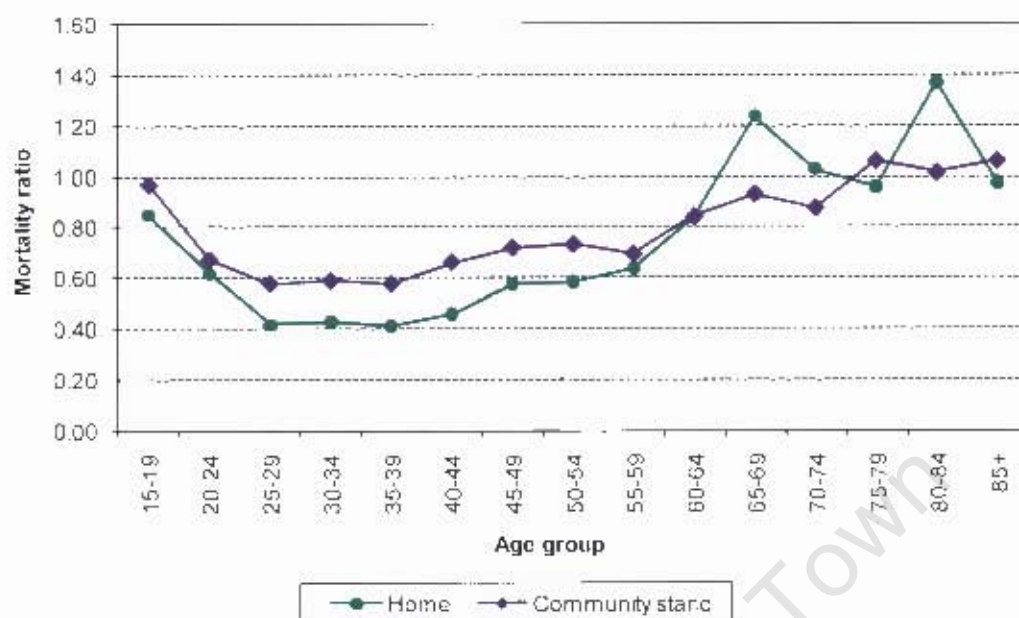


Figure 5.18 Ratio of mortality rates of those with access to piped water to those of those without access: females



For both males and females, mortality was lowest for those living in households with access to piped water by means of a tap on the household property over the age range 15-49 years. Mortality differentials once more appear to be greater for males than females.

Mortality differences also appear to exist between households that have access to piped water via a tap on the property to those that access piped water by means of a community stand. Access to piped water might be associated with urban-rural residence, or level of income.

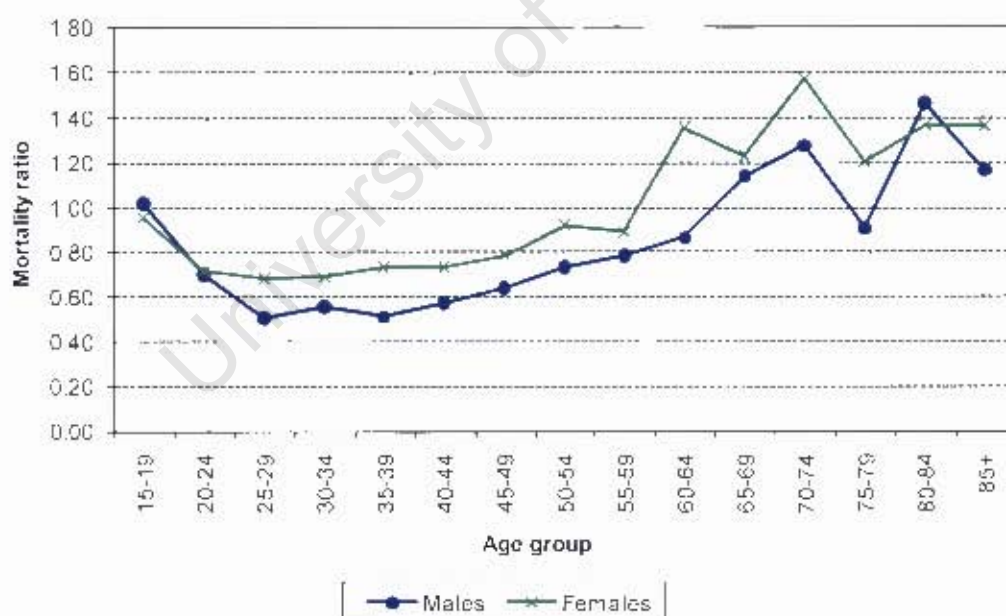
From the analysis above, the way in which households access drinking water impacts on adult mortality, with those that have access to piped water having lighter mortality than those without access to piped water.

5.4.5.2 Household toilet type

One would expect mortality for households with a flush toilet to be lighter than for those households with some other type of toilet, all other things equal, as poor waste management will result in unhealthy living conditions.

In order to investigate whether this is indeed true, adult mortality rates were estimated for adults from households with a flush toilet connected to either the sewerage system or septic tank and for those from households where there is a pit latrine, bucket or no toilet, using the toilet-type variable in the 10% sample data set. Mortality ratios for adults, by sex and type of toilet, follow in Figure 5.19.

Figure 5.19 Ratio of mortality rates of those in houses with a modern toilet to those without



Mortality differentials exist for both males and females with mortality being lighter for those households with access to a flush toilet. The mortality differential is largest for adult males aged 25-29 years where mortality for lives from households with a flush toilet is 50 per cent of

mortality for lives from households without a flush toilet. The results become unreliable for both males and females beyond 59 years owing to small numbers involved.

5.4.6 Wealth

As the income variable in the 10% sample is not without its drawbacks, a household asset index was developed as a proxy for wealth. This index was developed using principal components analysis, following the approach outlined by Filmer and Pritchett (1998), in which a weighted linear combination of the asset variables considered is generated (StataCorp, 2007).

Principal components analysis allows for a reduction in the number of variables, provided these variables are correlated, by developing one or more linear combinations of the correlated variables (StataCorp, 2007). The correlation matrix for household assets is given in Table 5.1.

Table 5.1 Correlation matrix for variables to be used in asset index

	Own radio	Own television	Own computer	Own refrigerator	Own telephone	Own cellphone
Own radio	1.000	0.379	0.164	0.361	0.254	0.253
Own television	0.379	1.000	0.258	0.626	0.434	0.374
Own computer	0.164	0.258	1.000	0.285	0.419	0.367
Own refrigerator	0.361	0.626	0.285	1.000	0.495	0.396
Own telephone	0.254	0.434	0.419	0.495	1.000	0.310
Own cellphone	0.253	0.374	0.367	0.396	0.310	1.000

Source: Census 2001 10% sample data set

Clearly access to many of the household goods in the table above can be expected to be highly correlated with whether the households have electricity as their main power source for lighting. As such, the development of an asset index to rank households in order to develop different socio-economic categories for the purpose of investigating possible mortality differences might be of little benefit due both to the fact that electricity as the main power source for lighting can be expected to be highly correlated to the assets and that electricity as the main power source for lighting was considered earlier in this chapter.

The first principal component was then calculated for each household in the 10% sample using the statistical software programme *STAT 10* (StataCorp, 2007). The index is linear and includes the mean of each of the individual asset variables. Assets used were re-coded as follows: one if the household had the item as at the census date and zero if not. By construction, the mean of the index is zero.

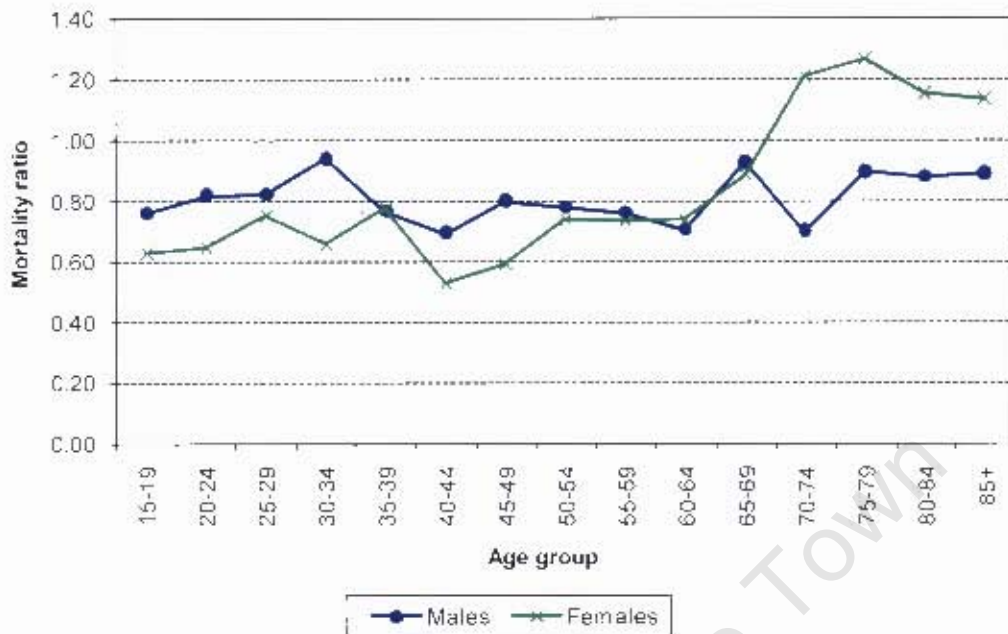
Households were divided into quartiles using the index. The wealthiest quartile for all households was identified. We then used the index to rank households based on household assets. The component loadings developed are given in Table 5.2 that follows.

Table 5.2 Component loadings and summary statistics for variables used in developing the first principal component, census 2001 10% sample

	Component loading	Mean	Standard deviation
Own radio	0.333	0.727	0.445
Own television	0.461	0.537	0.499
Own computer	0.344	0.086	0.280
Own refrigerator	0.476	0.512	0.500
Own telephone	0.428	0.245	0.430
Own cellphone	0.385	0.324	0.468

A check of the first principal component was undertaken by considering the proportion of households with those assets for each of the quartiles. As expected, the highest proportion of households in the 4th quartile had each asset. The asset index appears to be reasonable. Mortality rates were then estimated for both households in the highest quartile and households not in the highest quartile, by sex and age group.

Figure 5.20 Ratio of mortality rates of the wealthiest households to those of those in the rest of the households

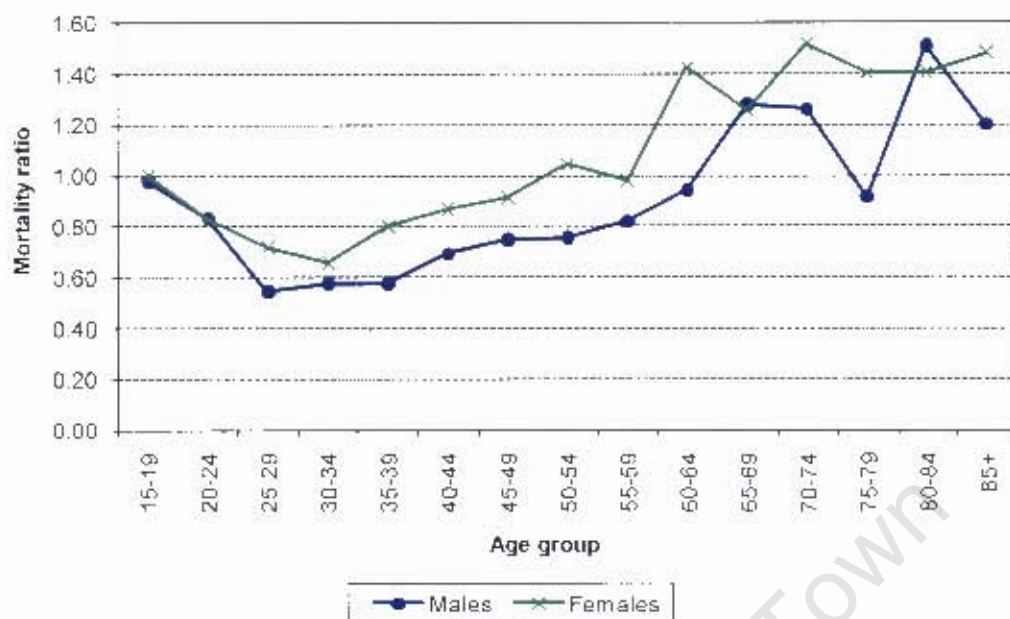


Investigation shows that mortality experience for males and females by household assets is not markedly different by household quartile for quartiles one, two and three. However, the mortality of males in the wealthiest quartile is about 80 per cent of that of the rest while that of females is relatively even lower below age 50 (Figure 5.20).

5.4.7 Household refuse collection

Household refuse collection method is used as a proxy for the extent of municipal and other services that might be available to a household, which would include access to health care facilities. One might expect medical and health facilities to be more accessible to people in households with formal refuse collection than those that do not. Mortality ratios of adult Africans for households with formal refuse removal to those without formal collection services are plotted in Figure 5.21.

Figure 5.21 Ratio of mortality rates of those with refuse collection to those of those without



There appears to be a mortality differential for both males and females for the working ages. Nevertheless, the ratios suggest that adults living in households with formal refuse collection have lighter mortality than their counterparts in households without formal refuse collection and hence that mortality does differ by access to public services.

5.5 Conclusion

Comparison of the number of deaths by sex and age from the vital registration system corrected for under-reporting with those reported by households in the census allows one to estimate adjustment factors that can be used to adjust deaths reported by households in various sub-groups, assuming that the adjustment factors derived at a national level are applicable to these sub-groups, to produce more reliable estimates of mortality rates than if the deaths reported by households had been used directly.

For the provinces, both KZN and EC have high mortality while the WC, GT and LP have low mortality, for both adult African males and females. The relatively high rates at the older ages for the WC suggest that mortality for the WC, at these ages, might be overstated. Mortality for adult males is greater than that for their female counterparts, all other things being equal.

Sex and age-specific mortality rates are produced by the following divisions: various household characteristics such as income level of the head of the household; employment status of the head of the household; and educational status of the head of the household; access to services; main power source used for cooking; toilet facilities; asset level of the household and

access to piped water. These rates are analysed univariately and mortality differences were found in all cases with the exception of whether or not the household used electricity for lighting. We see that mortality differentials are greatest for employment status, income level and completed education level of the household head and for mean income per household member.

We have only considered single variables in isolation, as random variation becomes significant, particularly at the older ages, due to sub-division of the data. It is difficult to derive robust mortality estimates by multiple variables simultaneously, making it difficult to identify possible differentials. The multivariate analyses in Chapter 6 are undertaken in order to identify which variables remain significant when one controls for other variables.

University of Cape Town

6.1 Using multivariate techniques

Socio-economic and other factors thought to be associated with differences in mortality were investigated in Chapter 5 where age- and sex- specific adult mortality rates for different sub-groups, adjusted for completeness, were derived and then compared to each other.

A multivariate approach allows us to investigate the impact of different socio-economic factors on mortality simultaneously. By considering different socio-economic factors simultaneously, we are also able to investigate whether a factor that appears significant when considered in isolation remains significant when we control for other factors.

6.2 Description of the approach

Initially, Poisson regression was applied to investigate the effect of specified socio-economic and other factors, identified in Chapter 5, on adult mortality rates. The response variable is an estimate of the central mortality rate, i.e. adjusted deaths divided by exposure. In the absence of specific weights for the mortality data, adult deaths are weighted using household weights then corrected for under-reporting using the adjustment factors.

However, Poisson regression assumes that the dependent variable is a non-negative count variable that has a Poisson distribution (StataCorp, 2007). But, the Poisson distribution has equal mean and variance and, where the variance of the dependent variable is greater than the mean, over-dispersion exists (*ibid*). Goodness of fit tests (using *STAT.10* statistical software package) indicated that the Poisson distribution was inappropriate and the Negative Binomial model, which is a generalisation of the Poisson regression and allows for extra variation (StataCorp, 2007), was applied instead.

Adult mortality rates are compared by calculating incidence-rate ratios (IRR). Effectively, these rate ratios give the mortality estimate for a particular group compared to the reference level. For example, an incidence-rate ratio of, say, 1.52 for the variable *sex*, where *sex* is a dummy variable that equals one where male and zero otherwise, suggests that adult male mortality is 1.52 times that of adult females, all other things being equal.

Further model assumptions and variables constructed for the purposes of these multivariate investigations are described in sub-sections that follow.

The likelihood ratio test, which requires that models be nested (StataCorp, 2007), and Aikaike's Information Criterion (AIC) are used to identify those models that are able to explain the most variation. Models are evaluated by comparing the AIC score of each model. The best fit

model is the one with the lowest AIC score. A likelihood ratio test is then used to confirm the model choice based on the AIC score.

6.2.1 Model assumptions

In developing these multivariate models, we assume that people alive on the census night 9/10 October 2001 and those that died in the 12 months prior to the census night contribute fully to the exposure for the 12 months prior to the census night. Using census data, we can track any in-movements, but not any out-movements. So, the above assumes that the in-migrants are more or less compensated by any out-migrants, i.e. net in-migration for the different sub-divisions is assumed to be negligible.

We also assume that the characteristics of adults and households, as reported at the census night, hold or are applicable for the prior 12 months. Lastly, care was taken to ensure that, for each level of sub-division, the number of deaths was non-zero, i.e. that actual mortality rates were positive, so as to improve model fit. While the inclusion of a significant independent variable would be expected to add explanatory power, an increase in the number of observations, all other things equal, might lead to a decrease in the proportion of variability explained by the model.

6.2.2 Description of variables for these multivariate analyses

Dummy variables which are one if a specific condition or set of conditions are true and zero otherwise, were developed for the independent variables both to aid interpretation of model outputs and to determine different discrete levels for socio-economic factors thought to be significant. For example, in order to improve interpretation of estimated incidence-rate ratios, using variable *sex* in the 10% sample data set, the dummy variable *sex* was developed and equals one if male and zero otherwise. Thus by construction, the reference level is female.

The construction of dummy variables might then be applied to categorical variables with more than two levels, say m levels. A reference level is chosen and $m-1$ dummy variables are created. For example, the Census 2001 the age variable *qp02* was sub-divided into five-year age bands by developing dummy variables. This was sub-divided into five-year age bands and dummy variables *age20_24*, *age25_29*, ..., *age55_59*, with reference level 15-19 years, were developed in order to analyse adult mortality by different age groups.

6.3 Findings

After allowing for age and sex, each province, with the exception of GT, is found to be significant at a one per cent level with the WC as the reference level. This means that mortality for each of the EC, NC, FS, KZN, NW, MP and LP differs significantly to mortality in the WC. This result supports our earlier findings. Each of the socio-economic and other factors identified

as being associated with different levels of mortality in Chapter 5, after allowing for age and sex, are also found to be significant at a one per cent level. For example, variable *flush_toilet* has an incidence rate ratio of 0.727 with p-value 0.000, i.e. the mortality of adults from households with a flush toilet is approximately 0.727 times that of adults from households without a flush toilet.

Similarly, the variables refuse collection and households where the chief source of power for cooking is electricity have incidence rate ratios of 0.779 and 0.709 respectively, with both significant at a one per cent level (p-value<0.000). The derived variable mean income per household member returned rate ratios of 0.634 and 0.488 for *mi3000_6999* and *mi7000+* respectively, also both at a one per cent level, i.e. households with mean annual household income of between R3 000 and R7 000 is approximately 0.634 times that of adults from assets where , all other things equal. Similarly, adults from households with average annual income in excess of R7 000 have 0.488 times the mortality of adults from the reference level.

Households with access to piped water by means of a tap either within the household structure or immediate household property have approximately 0.778 times the mortality of those without access to piped water.

This supports the findings of the earlier univariate analyses. However, when considered in combination, not all were significant. So, a model was fitted with age and sex and socio-economic factors were added to this model. Factors found to be significant were retained while those that were not were excluded.

After allowing for age and sex, we found enumeration area type to be significant at a one per cent level (p-value<0.01). Enumeration area type was considered first, after age and sex, as this is relatively easy to determine. We see that people living in urban areas have lighter mortality to those that live in rural areas, all other things being equal. The output suggests that those living in rural areas have mortality approximately 1.22 times the mortality of those living in urban areas. This might be expected given that those living in urban areas might be expected to have greater access to public service infrastructure and possible access to healthcare facilities.

Table 6.3 Negative binomial regression output for adult Africans using adjusted deaths, by sex, age, employment status and income level of the head and access to piped water, using Census 2001 10% sample data set

Negative binomial regression Number of obs = 160
 LR chi2(15) = 351.03
 Log likelihood = -2231.7097 Prob > chi2 = 0.000

deaths	IRR	Std. Err.	z	P>z	95% Confidence Interval	
sex	1.8141	0.1385	7.80	0.000	1.5620	2.1069
age20_24	3.1481	0.3797	9.51	0.000	2.4854	3.9876
age25_29	4.9855	0.6011	13.33	0.000	3.9363	6.3144
age30_34	6.4868	0.7837	15.48	0.000	5.1191	8.2198
age35_39	6.4428	0.7799	15.39	0.000	5.0820	8.1681
age40_44	6.6820	0.8080	15.71	0.000	5.2721	8.4689
age45_49	6.5275	0.7868	15.56	0.000	5.1540	8.2671
age50_54	7.9058	0.9513	17.18	0.000	6.2449	10.0085
age55_59	5.5836	0.7472	12.85	0.000	4.2953	7.2582
age60	17.4535	2.6425	18.89	0.000	12.9721	23.4831
employhhh	0.6794	0.0554	-4.74	0.000	0.5791	0.7971
incomehhh	0.7411	0.0400	-5.56	0.000	0.6667	0.8237
piped_water	0.8627	0.0464	-2.74	0.006	0.7763	0.9587
sex_employhhh	0.6245	0.0672	-4.38	0.000	0.5057	0.7711
age55_employhhh	2.9867	0.4198	7.78	0.000	2.2675	3.9341
exposure	(exposure)					
/lnalpha	-2.161	0.110			-2.376	-1.946
alpha	0.115	0.013			0.093	0.143

Likelihood-ratio test of alpha=0: $\chi^2(01) = 2.2e+07$ Prob>= $\chi^2 = 0.000$

From the output above, after allowing for age and sex, we see both the employment status and income level and access to piped water to be significant at a one per cent level. We see that, as before, people living in households where the head is employed and earning in excess of R9 600 p.a. will have lighter mortality than those where the head is unemployed and earning less than R9 600 p.a. In addition, with an IRR of 0.8627 we see that those that are able to access piped water from the household property have approximately 0.86 times the mortality of those that either access piped water via a community stand or who do not have access to piped water, all other things being equal.

Introducing interaction terms in the analysis above, the interaction between sex and employment status of the household head is significant at a one per cent level. The output indicates that a male in a household where the household head was employed has approximately 0.77 times the mortality of a female living in a household with a household head that is not employed, all other things being equal. Similarly, the output indicates that the mortality of a male with employed household head has approximately 1.13 times the mortality of a female with

employed household head, all other things being equal. However, should one not know the employment status of the household head, one would expect, from the output, the mortality of males to be approximately 1.81 times the mortality of females. So, the sex-employment status interaction reduces the mortality difference between males and females when one considers mortality of people from households where the household head is employed.

The age-employment status interaction for the open age group is also seen to be significant at a one per cent level. The output indicates that lives aged 55 years and older with employed household head have approximately three times the mortality of those aged 55 years and older with household head not employed, all other things being equal. This, at first, appears counter-intuitive. However, this might be due to the fact that one would expect lives in the open interval that are not employed to be in a higher socio-economic group to those that are employed as this would suggest some form of non-employment income, i.e. older lives that have reached retirement age that are able to access some form of non-employment income are less likely to need to work than those that do not have such income, all other things being equal. This interaction addresses partly one of the drawbacks of using employment status of the household head mentioned earlier in Chapter 5.

University of Cape Town

Table 6.4 Negative binomial regression output for adult Africans using adjusted deaths, by sex, age, employment status and income level of the head and access to piped water, using Census 2001 10% sample data set

Negative binomial regression Number of obs = 160
LR chi2(13) = 285.59
 Log likelihood = -2264.4266 Prob > chi2 = 0.000

deaths	IRR	Std. Err.	z	P>z	95% Confidence Interval	
sex	1.6938	0.1338	6.67	0.000	1.4509	1.9773
age20_24	3.1242	0.3689	9.65	0.000	2.4788	3.9378
age25_29	4.9489	0.5842	13.55	0.000	3.9266	6.2373
age30_34	6.4673	0.7650	15.78	0.000	5.1291	8.1546
age35_39	6.4647	0.7663	15.75	0.000	5.1245	8.1554
age40_44	6.6952	0.7926	16.06	0.000	5.3089	8.4436
age45_49	5.4506	0.7307	12.65	0.000	4.1911	7.0885
age50_54	6.4870	0.8884	13.65	0.000	4.9599	8.4843
age55_59	5.5985	0.7338	13.14	0.000	4.3303	7.2383
age60	17.3399	2.5715	19.24	0.000	12.9663	23.1889
incomehhh	0.7455	0.0394	-5.56	0.000	0.6722	0.8268
employhhh	0.6811	0.0541	-4.83	0.000	0.5828	0.7959
piped_water	0.8609	0.0454	-2.84	0.004	0.7764	0.9546
sex_employhhh	0.6214	0.0655	-4.52	0.000	0.5054	0.7639
age55_employhhh	2.9781	0.4098	7.93	0.000	2.2742	3.9000
sex_age45_54	1.4263	0.1881	2.69	0.007	1.1014	1.8471
exposure	(exposure)					
/lnalpha	-2.204	0.110			-2.419	-1.989
alpha	0.110	0.012			0.089	0.137

Likelihood-ratio test of alpha=0: $\chi^2(01) = 2.0e+07$ Prob>= $\chi^2 = 0.000$

Age-sex interaction was also tested for significance and was found to be significant at a one per cent level. The output indicates that a male aged 45 years has approximately 2.42 times the mortality of a female aged 45 years and approximately 7.77 times the mortality of a male aged 15 to 19 years, all other things being equal.

Further models were tested but no other socio-economic and other factors were significant when considered in combination with the above factors.

Mortality rates for adult Africans by age and sex, adjusted for completeness, are estimated for the 12 months prior to the 2001 census. These are very similar to those produced by Dorrington, Moultrie and Timæus (2004) who had access to the full census data set. The higher mortality at the older ages shown by Dorrington, Moultrie and Timæus (2004) is due to the rates of this research being ungraduated, while the higher mortality in the young adults is the result of better data used in this research. Thus this research provides a revision of mortality rates for adult Africans in the year prior to the 2001 census, against which future mortality estimates may be compared.

This research considers only adult mortality and complete life tables cannot be constructed in the absence of child mortality estimates. Should we wish to develop African life tables, we would need to consider a separate approach to estimating child mortality as we expect the level of under-reporting of child deaths to differ to that of adults. This research might also be extended to include an analysis of mortality of the other population groups not considered. It is of interest to understand whether the same socio-economic and other factors associated with different mortality levels for adult Africans apply for the different population groups. This would not be without difficulty though, given the failure to capture deaths of Whites in institutions and the smallness of death numbers, particularly for Indians (Dorrington, Moultrie and Timæus, 2004).

We derive mortality rates by province, applying the same adjustment factors developed to correct for the under-reporting of deaths for the country as a whole. We see that the mortality rates for the WC, GT and LP are low when compared to the country as a whole, although we are unable to fully explain the relatively high mortality at the older ages for the WC. This might be due to the completeness of reporting of household deaths in the census being a function of age, sex and province. We believe the rates at the older ages for the WC to be overstated. We also see that LP has relatively low mortality. Dorrington, Moultrie and Timæus (2004) also produce low mortality rates for LP, which they believe are anomalous. By general reasoning, it is difficult to argue that LP should have low mortality compared to the country as a whole. This result also suggests that the completeness of reporting of deaths might vary by province, all other things being equal.

As a check, in the multivariate analyses, provinces, after allowing for sex and age, with the exception of GT, are found to be significant at a one per cent level with the WC chosen as the reference level. This supports the earlier findings in Chapter 5.

The high mortality rates derived for adult Africans are due in no small part to HIV/AIDS. Unfortunately, we are unable to disaggregate the reported household deaths in the census into AIDS and non-AIDS deaths. That said, we see the impact of HIV in the distribution of reported deaths, provided by Statistics SA (personal correspondence) in calendar years 1997-2001 (see Chapter 4). In particular, we see the numbers of reported deaths over the age ranges 20-64 years for males and 15-59 years for females increase significantly over this period, while the numbers of deaths at the younger and older ages remain relatively stable. We also note the high mortality rates derived for KZN, which we expect given the high HIV prevalence in the province.

While we show that we are able to derive very similar mortality rates for the country as a whole using the 10% sample data set compared to the rates produced using the full census data set, we acknowledge that the rates for different socio-economic and other sub-groups experience greater random variation, due to the smallness of numbers. This is particularly noticeable at the older ages due to there being fewer deaths at these ages. In this respect, the full data set is preferable to the 10% sample in that we would expect mortality rates for different sub-groups derived using the full data set to experience less random fluctuation than those derived using the 10% sample.

We argue that the completeness of reporting of deaths is due, in part, to the possible disintegration of the household following a death in that household. With the disintegration of a household, remaining members either join existing households or form new households. Then, at the census, it is unlikely that the death will be reported as the household in which the death occurred no longer exists. Further, we believe that the disintegration of the household is more likely to occur for deaths at the older ages, i.e. for those people that are more likely to have contributed significantly to household income. It is important to understand more precisely the impact that household disintegration might have on the reporting of household deaths in the census. This, we believe, warrants further research.

Completeness of reporting of household deaths might also differ by socio-economic and other factors, while we have assumed in this research, in the absence of anything better, that the same adjustment factors may be used across the different socio-categories. These possible differences could also be explored in future research.

Mortality rates are produced for different socio-economic and other factors. These are: income level of the head; employment status of the head; completed education level of the head; access to services; main power source used for lighting and cooking; toilet facilities; mean income per household member; asset groups using a derived asset index and access to piped water. These

different socio-economic categories were investigated both on the basis that we believed that mortality differences might exist across these categories and due to data availability. Of course, the choice of variables that could be considered was limited by availability of data. In particular the use of household head characteristics has its problems. For one, these characteristics cannot be used as a proxy for the status other household members. For example, the educational attainment of the household head might be expected to be quite different from other household members. This is particularly problematic in South Africa given that Bantu education allowed only limited access to education for Africans in the past.

In addition households where the household head has reached retirement age will be included in the category of households where the household head is not employed. This may have implications for interpreting the results particularly where the household head has access to some form of post-retirement income, for example the state old age pension or some form of other post-retirement provision. That said, this research does not seek to use the household head variables as proxies of those of other household members, but rather to use these variables as an indicator of different socio-economic living conditions. From the univariate investigations we see that the largest mortality differentials exist for income level of the head, employment status of the household head, completed education level of the head and mean income per household member. By considering only the income level of the household head, one takes into account neither the incomes of other household members nor the number of household members. In order to address these issues, the analysis was extended to consider average household income per capita. As mentioned earlier, it is reasonable to expect the distribution of household income to be associated with the relative age of each member. However, this approach assumes that total household income is shared equally amongst all members and, as such, does not take into account the age composition of the household. The level of income of the head of the household and the average household income per capita yield similar results.

It was thought that the household assets from the 2001 census might be used to group households into different socio-economic categories. Due to the high correlation between some of the assets, a single index that used each of these assets was developed to rank broadly households. One was then able to compare households with a high index to those with a low index. To this extent, the index served its purpose. Of course, a single index that incorporates many of the other variables investigated in Chapters 5 and 6 might also be developed and the development of such an index is an avenue for possible future research.

The findings of the univariate analyses are also supported by the multivariate analyses. Using negative binomial regression techniques and, after allowing for sex and age, each of the

socio-economic and other factors thought to be mortality risk factors in the univariate analyses are found, individually, to be significant at a one per cent level. However, when considered in combination, and after allowing for age and sex, the best fit model consists of income level and employment status of the head and type of access to piped water. Possible interactions between the different socio-economic and other factors are considered and found to be significant. These are a sex-employment status, an age-employment status and an age-sex interaction. Interestingly, the IRR for the age55-employment status interaction of approximately 2.98 allows, in part, for one of the drawbacks of using employment status of the household head as an indicator of socio-economic level for older lives that have reached retirement age.

The best model is chosen by considering the AIC score of each model. When comparing two models, the model with the lowest AIC score is the best model, i.e. the one that explains the most variation. Model choice is then confirmed using a likelihood ratio test. The model with independent variables age, sex, employment status and income level of the head and access to piped water within the property is identified as the best model. More complex models were tested but no other socio-economic and other factors are significant when considered in combination with the above factors.

The model results mean that, after allowing for age and sex, mortality is dependent on employment status and income level of the head and type of access to piped water. In addition, it means that additional mortality differences may be explained by, say, the addition of employment status of the head, after already considering income level and type of access to piped water. We see that the addition of these variables contributes further information to mortality. More importantly, the multivariate results support the earlier findings of the univariate analyses, which indicate that mortality is a function of wealth. By general reasoning, we expect mortality rates to be lower for higher socio-economic and income groups, all other things being equal, and this research supports this argument.

The implications of this are as follows. An increase in (real) incomes and employment levels, together with improved service delivery, might be expected to result in lower adult mortality, i.e. successful measures to combat poverty might lead to a reduction in mortality rates.

Lastly, we might want to repeat this research on more recent data to derive mortality rates in more recent periods. A similar approach for estimating adult mortality using Community Survey 2007 data can be applied.

8 REFERENCES

- Anderson, B. A., and H.E. Phillips. 2006. *Adult mortality (age 15-64) based on death notification data in South Africa: 1997-2004*. Report No. 03-09-05. Pretoria: Statistics South Africa.
- Bah, S. 1998a. *A Critical, Analytical and Historical Review of South African Published Life Tables*. Discussion Paper no. 98-2. Revised version of paper presented at the Annual Conference of the South African Statistical Association, Dikololo, November 1997.
- Bah, S. 1998b. *The Making, Unmaking and Remaking of a National but Stratified Vital Statistics System in the Republic of South Africa*. Discussion Paper No. 98-3. Vital Statistics, Central Statistical Service, Pretoria.
- Bah, S. 2000. *A supplementary review of existing HIV/AIDS models with the view of adopting/adapting one or more models for national and provincial population projections in South Africa*. Discussion Paper No. 00-6. Statistics South Africa, Pretoria.
- Bah, S. 2005. "Technical Appraisal of Official South African Life Tables," pp. 181-217 in Zuberi, T., A. Sibanda and E. Udjo (ed.), *The Demography of South Africa*. London: M.E. Sharpe.
- Bennett, N. and S. Horiuchi. 1981. "Estimating the completeness of death registration in a closed population," *Population Index* **47** (2): 207-21.
- Bennett, N. and S. Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries," *Demography* **21** (2): 217-233.
- Booyesen, F. le R. 2003. *Chronic and Transitory poverty in the face of HIV/AIDS-related morbidity and mortality: Evidence from South Africa*. Paper presented at international conference on "Staying Poor: Chronic Poverty and Development Policy", University of Manchester, 7-9 April 2003, United Kingdom.
- Bourne, D. "Sources of South African mortality data, 1910-1992," *South African Medical Journal* 1995: **85** Issue 10
- Bradshaw, D., P. Groenewald, R. Laubscher, N. Nannan, B. Nojilana, R. Norman, D. Pieterse and M. Schneider. 2003. *Initial Burden of Disease estimates for South Africa, 2000*. Cape Town: Medical Research Council. ISBN: 1-919809-64-3.
- Bradshaw, D., R. Laubscher, R.E. Dorrington, D. Bourne and I.M. Timæus. 2004. "Unabated rise in number of adult deaths in South Africa," *South African Medical Journal* **94**: 278-9.

- Brass, W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill, NC: University of North Carolina, Chapel Hill in United Nations, *Manual X Indirect Techniques for Demographic Estimation*. Department of Economic and Social Affairs, Population Studies, No. 81. New York: United Nations.
- Brass, W. and K. Hill. 1973. "Estimating adult mortality from orphanhood," *Proceedings of the International Population Conference, Liege, 1973*. Vol. 3, pp. 111-123 in United Nations, *Manual X Indirect Techniques for Demographic Estimation*. Department of Economic and Social Affairs, Population Studies, No. 81. New York: United Nations.
- Brass, W. and E. Bamgboye. 1981. *The Time Location of Reports of Survivorship: Estimates for Maternal and Paternal Orphanhood and Ever-Widowed*. Centre for Population Studies Working Paper, No.81-1. London: London School of Hygiene and Tropical Medicine in United Nations, *Manual X Indirect Techniques for Demographic Estimation*. Department of Economic and Social Affairs, Population Studies, No. 81. New York: United Nations.
- Bogue, D. J. 1965. "Pros and Cons of 'Self-Enumeration'," *Demography*, Vol. 2: 600-626.
- Budlender, D. 2003. *The debate about household headship*. Discussion Document. Pretoria: Statistics South Africa.
- Chackiel, J. and H. Orellana. 1985. "Adult female mortality trends from retrospective questions about maternal orphanhood included in censuses and surveys," *International Population Conference, Florence 1985*, Vol. 4: 39-52 in Timæus, I. M., "Measurement of Adult Mortality in Less Developed Countries: A Comparative Review," *Population Index*, Vol. 57, No. 4. (Winter, 1991): 552-568.
- Cutler, D. M., A. S. Deaton and A. Ileras-Muney. 2006. *The Determinants of Mortality*. NBER Working Paper No. 11963 JEL No. I1, J1, O1, N3.
- Dorrington, R.F., D. Bradshaw and T. Wegner. 1999. *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*. University of Cape Town.
- Dorrington, R.E., D. Bourne, D. Bradshaw, R. Laubscher and I.M. Timæus. 2001. *The impact of HIV/AIDS on adult mortality in South Africa*. Technical Report. Cape Town: Burden of Disease Unit, Medical Research Council.
- Dorrington, R. E., T. A. Moultrie and I. M. Timæus. 2004. *Estimation of mortality using the South African Census 2001 data*. CARE Monograph, 2004.

- Drever, F. and M. Whitehead (ed.). 1997. *Health Inequalities Decennial supplement*. Office for National Statistics. Series DS No. 15. London: The Stationery Office.
- Filmer, D. and L. Pritchett. 1998. *Estimating wealth effects without expenditure data – or tears: An application to educational enrolments in States of India*. World Bank Policy Research Working Paper No. 1994. Washington, DC: World Bank.
- Gaspari, K. C. and A. G. Woolf. 1985. “Income, Public Works and Mortality in Early Twentieth-Century American Cities,” *Journal of Economic History*, Vol. **XLV**, No. 2.
- Gakidou, E., M. Hogan and A. D. Lopez. 2004. “Adult Mortality: time for a reappraisal,” *International Journal of Epidemiology*, **33**: 710-717.
- Gjonca, E. 2003. *Education and Social Class differences in survival at old age in England Wales: Results from the 1% Longitudinal Study of England and Wales*. Max Planck Institute for Demographic Research.
- Groenewald, P., D. Bradshaw, J. Daniels, R. Matzopoulos, D. Bourne, N. Shaikh, D. Blease, N. Zinyakira and N.T. Naledi. 2007. *Cause of death and premature mortality in Cape Town, 2001-2004*. Cape Town: Medical Research Council.
- Hill, K and J. Trussell. 1977. “Further developments in indirect mortality estimation,” *Population Studies*, **31**: pp. 313-34.
- Hill, K. 1977. “Estimating Adult mortality levels from data on widowhood,” *Population Studies*, **31** (1): 75-84.
- Hill, K. 1987. “Estimating Census and Death Registration Completeness,” *Asian and Pacific Population Forum*, Vol. **1**, No. 3: 8-13, 23-24.
- Hill, K. 2001. *Methods for Measuring Adult Mortality in Developing Countries: A Comparative Review*. Paper presented to the International Population Conference, Salvador, Brazil.
- Hill, K. 2003. *Adult mortality in developing world; what we know and how we know it*. Workshop on HIV/AIDS and Adult Mortality in Developing Countries, 8-13 September, Population Division, Department of Economic and Social Affairs, United Nations Secretariat, New York. UN/POP/MORT/2003/1.
- Hill, K., Y. Choi and I.M. Timæus. 2005. “Unconventional approaches to mortality estimation,” *Demographic Research*, Max Planck Institute for Demographic Research, Vol. **13**, Article 12: 281-300.

- Hosegood, V., A-M. Vanneste and I. M. Timæus. 2004. "Levels and causes of adult mortality in rural South Africa: the impact of AIDS," *AIDS* **18**: 663-671.
- Hosegood, V., N. McGrath, K. Herbst and I.M. Timæus. 2004. "The impact of adult mortality on household dissolution and migration in rural South Africa," *AIDS* **18**: 1585-1590.
- Huisman, M., A.F. Kunst, M. Bopp, J-K. Borgan, C. Borrell, G. Costa, P. Deboosere, S. Gadeyne, M. Glickman, C. Marinacci, C. Minder, F. Regidor, T. Valkonen and J.P. Mackenbach. 2005. "Education inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations," *Lancet* **365**:493-500.
- Igumbor, E.U., D. Bradshaw and R. Laubscher. 2003. *Mortality Profile from Registered Deaths for Limpopo Province, South Africa 1997-2001*. Cape Town: South African Medical Research Council, University of Venda.
- Kahn, K., S.M. Tollman, M. Garenne and J. S.S. Gear. 1999. "Who dies from what? Determining cause of death in South Africa's rural north-east," *Tropical Medicine and International Health* Vol. **4**, No. 6: 433-441.
- Kahn, K., M.L. Garenne, M.A. Collinson and S.M. Tollman. 2007. "Mortality trends in a new South Africa: Hard to make a fresh start," *Scandinavian Journal of Public Health*. **35** (Suppl 69): 26-34
- Khalfani, A.K., T. Zuberi, S. Bah and P.J. Lehohla. 2005. "Population Statistics," pp. 90-119 in Zuberi, T., A. Sibanda and E. Udjo (ed.), *The Demography of South Africa*. London: M.F. Sharpe.
- Lopez, A.D., O.B. Admad, M. Guillot, B.D. Ferguson, J.A. Solomon, C.J.L. Murray and K.H. Hill. 2002. *World Mortality in 2000: Life Tables for 191 Countries*. Geneva: World Health Organisation.
- Martikainen, P., T. Valkonen and T. Martelin. 2001. "Change in male and female life expectancy by social class: decomposition by age and cause of death in Finland 1971-95," *Journal of Epidemiology and Community Health*, **55**: 494-499.
- Oni, S.A., C.L. Obi., A. Okorie, D. Thabede and A. Jordan. 2002. "The Economic Impact of HIV/AIDS on Rural Households in Limpopo Province," *S.AJE* **70** (7).
- Preston, S., A. J. Coale, J. Trussell and M. Weinstein. 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable," *Population Studies*, Vol. **46**, No. 2: 179-202.

- Sadie, J.L. 1973. *Projeksies van die Suid-Afrikaanse bevolking, 1970-2020*. Johannesburg, Nywerheid-Ontwikkelingskorporasie van Suid-Afrika Beperk.
- Smith, M. K. 2002. "Gender, Poverty and Intergenerational Vulnerability to HIV/AIDS," *Gender and Development*, Vol. **10**, No. 3: 63-70.
- StataCorp, 2007. *Stata Base Reference Manual, Vol. 3, Release 10*. College Station, Texas, StataCorp LP.
- Statistics South Africa. 1998. *Census 1996 10% Sample of records*. Pretoria: Statistics South Africa.
- Statistics South Africa. 2000a. *South African Life Tables, 1985-1994 and 1996*. Report No. 02-04-04 (1985-1994 & 1996). Pretoria: Statistics South Africa.
- Statistics South Africa. 2000b. *Recorded deaths, 1996*. Report No. 03-09-01 (1996). Pretoria: Statistics South Africa.
- Statistics South Africa. 2002. *Causes of Death in South Africa, 1997-2001*. Statistical release P0309.2. Pretoria: Statistics South Africa.
- Statistics South Africa. 2003a. *Census 2001: Concepts and Definitions*. Pretoria: Statistics South Africa.
- Statistics South Africa. 2003b. *Census 2001: Metadata*. Pretoria: Statistics South Africa.
- Statistics South Africa. 2005. *Mortality and causes of death in South Africa, 1997-2003: Findings from death notification*. Statistical release P0309.3. Pretoria: Statistics South Africa.
- Timæus, I. M. 1991. "Measurement of Adult Mortality in Less Developed Countries: A Comparative Review," *Population Index*, Vol. **57**, No. 4: 552-568.
- Timæus, I. M. and M. Jasseh. 2004. "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys," *Demography*, Vol. **41**, No. 4: 757-772
- Tollman, S. M., K. Kahn, M. Garenne and J. S. S. Gear. 1999. "Reversal in mortality trends: evidence from the Agincourt field site, South Africa, 1992-1995," *AIDS* **13**: 1091-1097.
- Udjo, E.1997. *Fertility and mortality trends in South Africa: the evidence from the 1995 October household survey, and implications on population projections*. Pretoria: Central Statistical Service.
- Udjo, E. 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 1999.

- Udjo, E. 2005. "An Examination of Recent Census and Survey Data on Mortality within the Context of HIV/AIDS," pp. 90-119 in Zuberi, T., A. Sibanda and E. Udjo (ed.), *The Demography of South Africa*. London: M.E. Sharpe.
- United Nations. 1983. *Manual X Indirect Techniques for Demographic Estimation*. Department of Economic and Social Affairs, Population Studies, No. 81. New York: United Nations.
- United Nations. 2002. *Methods for Estimating Adult Mortality*. Department of Economic and Social Affairs, Population Division. ESA/P/WP.175. New York: United Nations.
- United Nations. 2003. *World Population Prospects: The 2002 Revision*. Department of Economic and Social Affairs, Population Division. New York: United Nations.
- Urassa, M., J. T. Boerma, R. Isingo, J. Ngalula, J. Ng'weshemi, G. Mwaluko and B. Zaba. 2001. "The impact of HIV/AIDS on mortality and household mobility in rural Tanzania," *AIDS* **15**: 2017-2023
- van Eeden, I.J. and J.L. van Tonder. 1975. *Abridged Life Tables for All the Populations Groups in South Africa*. Report No. S-34. Pretoria: Human Sciences Research Council.
- Yamano, T., T.S. Jayne and M. McNeil. 2002. *Measuring the impacts the prime-age adult death on rural households in Kenya*. IAAE Conference, October 2002.

9 APPENDIX 1: DERIVED ADULT AFRICAN MORTALITY RATES

Figure 9.1 Estimated adult African male central mortality rates for the 12 months prior to 9/10 October 2001, by age and province

Males	WC	EC	NC	FS	KZN	NW	GP	MP	LP
15-19	0.0033	0.0027	0.0014	0.0018	0.0030	0.0018	0.0023	0.0026	0.0015
20-24	0.0055	0.0094	0.0065	0.0041	0.0086	0.0046	0.0040	0.0056	0.0024
25-29	0.0071	0.0184	0.0129	0.0138	0.0181	0.0078	0.0053	0.0111	0.0079
30-34	0.0072	0.0247	0.0124	0.0163	0.0269	0.0124	0.0081	0.0122	0.0112
35-39	0.0075	0.0252	0.0116	0.0180	0.0285	0.0151	0.0086	0.0161	0.0140
40-44	0.0099	0.0240	0.0141	0.0224	0.0258	0.0145	0.0084	0.0159	0.0132
45-49	0.0126	0.0262	0.0163	0.0239	0.0287	0.0191	0.0125	0.0181	0.0179
50-54	0.0113	0.0323	0.0262	0.0260	0.0340	0.0212	0.0150	0.0210	0.0174
55-59	0.0153	0.0369	0.0274	0.0262	0.0361	0.0204	0.0235	0.0282	0.0186
60-64	0.0275	0.0419	0.0612	0.0293	0.0452	0.0284	0.0293	0.0267	0.0291
65-69	0.0585	0.0418	0.0657	0.0399	0.0551	0.0387	0.0389	0.0511	0.0280
70-74	0.0708	0.0563	0.0727	0.0587	0.0751	0.0530	0.0636	0.0515	0.0421
75-79	0.0848	0.0683	0.0401	0.0833	0.0863	0.0739	0.0574	0.0524	0.0603
80-84	0.1434	0.0863	0.1707	0.1151	0.0944	0.0947	0.1123	0.0648	0.0673
85+	0.3057	0.0207	0.2405	0.1907	0.1849	0.1611	0.1776	0.1101	0.1347

Figure 9.2 Estimated adult African female central mortality rates for the 12 months prior to 9/10 October 2001, by age and province

Female	WC	EC	NC	FS	KZN	NW	GP	MP	LP
15-19	0.0013	0.0025	0.0025	0.0024	0.0031	0.0019	0.0021	0.0032	0.0010
20-24	0.0028	0.0080	0.0101	0.0082	0.0102	0.0078	0.0042	0.0064	0.0031
25-29	0.0047	0.0156	0.0154	0.0160	0.0187	0.0134	0.0055	0.0121	0.0056
30-34	0.0053	0.0176	0.0111	0.0170	0.0183	0.0166	0.0083	0.0126	0.0076
35-39	0.0065	0.0148	0.0230	0.0169	0.0156	0.0134	0.0084	0.0130	0.0074
40-44	0.0107	0.0141	0.0126	0.0156	0.0163	0.0123	0.0072	0.0106	0.0058
45-49	0.0095	0.0140	0.0105	0.0194	0.0140	0.0163	0.0081	0.0090	0.0076
50-54	0.0129	0.0140	0.0078	0.0137	0.0154	0.0134	0.0099	0.0135	0.0081
55-59	0.0081	0.0132	0.0109	0.0144	0.0150	0.0133	0.0125	0.0134	0.0074
60-64	0.0360	0.0176	0.0272	0.0162	0.0185	0.0225	0.0180	0.0173	0.0104
65-69	0.0337	0.0228	0.0289	0.0280	0.0255	0.0200	0.0234	0.0195	0.0119
70-74	0.0588	0.0304	0.0485	0.0342	0.0358	0.0345	0.0410	0.0316	0.0192
75-79	0.1375	0.0488	0.0433	0.0407	0.0431	0.0307	0.0484	0.0339	0.0220
80-84	0.1066	0.0653	0.0495	0.0687	0.0635	0.0527	0.0592	0.0589	0.0344
85+	0.2362	0.1403	0.1106	0.1657	0.1414	0.1235	0.1596	0.0915	0.0978

10 APPENDIX 2: REPORTED DEATHS BY YEAR OF DEATH, 1997– 2001

Figure 10.1 Adult African male deaths 1997-2001, Statistics SA (personal communication)

Males	1997	1998	1999	2000	2001	Total
0-4	2 759	7 189	12 937	14 652	15 210	52 747
5-9	322	701	1 230	1 439	1 496	5 188
10-14	325	670	1 022	1 187	1 200	4 404
15-19	637	1 515	2 660	2 869	2 998	10 679
20-24	1 220	3 232	5 552	6 241	6 247	22 492
25-29	1 604	4 692	9 030	10 762	12 111	38 199
30-34	1 796	5 197	10 709	12 854	14 774	45 330
35-39	1 694	5 173	10 450	12 682	14 546	44 545
40-44	1 629	4 744	9 353	11 278	12 753	39 757
45-49	1 718	4 733	8 839	10 215	11 281	36 786
50-54	1 534	4 018	7 568	9 140	10 271	32 531
55-59	1 855	4 196	7 420	7 755	8 036	29 262
60-64	1 523	3 647	6 115	7 503	7 979	26 767
65-69	1 884	3 827	5 792	6 063	6 242	23 808
70-74	1 628	3 761	5 984	6 562	6 878	24 813
75-79	1 803	3 276	4 777	4 847	4 974	19 677
80-84	881	2 224	3 356	4 006	4 419	14 886
85+	854	1 839	2 838	3 204	3 385	12 120
Total	26 180	65 508	116 552	134 081	145 517	487 838

Figure 10.2 Adult African female deaths 1997-2001, Statistics SA (personal communication)

Female	1997	1998	1999	2000	2001	Total
0-4	2 473	6 371	11 933	13 297	13 791	47 865
5-9	246	554	966	1 144	1 151	4 061
10-14	237	487	800	877	998	3 399
15-19	410	1 085	2 103	2 416	2 726	8 740
20-24	843	2 645	5 575	7 072	7 777	23 912
25-29	1 153	3 715	8 468	11 316	13 613	38 265
30-34	1 104	3 646	8 143	11 265	13 129	37 287
35-39	1 026	3 169	6 894	9 372	10 871	31 332
40-44	905	2 722	5 514	7 279	8 484	24 904
45-49	863	2 504	5 022	5 985	6 789	21 163
50-54	859	2 113	4 189	5 456	5 923	18 540
55-59	1 175	2 750	4 609	4 942	5 074	18 550
60-64	1 364	3 063	5 186	6 390	6 669	22 672
65-69	1 796	3 963	6 400	6 525	6 936	25 620
70-74	1 512	3 577	5 944	7 710	8 187	26 930
75-79	2 014	3 695	5 158	5 513	5 620	22 000
80-84	1 154	3 209	5 082	6 220	6 914	22 579
85+	1 629	3 509	5 211	6 127	6 535	23 011
Total	21 183	53 421	97 907	119 557	131 730	423 798