Old age mortality in South Africa, 1985-2011

Ronald David Richman
Centre for Actuarial Research (CARe)

Dissertation submitted for the degree of Master of Philosophy to the Faculty of Commerce, University of Cape Town

March 2017

PLAGIARISM DECLARATION

I, Ronald David Richman, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: 2017/03/10

ACKNOWLEDGEMENTS

I wish to express sincere appreciation to Professor Rob Dorrington. Firstly, for generously sharing from his knowledge of demographic methods in a manner that inspired me to pursue this research. Secondly, for his tireless efforts and the time spent supervising and supporting this research with invaluable feedback, from inception to finish. Thirdly, for his patience as I explored the many methodological dead-ends that eventually led to the final technique employed in this dissertation.

I wish to thank Gila, Batya and Avigayil for their love and dedication during the many hours that I spent preparing this dissertation.

ABSTRACT

Estimating the level and trend in population mortality rates at advanced ages in South Africa is complicated by problems with both the population and death data. Population and death data, particularly in developing countries, often suffer from age misreporting – age exaggeration and digit preference. Also, censuses may under- or overestimate the population and registration of deaths is usually incomplete in developing countries (Dorrington, Moultrie and Timæus 2004). To avoid these problems, the research in this dissertation relies on the method of extinct generations and its extensions (Thatcher, Kannisto and Andreev 2002) to re-estimate the population using only the death data, which is often recorded more accurately than the population data.

Since deaths are not reported completely in South Africa, the death data must be corrected before use. Death Distribution Methods (Moultrie, Dorrington, Hill *et al.* 2013) are used to correct the death data for incomplete registration of deaths. After correction, Near Extinct Generation methods (NEG) are used to re-estimate the population by projecting future deaths of nearly extinct cohorts.

After showing that mortality rates produced using the original NEG methods are biased because of age and year of birth heaping present in the South African death data, the NEG methods are adapted to the South African context. The adapted NEG model smooths the age and year of birth heaping in the death data and produces mortality rates that are less biased than the original NEG methods. This model - referred to as the NEG-GAM model in this research - is used to re-estimate the population at each age from 70 and above and to calculate mortality rates since 1985.

The population estimates aged 70+ produced using the NEG-GAM model match those from the 2011 census well. It is found that both the population and death data suffer from the same pattern of heaping, that the population and death data are affected by age exaggeration and that the death data are less affected by age exaggeration than the population data. The level and trend in mortality rates calculated using the NEG-GAM model are discussed and compared to the mortality rates in the Human Mortality Database and other studies of South African mortality. The mortality rates produced for the African and Coloured population groups appear too low at the older ages due to age exaggeration in the death data, while those for the Indian and White population groups appear to be reasonable over the entire age range. Mortality appears to be improving in

the age range 70-79 for the Coloured, Indian and White population groups and deteriorating slowly for the African population group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

Knowledge of mortality rates is fundamental to understanding the demography of a population and for performing actuarial calculations for pensions and life insurance. Although important at the younger ages, at the older ages (defined as 70 and older in this research) when mortality is higher, mortality rates become the essential information needed for most demographic and actuarial investigations.

Mortality rates are key inputs into actuarial valuations of cashflows contingent upon survival. While mortality rates for the insured population are generally available in standard mortality tables for most of the calculations performed by actuaries, actuaries often need to rely on population mortality rates at the older ages. For example, the mortality improvement model published by the Institute and Faculty of Actuaries (Continuous Mortality Investigation 2016) was calibrated on population mortality data. Another example is the calibration of the solvency capital requirement for longevity risk, prescribed for insurers in the European Union under Solvency II, that used population mortality data (Towers Perrin 2009).

The size of the aged population, together with a knowledge of mortality rates, is required to budget for the provision of services required by the elderly such as old age grants and pensions (Dorrington, Bradshaw and Wegner 1999), geriatric healthcare and in-patient hospital care (Bhat 1992). Knowledge of the level of mortality rates is necessary to indicate the "demographic well-being" (Dorrington, Moultrie and Timæus 2004: 1) of a population, to evaluate progress in improving public health and for population projections (Timaeus 1991).

Population ageing is a process driven by changes in the major demographic forces that shape a population – fertility, mortality and migration (Preston and Stokes 2012). Changes in the mortality rates of either the old-age population or the younger population will lead to a change in the proportion of the elderly in the population. A step towards understanding population ageing is gaining an understanding of the mortality rates to which an aged population is subject, and how these may be varying over time.

However, despite the importance of accurate estimates of old-age mortality rates and of the size of the aged population, and despite the fact that South Africa has the highest proportion of the population that are elderly of all countries on mainland Africa (Cohen, Menken, Hosegood *et al.* 2006; Stats SA 2012), little research focussing specifically on the aged population of South Africa has been performed, with the exception of the study of Machemedze (2009). Since estimating the level and trend in population mortality rates at advanced ages in South Africa is complicated by potential problems with both the population data and death data, it is likely that these difficulties are the reason for the limited research into mortality at the advanced ages.

As a consequence of the considerable uncertainty to which mortality rates for the old-age population in South Africa are subject, actuaries are sometimes forced to rely on out-dated studies of mortality or professional judgment. For example, the base life tables recommended for the determination of compensation in cases of damages by Koch (2016), and in common use at the time of this research, are the dated South African Life Tables of 1985 (SALT85), with adjustments to the mortality rates that depend on the earnings of the claimant. It is unclear if the mortality rates in these tables at the oldest ages are still applicable, given the history of mortality improvement in other countries. As another example, the data on insured lives in South Africa have not yielded reasonable estimates of mortality improvement (CSI 2012; Dorrington and Tootla 2007) and mortality improvements in other countries are often used as a proxy for the South African experience (Cooper-Williams, Albertyn and Lewis 2012). Lastly, the capital requirements for longevity risk for South African insurers (amounting to approximately R10 billion in the second quarter of 2016 (FSB 2016)) were set subjectively in the Solvency Assessment and Management (SAM) insurance capital regime on the basis of the European requirements (FSB 2014). It is therefore in the public interest to reduce the uncertainty surrounding old-age mortality in South Africa.

Whereas studies of adult mortality in developed countries are facilitated by the easy availability of reasonably accurate demographic data, for example, in the Human Mortality Database (Human Mortality Database 2008), in developing countries, such studies are hampered by inaccurate data arising from incomplete death registration and censuses of poor quality (Hill 2003; Hill, Choi and Timæus 2005). Even in developed countries, though, studies of old age mortality are hampered by the overstatement of age in population censuses and, to a lesser extent, in age at death data (Rosenwaike 1981). It is well established in the demographic literature that population data may suffer from the inaccurate recording of the age of the elderly, often manifesting as an exaggeration of the population at older ages (Preston, Elo and Stewart 1999). Age exaggeration has been identified in developed countries (see for example Coale and Caselli (1990)) and in developing countries (for example, age exaggeration in the South African population data has been observed by Machemedze (2009)). If population data suffering from age exaggeration are used to estimate mortality, then the mortality rates at the older ages will be biased downwards (Coale and Kisker 1990; Das Gupta 1991).

An alternative to using population data is to use demographic methods to reconstruct the old age population. A technique often used for such population reconstruction is the method of extinct generations (Vincent 1951) and its extensions to Nearly Extinct Generations (NEG) (Thatcher, Kannisto and Andreev 2002). These techniques rely on death data (which, when complete, are usually more accurate than census and survey data) to re-estimate the population and therefore provide an alternative measure of exposure in the calculation of mortality rates. Since the method allows one to estimate the population in each year and not only the census years, it allows one to study the change in mortality year by year.

While the method of extinct generations has been applied quite widely to developed country populations (Thatcher, Kannisto and Andreev 2002), the application of the method to developing country data is complicated by incomplete registration of death data. Since deaths in South Africa are not completely registered (Dorrington, Moultrie and Timæus 2004; Stats SA 2014), the data need to be corrected for under-reporting before estimating mortality rates. Other problems with the death data that need correction are age heaping, in which too many of those dying are reported to have been born in years of birth ending in a '0' (Bradshaw, Schneider, Laubscher *et al.* 2002; Machemedze 2009) and, potentially, post retirement age exaggeration.

Recently, a detailed study of South African old age mortality was performed by Machemedze (2009) who applied the NEG methods to re-estimate the population and calculate mortality rates at the national level. The current research seeks to update and extend this previous work by addressing some of the limitations Machemedze identified in his research. In particular, the study performed by Machemedze (2009) relied on data on deaths from 1997 to 2006, while updated data on deaths from 1984 to 2013 are available for this study. These updated death data contain, in addition, the year of birth as well as the year of death for the deaths occurring in 2012 and 2013, allowing the NEG methods to be applied to year of birth cohorts, which was an area of future research suggested by Machemedze. Lastly, the current research will be performed for each of the population groups (African, Coloured, Indian and White) separately as well as at a national level[1], addressing a limitation of Machemedze's research.

A specific area of focus in this research is how the NEG methods can be adapted to produce unbiased mortality rates despite the age and year of birth heaping that has been identified in the South African death data by Bradshaw, Schneider, Laubscher *et al.* (2002) and Machemedze (2009).

1.2 Research Question

This research aims to estimate the mortality rates, at individual ages, of the South African population and each of its major population groups (African, Coloured, Indian and White), from age 70 and older, for males and females separately, for each year 1985-2011 and to investigate the trend in these rates over this period.

1.3 Specific Objectives of the Research

The research aims to achieve the following specific objectives:

1. Re-estimate the population aged 70 and above by single age in each of the years 1984-2011. The objective is to produce estimates that are free from bias introduced by age misreporting (in particular age exaggeration) and under- or over counting, separately for each population group and each sex.
2. Estimate the number of deaths occurring in this population in each of the years in this time period, free of under-registration and age misreporting present in the underlying vital registration data. The number of deaths, for each population group and each sex, in single years of age from age 70 for each of the years 1985-2011 will be produced.
3. Produce age-specific mortality rates for single years of age from age 70, for each population group, each sex and each of the years 1985-2011.
4. Estimate the trend in the mortality rates over time.

[1] While there are arguments against disaggregating the data on the basis of race in much research, this research follows Moultrie and Dorrington (2012: 14), who argue that "in the case of a country with limited vital registration, official and other data, race offers an important proxy for a host of other socio-economic differences". From a technical perspective, a further reason for using data classified by race is that the actuarial and demographic techniques applied in this study rely on homogeneous data to produce accurate results.

1.4 Summary of the Dissertation

The balance of this dissertation is structured into the following four chapters.

Chapter 2 is a comprehensive review of the literature that examines problems with demographic data at the older ages, the method of extinct generations and its extensions to nearly extinct generations, the death distribution methods, the question of whether mortality increases exponentially with age and previous studies of mortality in South Africa.

Chapter 3 is concerned with the application of the Death Distribution Methods to the South African death data, nationally and for each population group. After discussing estimates of completeness of reporting and of population group in the intercensal periods between 1985 and 2011, the chapter discusses how annual estimates of completeness were derived.

Chapter 4 covers the application of the NEG methods. The discussion begins with an illustration of the problems encountered applying the original NEG methods to the South African data. A new NEG method based on a regression model (the NEG-GAM) model is then proposed, and the results of fitting this model are presented and compared to other studies.

Chapter 5 discusses to what extent the goals of the research were achieved, draws conclusions from the results and suggests areas of future research.

In addition, several appendices at the end of the dissertation provide mathematical proofs relating to the NEG methods and detailed results of the application of the DDMs and NEG methods.

2 LITERATURE REVIEW

2.1 Introduction

This chapter begins by discussing the errors that are prevalent in population and death data, the studies that have identified these errors and the circumstances that lead to such errors. The subsequent three sections focus on techniques that can be used to correct these data for errors. The first of these sections deals with the method of extinct generations, which is used by demographers to provide more accurate population estimates at older ages. The next section covers the death distribution methods, which are used to correct death data for incompleteness of reporting. The last part discusses other methods which may be applied to adjust demographic data for age exaggeration, age heaping and other errors. This is followed by a discussion of whether old age mortality rates increase exponentially with age. The last section concentrates on the various estimates of South African old age mortality and estimates of the size of the elderly population over time.

2.2 Patterns of errors in old-age demographic data

2.2.1 Description of patterns of age misreporting

Two types of age misreporting often occur in population and death data: age exaggeration, when people or deaths are incorrectly reported to be older than they are and age heaping (also called digit preference), when people or deaths are incorrectly reported at preferred ages or dates of birth, which are often those ending in a zero or a five. Although age exaggeration and age heaping are different types of age misreporting, they do not occur independently. Coale and Kisker (1986) show a strong linear correlation between a measure of age heaping at age 70 and the proportion of the elderly population aged 95 and older, a measure of age exaggeration. Thus age-heaping is a symptom of wider age misreporting that likely includes age exaggeration.

Age exaggeration occurs when there is a tendency for reported ages to be older than actual ages and results in too many reported at older ages and too few at younger ages. Preston, Elo and Stewart (1999) show that the same level of age exaggeration in both the population and death data will cause a greater bias in the population data than the death data.

Age heaping occurs when too many are recorded at the particular ages where a reporting preference exists and correspondingly too few at the other ages. Patterns of age heaping may be determined by reference to years of birth that are significant within a culture (Myers 1976), to years before or after significant events (Machemedze 2009) and to different calendars (Myers (1976), and may change depending on how age was elicited (Coale and Kisker 1986). "Reverse heaping of ages" (Myers 1976: 577), or year of birth heaping as it is referred to in this research, relates to the situation when age is calculated from year of birth, and convenient years of birth have been reported instead of the actual year of birth, leading to age heaping dependent on the reported years of birth.

2.2.2 Determinants of quality of age reporting

The accuracy of age reporting is often determined by the social institutions of those whose age is being recorded.

Contact with age-linked institutions, such as schools, the military and the government which enforce the importance of age in society, was shown by Hill, Preston, Elo *et al.* (1997) to predict the quality of reporting of age at death. The existence of birth records was shown by Rosenwaike and Hill (1996) to be a significant predictor of agreement between death certificates and Social Security records in the United States. Kannisto, Lauritsen, Thatcher *et al.* (1994) speculate that identity cards listing date of birth or the existence of a population register serve to reduce age misstatement in the Nordic countries.

Members of cultures which attach significance to date of birth, for example, due to the importance of astrology, will generally know their date of birth even if they are illiterate (Coale and Kisker 1986; Coale and Li 1991), leading to improved data about age. However those cultures which do not attach significance to date of birth may have poor data; for example, Caldwell and Igun (1971: 287) attributed the poor data on age collected in the Nigerian census of 1963 to the lack of cultural significance of "exact age in the Western sense" to those enumerated in traditional societies. National populations composed of different cultures may exhibit patterns of age exaggeration that are best dealt with by treating the data from each ethnic group separately (see, for example, Coale and Li (1991)).

However, Hill, Preston, Elo *et al.* (1997) caution that ethnic explanations of misreporting may be oversimplifications, since access to age-linked institutions may be determined by race and class. For example, Jowett and Li (1992) found that seemingly reasonable ethnic explanations for age heaping fail to be verified empirically in the Chinese context.

Age may be misreported due to a lack of knowledge on the part of the respondent of the actual age of the person being asked about. In censuses, it is often the case that a single respondent provides information for all members of a household, and the respondent may not know the exact ages of all members of the household (Coale and Kisker 1986; Rosenwaike and Logue 1983).

Age heaping in particular, has been attributed to the respondent's lack of knowledge of their exact age in single years (Pullum 1991). Even if age is known, there may be motives for consciously exaggerating age, such as the prestige endowed upon the aged (Bhat 1995; Kannisto 1988) or to qualify for benefits due to the aged Medvedev (1974). However, age could also be understated due to vanity (Preston, Elo, Rosenwaike *et al.* 1996) or because of greater societal value attached to younger women (Bhat 1995). It may be expected that age reported at death will be more accurate than the age reported in censuses (Condran, Himes and Preston 1991) since some of the incentives to misreport age do not exist after death.

2.2.3 Examples of age misreporting

Age misreporting has been documented in numerous studies over an extended period. Age misreporting has been identified directly through examination of death certificates and population records in matching studies and indirectly through the application of demographic techniques.

2.2.3.1 Matching studies

Misreporting of age at death has been identified directly through studies that match age of death recorded on death certificates with the age at death implied by other sources of data. Different patterns of bias have been found, and the extent of the misreporting has often been linked to the age and race of the decedent.

Rosenwaike and Logue (1983) found that the accuracy of death certificates of residents of Pennsylvania and New Jersey who died at ages 85+ was strongly dependent on race. Whereas 72 per cent of records relating to Whites matched exactly or had an age difference of less than one year, this was true for only 40 per cent of African-Americans. When age misreporting was identified, both Whites and African-Americans were more likely to have older ages at death reported than younger ages relative to the age calculated from the census records and the extent of the misreporting increased with age.

Focussing on African-American deaths in 1985, Preston, Elo, Rosenwaike *et al.* (1996) found severe misreporting for those aged 65+ at death, with only 44.6 per cent of records for females and 50.7 per cent of records for males showing no age discrepancy. A pattern of net underreporting of age on death certificates for the population aged 65 to 90, and a pattern of overstatement of age above the age of 90. The pattern of age understatement found in this study is different from the pattern of age exaggeration found by Rosenwaike and Logue (1983).

Hill, Preston and Rosenwaike (2000) matched records for white Americans with a reported age at death of 85 and older and found that minor age misreporting occurred, with exact age agreement for 91.8 per cent of the records, and two-thirds of the inconsistencies falling within one year. A pattern of underreporting of age for decedents aged 85-89 and exaggeration of age for decedents older than 90 was found. Age exaggeration worsened at older reported ages of the decedent on the death certificate.

Ortega and Garcia (1986), cited in Dechter and Preston (1990), performed a study matching death certificates to birth certificates in the Puriscal and Coronada regions of Costa Rica. They report a pattern of greater age understatement than exaggeration until the ages of 80-84, whereafter they report more age exaggeration than age understatement.

In contrast to these studies, Bourbeau and Desjardins (2002) found almost no age exaggeration when examining the death certificates of French-speaking centenarian decedents born in Quebec. Of the 209 records sampled, only four cases were shown to be erroneous.

Likely patterns of age misreporting in other populations can be identified by comparison with the cases identified above. For example, Dechter and Preston (1990) use the pattern of age exaggeration found by Ortega and Garcia (1986) to correct demographic data from a number of Latin American countries.

2.2.3.2 Indirect evidence of age misreporting

Age heaping is immediately discernible from the examination of aggregate population and death data, as noted in Myers (1954), Myers (1976) and Whelan (2009a), but the evidence of misreporting is still indirect in the sense that individual records were not examined in these studies. Age exaggeration, on the other hand, is not immediately discernible from the data and is generally diagnosed using demographic techniques such as population projections (Condran, Himes and Preston 1991; Dechter and Preston 1990), variable-*r* techniques[2] (Bhat 1995) and the method of extinct generations and its variants (Coale and Caselli 1990; Rosenwaike 1968). This section will focus on the evidence of age misreporting found using the method of extinct generations and will summarise the other findings.

The method of extinct generations (Vincent 1951) and its various extensions have been used in developed countries to estimate the size of the elderly population using data on deaths alone. The reconstructed population is considered to be more accurate than the enumerated population because age misreporting on death certificates is often less severe than that in censuses (Condran, Himes and Preston 1991; Rosenwaike 1981). A comparison of the reconstructed population to the enumerated population is therefore used to diagnose age misreporting in the enumerated population.

For example, studies performed by Rosenwaike (1968, 1979, 1981) calculated extinct generations estimates to assess the accuracy of the United States censuses and found evidence of age exaggeration from age 95 in the census of 1950 and from age 85 (for non-whites) in the census of 1960.

Das Gupta (1991) estimated the elderly population in 1980 to provide better estimates of the elderly population than were available from the census. He estimated that the proportion of the population aged 85-89 was higher and the proportion of the population aged older than 90 was lower, than implied by the census age distribution and found that the number of centenarians was exaggerated by 40 per cent compared to the census.

Coale and Caselli (1990) found evidence of age exaggeration for both sexes from the age of 94 in the 1971 and 1981 censuses of England and Wales using the Synthetic Extinct Generations method. The same technique applied to Italian census counts of the female population in 1971 and 1981, showed age exaggeration from the ages of 84 and 86 respectively. Similar findings of age exaggeration in England and Wales were apparent in a long run study by Thatcher (1992) who found that the extinct generations estimates were persistently higher than the census populations in the age group 80-89 (and lower in the older age groups) for all of the censuses from 1911-1981.

---

[2] The variable-*r* techniques are based on the generalisation of stable population theory (which assumes a constant growth rate) to apply to all populations, that is due to Preston and Coale (1982).

Terblanche and Wilson (2014) estimated the population of Australia aged 85 years and older for the years 1976-2011 and found that their estimates were close to the official population estimates in 2011, except for 18.5 per cent fewer male centenarians than the official estimates at that date.

Fewer studies have applied the method of extinct generations in developing countries, relative to developed countries. Machemedze (2009), whose study is discussed in detail in section 2.7.1, estimated the population of South Africa from the age of 75 and older, in the years of 1996, 2001 and 2007, and showed that the population numbers were exaggerated for ages 95 and above. Gomes and Turra (2009) applied the method of extinct generations using death data corrected for incomplete registration, to re-estimate the number of centenarians in Brazil in 1991 and assess whether the counts of centenarians in the census of 1991 were reasonable. They found that the estimates of centenarians were 66 per cent lower than the enumerated population, which they attributed to age exaggeration as well as imprecise imputation of missing census data.

Studies relying on other demographic techniques include: Bhat (1990) and Bhat (1995) who identified age exaggeration from age 70 for both males and females in the Indian censuses of 1971 and 1981 using variable-*r* techniques, Siegel and Passel (1976) who found that the number of centenarians were exaggerated in the United States in the years 1950-1970 and Demeny and Shorter (1968) who found evidence of age exaggeration in the Turkish population in the censuses of 1955 and 1960. Further studies include: Lee and Lam (1983) who found age exaggeration from the age of 65 in the decennial English censuses conducted from 1841 to 1931 from age 65, Condran, Himes and Preston (1991) who assessed consecutive censuses for 16 low-mortality countries for censuses covering the period 1950-1985 and found evidence of age misreporting in those countries without population registers and Dechter and Preston (1990) who found age misreporting in the death and population data of four Latin American countries.

Instead of relying on demographic techniques, some studies have instead compared the demographic experience of countries with poor data quality (for example, in the developing world, the study of Coale and Kisker (1986) or the developed world, the studies of Kannisto (1988) and Jdanov, Jasilionis, Soroko *et al.* (2008)) to countries with superior data quality, such as Sweden, to check if the data are accurate.

In summary, population and death data from both developed and developing countries are affected by age misreporting. Matching studies are the most direct way of quantifying age exaggeration in demographic data, but are expensive and suffer from sampling error. Instead of matching studies, demographic data can be checked for age misreporting using indirect techniques. If evidence of age exaggeration is found, the data should be corrected, and techniques for correcting the data are discussed next.

2.3 The method of extinct generations

2.3.1 Vincent's method of extinct generations

Demographers have often relied on the method of extinct generations (Vincent 1951) to produce population estimates that are less biased by age misreporting at advanced ages. Vincent's insight was that, assuming the population is closed to migration (a reasonable assumption given that the method is usually applied at the advanced ages), all of the future deaths of a cohort must arise from the current population of that cohort. The population can, therefore, be re-estimated using only the death data for that cohort. Thus, a drawback of the method is that it can only be applied once all of the deaths relating to the population of interest have occurred. Once these have occurred, however, there are several advantages of this procedure: the resulting population estimates may be more accurate than the enumerated population because it has been observed that age at death is recorded more accurately than the age reported in censuses or population surveys (Condran, Himes and Preston 1991; Humphrey 1970; Rosenwaike 1981), the population estimates will be guaranteed to correspond to the deaths (Humphrey 1970; Whelan 2009a), and it is possible to produce a time series of the population which allows one to produce a time series of mortality rates.

A variety of notations have been utilised in the literature to define Vincent's method mathematically, and no standard notation has emerged. For example, Humphrey (1970) used an age-cohort notation, Siegel and Passel (1976) and Thatcher (1992) used an age-year notation, and Thatcher, Kannisto and Andreev (2002) used a cohort-year notation. In this dissertation, unless otherwise indicated, the age-year notation will be used, with age as age last birthday on previous 1 January, since the death data are most often available in this form.

Defining $D(x,t)$ as the number of deaths in calendar year t of those aged x last birthday on the previous 1 January and $N(x,t)$ as the population on 1 January of the calendar year t aged x last birthday, then $N(x,t)=\sum_{i=0}^{\omega-x} D(x+i,t+i)$, where ω[3] is the age at which the last death occurs.

To apply the method, one requires the death data classified by age last birthday on previous 1 January, but tabular death data are usually recorded by age last birthday at death. In such cases (i.e. when date of birth has not been recorded, as is the case for most of the death data in South Africa until 2012), there are a number of approximations suggested in the literature for deriving the data by year of birth cohorts. Most common is to assume that half of those dying at age x last birthday in calendar year t were born in the earlier of the two possible calendar years, which are $t-x-1$ and $t-x$. On this assumption, $D(x,t)$ can be approximated by $D(x,t)=\frac{1}{2}*(d(x,t)+d(x+1,t))$ (Rosenwaike 1981), where $d(x,t)$ is the number of deaths reported at age x last birthday in calendar year t. More complicated (and presumably, more accurate) approximations of $D(x,t)$, based on death data from countries with more complete date of birth and date of death data, are given in Kannisto (1988) and Condran, Himes and Preston (1991) and other authors provide approximations suited to other specific definitions of age at death (Humphrey 1970; Thatcher 1992; Whelan 2009a).

2.3.2 Near extinct generations

Vincent's identity relies on the availability of the full time series of deaths until the age ω. Several extensions of the method have been devised to apply the method to cohorts that are 'almost' extinct. These are referred to as the near extinct generations (NEG) methods. The principle underlying these methods is that the age structure of deaths in past cohorts is representative of how future cohorts will experience deaths, and using this information, the future deaths of a cohort can be predicted. Since the NEG methods are generally applied at advanced ages where the population that must be estimated is small relative to the population being estimated at younger ages from the

---

[3] A procedure for estimating the maximum age at death ω, where this cannot be measured directly with any degree of accuracy, is given by Wilmoth, Andreev, Jdanov et al. (2007). This procedure calculates a measure of the number of deaths in the most recent l cohorts with death data available beyond age x by $\tilde{D}(x,t,l)=\frac{1}{l}*\sum_{j=0}^{l}\sum_{i=0}^{l-j} D(x+i,t-j)$ and the maximum age x is chosen as the age such that $\tilde{D}(x,t,l)<\frac{1}{2}$.

series of deaths, deviations from the projected series of deaths are not expected to bias the estimates of the population at the less advanced ages significantly (Wilmoth, Andreev, Jdanov *et al.* 2007).

The following NEG methods have been proposed:

1. The Das Gupta method (Das Gupta 1991) and a similar method proposed by Meslé and Vallin (2002)
2. The Survivor Ratio method (Dépoid 1973; Vincent 1951)
3. The curve fitting method of Siegel and Passel (1976)
4. The mortality decline method (Andreev 1999) and a special case proposed by Whelan (2009b)
5. The Das Gupta Advanced method (Andreev 2004)
6. The Survivor Ratio Advanced method suggested in Andreev (1999) and implemented by Terblanche and Wilson (2015)

These methods can be grouped into two types – the first three methods do not explicitly allow for improving mortality over time, whereas the next three methods explicitly allow for improving mortality. All of these methods use the history of deaths in past cohorts to predict the future deaths that will occur to current cohorts.

Related to these methods is the Synthetic Extinct Generation (SEG) method of Bennett and Horiuchi (1981, 1984) and the extension proposed by Coale and Caselli (1990). As described in the section on the Death Distribution Methods, the SEG method converts the deaths occurring in a period into the population in the middle of the period through a schedule of population growth rates (and net migration rates). In their review of the NEG methods, Terblanche and Wilson (2015) include the SEG method as a NEG method. However, this is erroneous since the SEG method does not project the future deaths that will occur in a cohort, which is the fundamental principle of the NEG methods. Although it can be shown in the special case when mortality (and net migration) rates remain constant until the extinction of a cohort, that the SEG estimate of the population is equivalent to the estimate produced by projecting current deaths forward using these rates, this is nothing more than assuming that the population is stable[4] above a certain age (Dorrington 2013c). Thus, the SEG method has not been included in this review of NEG methods and is instead discussed in Section 2.4.2.

---

[4] The assumption of stability is the basis of the death distribution method proposed by Preston, Coale, Trussell et al. (1980: eq 11) which projects the future deaths to be experienced by a cohort (stability being a reasonable assumption for some elderly populations as shown by Horiuchi and Coale (1982)).

2.3.2.1 The cohort ratio method of Das Gupta[5]

Das Gupta (1991) published his method as an output of research to estimate the age structure of the population aged 85 and older in the United States census of 1980. The method assumes that the ratio of deaths at consecutive ages does not vary much by cohort and that once a series of ratios has been estimated using data from earlier cohorts, the deaths in the current cohort can be predicted from these ratios. The experience of several previous cohorts is used to estimate the ratio to remove the effect of random fluctuations from the calculation.

Expressed mathematically, $\hat{D}(x,t) = D(x-1,t-1) * CR(x,t,m)$ where $\hat{D}(x,t)$ represents an estimate of the deaths of those aged x last birthday of 1 January in year t, $D(x-1,t-1)$ is the recorded deaths of those aged $x-1$ last birthday of 1 January in year $t-1$ and $CR(x,t,m)$ – the cohort ratio for age x in year t estimated using m cohorts – is the ratio of deaths at age x to deaths at age $x-1$ based on the experience of the m cohorts prior to the current cohort that was born in the year $t-x-1$.

$CR(x,t,m)$ is estimated by $\dfrac{\sum_{j=1}^{m} D(x,t-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)}$ (Das Gupta set m equal to three) and the population is then estimated as

$$\hat{N}(x,t) = D(x-1,t-1) * \sum_{i=1}^{\omega-x+1} \prod_{j=1}^{i} CR(x+j-1,t+j-1,m).$$

Das Gupta further assumed that the total population aged 85 and older enumerated in the 1980 census was correct (Das Gupta 1991: 155), and therefore that more accurate estimates of the population at ages 85+ could be produced by constraining the estimates of the population at ages 85+ so that in aggregate the re-estimated population would be equal to the estimate of the population 85+ from the census. He therefore calculated the ratio of the enumerated population in the open interval aged 85+ to the re-estimated population aged 85+, and multiplied the re-estimated population uniformly by this ratio[6].

---

[5] A similar algorithm is well known in actuarial science as the "chain-ladder method" and is widely applied to estimate outstanding claims reserves for short-term insurers, see for example Wüthrich and Merz (2008).

[6] As discussed in the following section, Thatcher, Kannisto and Andreev (2002) suggest this adjustment to counteract the effect of falling mortality rates.

In a discussion of the Das Gupta method, Andreev (1999: 10) notes that the cohort ratios can be used to estimate mortality directly (i.e. without re-estimating the population). Defining $q_{x,t}$ as the probability of death within the following year, at age x last birthday on 1 January of year t, then $\hat{q}_{x,t} = \frac{1}{\left(1 + \sum_{i=2}^{\omega - x+1} \prod_{j=2}^{i} CR(x+j-1, t+j-1, m)\right)}$. A simple proof for this expression is provided in Appendix A.1.

The Das Gupta method was criticised by Andreev (1999) on two counts. By comparing the mortality rates implied by the cohort ratios used in Das Gupta's study to the mortality for the same period of time in the Kannisto-Thatcher database, Andreev found that mortality was over-estimated by the Das Gupta method at all ages from 85-105, by between 5 per cent and 15 per cent (Andreev 1999: Figure 1.5). The cause of the exaggeration of mortality was that the cohort ratios used by Das Gupta were too low because they were estimated on the assumption that they were constant when in fact, there was a trend in the ratios over time.

The second criticism of the Das Gupta method was that abnormally high mortality within a cohort is projected forward by the Das Gupta method. For example, if $D(x-1, t-1)$ was too high due to excess mortality from a harsh winter, then the projected number of future deaths will be correspondingly too high[7].

Meslé and Vallin (2002) proposed a variation of the Das Gupta method where, instead of assuming that deaths at consecutive ages are in constant proportion to each other, it is assumed that $D(x,t)$ is a constant proportion of $D(x-1, t-1) + D(x, t-1)$.

2.3.2.2 The Survivor Ratio method

In their review of three near-extinct generations methods, Thatcher, Kannisto and Andreev (2002) attribute the Survivor Ratio method to Vincent (1951) and Dépoid (1973). The method assumes that the ratio of the survivors of a cohort to the sum of deaths experienced by cohorts in the previous several years is relatively constant and that once this ratio has been estimated, it can be applied to the current cohort to estimate the survivors. The experience of several previous cohorts is used to estimate the ratio to remove the effect of random fluctuations from the calculation.

[7] Actuaries reserving for short-term insurance liabilities often avoid using the chain-ladder method (analogous to the Das-Gupta method) if the reported claims which are to be projected are volatile, preferring the Bornhueter-Ferguson or Cape-Cod methods in these instances. Further, it is shown in Appendix A.4.2 that the Survivor Ratio method projects forward a smoothed estimate of the deaths, thus avoiding this problem.

Expressed mathematically, the population on 1 January in year t aged x last birthday, $\hat{N}(x,t)=\sum_{i=1}^{k} D(x-i,t-i) * SR(x,t,k,m)$ where $SR(x,t,k,m)$ – the Survivor Ratio – is the ratio of survivors at age x to the sum of deaths at ages $x-1$ to $x-k$ based on the experience of the m cohorts prior to the current cohort that was born in the year $t-x-1$. $SR(x,t,k,m)$ is estimated by $\frac{\sum_{j=1}^{m} N(x,t-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$ (Wilmoth, Andreev, Jdanov *et al.* 2007), which is a weighted average of the experience of the previous cohorts, using the sum of the deaths at the k previous ages in each cohort as weights, as shown in Appendix A.2. Using deaths as weights gives more weight to cohorts with more deaths. Thatcher, Kannisto and Andreev (2002) suggest that the number of years of deaths, k, and that the number of cohorts, m, be taken as five, although their tests of different combinations of these parameters did not reveal any optimal combination.

An alternative but equivalent form of the method (shown in Appendix A.3) is given in Andreev (1999: 5). The k-ages survival using the last m cohorts is defined as $\tilde{s}(x-k,t-k,k,m)=\frac{\sum_{i=1}^{m} N(x,t-i)}{\sum_{i=1}^{m} N(x-k,t-k-i)}$ for those aged x last birthday on 1 January of the year $y-k$.The population is then estimated as $\hat{N}(x,t)=\sum_{i=1}^{k} D(x-i,t-i) * \frac{\tilde{s}(x-k,t-k,k,m)}{1-\tilde{s}(x-k,t-k,k,m)})$.

Declining cohort mortality rates undermine the assumption that the survivor ratio derived from previous cohorts will be directly applicable to the next cohort (Thatcher, Kannisto and Andreev 2002: 5). To account for this, the survivor ratio can be modified by building on the fundamental assumption of the method that the number of survivors of a cohort are proportional to the deaths experienced in the last k years. If a cohort has experienced improved mortality relative to the previous cohort but the same absolute numbers of deaths are experienced in both the current and previous cohort, then the initial population of the current cohort must necessarily be larger than the previous cohort. Expressed mathematically, the population estimate is redefined as

$\hat{N}(x,t)=\sum_{i=1}^{k} D(x-i,t-i)*SR(x,t,k,m)*c$ (Andreev 1999: 7), where c is a factor that is greater than one when mortality rates are falling, and which has the effect of increasing the projected survivor ratio. (It is shown in Appendix A.5 that the mortality rates implied by the Survivor Ratio method are reduced in inverse proportion to c.) The mortality improvement factor c is estimated implicitly by constraining the re-estimated population to agree with an accurate estimate of the population in an open interval, or is calculated as the value that will ensure that the series of estimates of the population derived from the Survivor Ratio method merge with the official population estimates, at the age at which the official population estimates are believed to be more accurate than those produced by the Survivor Ratio method (Thatcher, Kannisto and Andreev 2002: 6). However, if reporting of deaths at older ages is not complete or if the population estimates are not accurate then calculating c will correct for these issues as well and therefore it is hard to interpret c as correcting solely for mortality improvements. Furthermore, since the same adjustment c is applied equally to all of the re-estimated cohorts, the effect of the adjustment is a correction for the average mortality improvements experienced by the cohorts over the period of estimation, on the possibly strong assumption that mortality improvement is the only reason for a difference between the re-estimated population and the official estimates.

The unconstrained version of the Survivor Ratio method has been shown to produce estimates of the population that are too low (Andreev 2004; Thatcher, Kannisto and Andreev 2002) and therefore the constrained version is recommended to achieve greater accuracy (Thatcher, Kannisto and Andreev 2002). However, an inaccurate constraint will lead to inaccurate estimates (Andreev 2004; Terblanche and Wilson 2015). Constraining the population to an independent estimate of the population will ensure that the size of the population estimated from the deaths is correct for that age group, however, since the adjustment is applied equally to all ages, the age structure of the estimates in the age group could still be inaccurate.

2.3.2.3 Consistency of the Das Gupta and Survivor Ratio methods

Terblanche and Wilson (2015) provide a proof of the equivalence of the Das Gupta method to the Survivor Ratio method, when a single year of deaths is used to estimate the survivor ratio. On the basis of this proof, Terblanche and Wilson conclude that the Das Gupta method is a variant of the Survivor Ratio method. As shown in Appendix A.4, this relationship can be generalised to m cohorts of deaths.

Since the proof shows that the future deaths are projected in the same way in both the Das Gupta and Survivor Ratio methods, the first criticism of the Das Gupta method levelled by Andreev (1999) – that mortality may be over-estimated - is equally applicable to the Survivor Ratio method. However, the second criticism – that if $D(x-1,t-1)$ is high due to volatility, for example, after a harsh winter, then the projection of $D(x-1,t-1)$ will result in an incorrect population estimate – is not applicable to the Survivor Ratio method, since the sensitivity of the projections is decreased by implicitly smoothing the estimates of the population before projection (see Appendix A.4.2).

2.3.2.4 Curve fitting method

In their study of centenarians in the United States, Siegel and Passel (1976) used a novel NEG method to produce estimates of the number of centenarians in 1970 to compare to the census counts of centenarians in that year. They investigated the statistical relationship between the number of deaths occurring to those aged 100 and older at the start of 1950, from 1950 until 1960, to the youngest age at death in that cohort. For example, they found that 240 deaths occurred at ages 100 and older in 1950, with the youngest age in that year being 100. They fitted a regression of the form $\ln(D(100+,t)) = m*(a-100)+b$, where $D(100+,t)$ is defined as the number of deaths occurring to the cohort aged 100 and older, where the cohort is measured from year t, and a is the youngest age at death in the cohort.

Although Siegel and Passel's calculation is somewhat specialised to deal with the aggregated number of deaths at age 100 and older, their principle can be applied by fitting a curve to death data that has not been aggregated, and calculating the number of future deaths at each age by extrapolation. Indeed, the differential of a function $f(a)$ describing the number of deaths at age a of a cohort, $\frac{df(a)}{da}$, is an estimate of the cohort ratio at age a, since it describes how the number of deaths changes from age to age.

2.3.2.5 Mortality decline method

Andreev (1999) noted that to predict accurately the number of survivors in a cohort, it is necessary to predict the mortality rates that will be experienced by that cohort. He therefore proposed an explicit model for the logarithm of the force of mortality consisting of two factors, with the first factor representing the level of mortality for age

x and the second representing a time effect that reduces the level of mortality over the years t. The model is a log-linear regression expressed mathematically for each age x separately as $\ln \mu_{x,t} = \alpha_x + \beta_x * t$, where $\mu_{x,t}$ is the force of mortality for those aged x in year t, fit through maximum likelihood estimation on the assumption that deaths follow a Binomial distribution. This model is fit to every age group from a minimum chosen age, such as age 80, expressed as x_0 in the year t_0. The model is then used to project the force of mortality for each age x via $\mu_{x,t} = e^{\alpha_x + \beta_x *(t)}$. Using the identity $q_{x,t} = 1 - e^{-\mu_{x,t}}$, where $q_{x,t}$ is the probability of those aged x dying in year t, the "Mortality Decline" Survivor Ratio applicable to the next cohort can be predicted as

$$SR(x,t)^{MD} = \frac{\prod_{i=0}^{x-x_0-1}(1-q(x_0+i,t_0+i))}{1-\prod_{i=1}^{x-x_0-1}(1-q(x_0+i,t_0+i))}$$

, where the numerator is equal to the proportion of the cohort that survives from age x_0 to x and the denominator is the proportion of the cohort that dies from age x_0 to x (Thatcher, Kannisto and Andreev (2002)). The surviving population is then estimated as $\hat{N}(x,t) = SR(x,t)^{MD} * \sum_{i=0}^{x-x_0-1} D(x_0+i,t_0+i)$. The procedure is applied recursively from the maximum age ω to re-estimate the size of the cohorts of interest.

A disadvantage of the Mortality Decline method is the large volume of data required to calibrate the regression models (Andreev 2004).

Whelan (2009b) proposed that instead of utilising a log-linear model of mortality, the Kannisto model for the force of mortality ($\mu_x = \frac{ae^{bx}}{1+ae^{bx}}$ (Thatcher, Kannisto and Vaupel 1998)), which he observed to give a good fit to mortality rates at higher ages for the Irish population, could be fit to the mortality experience of an extinct cohort. The modelled rates are then assumed to apply to the first cohort that has not yet become extinct, as an initial guess of the mortality rates applying to this cohort (if rapid mortality change is occurring, and it is thought that the mortality of the older cohort will be too heavy for the younger cohort, the coefficients of the Kannisto model should be forecast to derive a new set of coefficients applicable to the younger cohort) The initial exposed to risk in this cohort is estimated via $\hat{N}(x,t) = \dfrac{\sum_{i=1}^{\omega-1} D(x+i,t+i)}{1-\prod_{i=1}^{\omega-1} p_{x+i,t+i}}$ where $p_{x,t}$ is the probability of a person aged x exactly in year t surviving a further year, and the probabilities are evaluated from the fitted Kannisto model. The series of deaths and the initial exposed to risk can be used to construct a series of crude mortality rates for the current cohort, and proceeding in this manner, the Kannisto model is then fit to each cohort. Whelan, however, found that his method was very sensitive to errors in the data.

2.3.2.6 Das Gupta Advanced method

Andreev (2004) extended the Das Gupta method to allow for mortality improvements by incorporating an explicit adjustment to the cohort ratios into the procedure, which he termed the Das Gupta Advanced method. He was motivated by the need for an NEG method which would perform well even when unconstrained by an independent estimate of the elderly population, because the population estimates used to constrain NEG estimates at the relatively old ages where these methods are applied may themselves be inaccurate. Since population estimates at younger ages are expected to be more accurate (Andreev 2004: 233-234), these estimates at younger ages can be used to calibrate adjustments to the cohort ratios that enhance the accuracy of the procedure.

Defining the cohort ratio as before, the modified cohort ratio proposed by Andreev is $CR(x,t,m)^{DGA} = CR(x,t,m) * e^{\alpha+\beta*(x-x_o)}$, where x_0 is the age at which the cohort ratio sequence starts. The parameters α and β are estimated by least squares through minimising $\sum_{\forall i,j} [N(x+i,t+j) - \hat{N}(x+i,t+j)]^2$ where $\hat{N}(x,t)$ is the population estimated by applying the modified cohort ratios and i and j are chosen to specify the ages and years of accurate population counts.

Terblanche and Wilson (2015) note that the assumptions of a steady decrease in mortality rates over time, which lessens with age, do not always hold. They found that these assumptions were not supported by the mortality experience of Australia and New Zealand in the years 1976-1996, which in some years showed rates of mortality decline that increased with age.

2.3.2.7 Survivor Ratio Advanced method

Terblanche and Wilson (2015) proposed a similar model to the Mortality Decline method, but instead of allowing for a decline in mortality rates over time using a log-linear regression, they used the Survivor Ratio method, modified by applying a 10-period moving average of the rate of change in survivor ratios to the survivor ratios, keeping age constant and averaging survivor ratios over cohorts. Using the k -period survival defined above, $\tilde{s}(x-k, y-k, k, 1) = \frac{N(x,t)}{N(x-k, t-k)}$, they define the rate of change in survivor ratios at age x for the cohort born in the year $x-t$ using n previous cohorts of data as $r(x,t,n,k) = \left(\sum_{j=1}^{n-1} \frac{\tilde{s}(x-k, t-k-j, k, 1)}{\tilde{s}(x-k, t-k-j-1, k, 1)} - 1 \right) / (n-1)$. The modified *k*-period survival is then computed as $\tilde{s}(x-k, y-k, k, m)^{SRA} = \tilde{s}(x-k, y-k, k, m) * (1 + r(x,t,n,k))^{(m+1)/2}$, where the rate of change is applied from the mid-point of the m cohorts used in the calculation.

The Survivor Ratio Advanced method was shown not to perform as well as the constrained Survivor Ratio method when applied to data from Australia and New Zealand (Terblanche and Wilson 2015).

2.3.2.8 Choice of optimal Near Extinct Generations method

Thatcher, Kannisto and Andreev (2002) considered the performance of the Das Gupta method, the Survivor Ratio method and the Mortality Decline method in estimating the population aged 95+ and 100+ in several developed countries. The Survivor Ratio method was tested using five years of deaths and five cohorts of data and the Das Gupta method was tested using alternatively five cohorts or three cohorts of data. The Mortality Decline method was fit using ten years of data. The unconstrained versions of the NEG methods were tested together with two constrained versions, which were either constrained to the total population aged 90 and older or constrained to the population at age 90.

For all three of the NEG methods tested, the versions constrained to the population aged 90 and older outperformed the other two versions. It was found that the Survivor Ratio method, constrained to the population aged 90 and older and utilising the experience of five previous cohorts and five years of deaths outperformed the other methods. The more complicated Mortality Decline method did not produce better estimates of the population.

A second test of the methods compared performance when no constraint was used. When there was enough data to allow the use of the Mortality Decline method, it outperformed the other two methods for both sexes, at the ages 90+ and 95+. The Survivor Ratio method performed better than the Mortality Decline method at ages

100+. When less data were available, the results were split evenly between the Survivor Ratio method and the Das Gupta method, using the experience of the last 5 cohorts of data. It was found that both constrained versions of the NEG methods over-estimated the population by 1-5 per cent. The unconstrained versions tended to under-estimate the population by around 5-15 per cent.

Thatcher, Kannisto and Andreev (2002) did not offer any theoretical explanations why the constrained Survivor Ratio method would be expected to perform best and therefore it is difficult to generalise the conclusions of that study to other applications. The death and population data utilised in the study were of a high quality and the method was only applied at ages older than 90. In situations where the data are of lesser quality, as may be encountered in developing countries, or when the method must be applied at younger ages, it could be that the constrained Survivor Ratio method is not the optimal choice.

In another empirical study, Andreev (2004) compared the performance of the Das Gupta Advanced (DGA) method to the unconstrained Survivor Ratio and Das Gupta methods. Four years of death data and population data for the ages 81-89 were used to estimate the parameters for the DGA method. The same variants of the Survivor Ratio and Das Gupta methods as used by Thatcher, Kannisto and Andreev (2002) were used to estimate the populations at ages 90 and older for a selection of countries in the Human Mortality Database, for both males and females. The relative errors of the three methods compared to the extinct generations estimates for the same population were calculated. For larger countries (England and Wales, France, Japan and the United States), it was found that the DGA method outperformed the other methods. For smaller countries (Denmark, Finland, the Netherlands, Norway, Sweden, and Switzerland) it was found that, despite the fact that the relative errors produced by the DGA method were low with the largest error recorded being 1.7 per cent in these countries, the other methods outperformed the DGA method[8]. Compared to the unconstrained NEG methods which were prone to underestimate the population significantly, the DGA method did not produce a consistently biased relative error.

Terblanche and Wilson (2015) compared the performance of the constrained and unconstrained versions of the Survivor Ratio and Das Gupta methods, for various combinations of the number of years of deaths and number of cohorts, as well as two versions of their Survivor Ratio Advanced method. The methods were used to estimate the populations of Australia and New Zealand, at ages 90 and older for both sexes separately, in the years 1976-1996. The performance of the NEG methods was evaluated by comparing the results to population estimates produced using the method of extinct generations, for cohorts that were extinct. For both countries and sexes, in both age groups, it was found that the constrained Survivor Ratio was the optimal NEG method. The optimal choice parameters of the method, however, differed depending on whether the Australian or New Zealand population was being considered, and the open interval of the population being estimated. It was also found that the accuracy of the constrained NEG methods depended on the accuracy of the constraint used. For example, the population estimate used as a constraint for male centenarians was shown to be highly inaccurate when compared to the extinct generations estimate, and therefore it is not surprising that the unconstrained estimates were better performing than the constrained estimates in this case. The Das Gupta method had accuracy close to that of the Survivor Ratio method in larger population (Australians, females and those aged 90+), but performed poorly in smaller population (New Zealanders, males and those aged 100+)[9]. Regarding the size of the errors, it was found that the NEG methods have larger errors when used to estimate male populations compared to female populations and older populations compared to younger populations; again probably because of the greater volatility in these populations arising from smaller volumes of data.

In conclusion, when the NEG methods are not constrained, they tend to underestimate the population. Of the constrained NEG methods, the Survivor Raito method is the optimal choice, but it is only as accurate as the constraint that is used. Andreev's study indicates that in the absence of reliable population estimates at older ages to function as a constraint for the Survivor Ratio and Das Gupta methods, the DGA method should be considered for use (due to the consistently low errors and lack of bias in estimating the population). However, the DGA method can only be used if

---

[8] The better performance of the DGA method in larger countries can be explained in terms of the proof in Appendix A.4.2 that the Survivor Ratio method is equivalent to the Das Gupta method applied to smoothed data. Death data in larger countries are less variable, and therefore projections of these data using the Das Gupta method will be less prone to error than in smaller countries, where the Survivor Ratio method will be more accurate.

[9] This is probably due to the smoothing effect of the Survivor Ratio method discussed previously.

accurate population data at younger ages (i.e. younger than the age at which a constraint would be applied), which are required to calibrate the DGA model, are available.

2.3.2.9 Problems encountered when applying the method of Extinct Generations

All of the extinct generations and near extinct generations methods described above rely on the assumptions that the deaths are completely reported and that age is reported more accurately in the death data than in the population (at least above a certain age). Inaccurate death data will lead to inaccurate mortality estimation, as shown by the following examples.

Rosenwaike (1981) calculated mortality rates for both sexes, and Whites and non-Whites, for each of the years from 1951 to 1965 using extinct generations population estimates. It was found for non-Whites that the average rates calculated using extinct generations were higher than the official mortality rates. It was noted, however, that rates corrected based on estimates of age exaggeration from a record matching study were considerably higher than the rates calculated using the extinct generations estimates, leading Rosenwaike to conclude that the method of extinct generations provides some correction for age exaggeration, but may not completely solve the problem.

In a study of African-American mortality, Coale and Kisker (1990) applied the synthetic extinct generations method to re-estimate the population to eliminate the inconsistency in age reporting between the censuses and the vital registration. Mortality estimates based on the re-estimated population were higher than the official life table estimates for some ages. However, these estimates eventually fell beneath the life table rates and then failed to increase with age from the age of 90. The failure of the rates to increase with age as expected was attributed to exaggeration in the recorded age at death. Similar results were reported by Elo and Preston (1994) who applied the procedure of Bennett and Horiuchi (1984) to African-American death data in the United States. The mortality schedules produced were too low at older ages, a fact that was again attributed to overstatement of age at death in the death data.

A simulation study performed by Preston, Elo and Stewart (1999) used the age misreporting transition matrix estimated in Preston, Elo, Rosenwaike *et al.* (1996) to simulate patterns of reported deaths. The transition matrix was modified to produce patterns of symmetric age misstatement, net age understatement and net age overstatement, thus producing three sets of simulated deaths. The method of extinct generations was then applied to the three sets of deaths, and schedules of life expectancies were calculated. It was found that all three patterns of age misreporting produced schedules of life expectancy that were overstated. It is, however, noted in Bourbeau and Lebel (2000: 8) that the pattern of age misreporting used in Preston, Elo and Stewart (1999) to show that the method of extinct generations produced unreliable results in the presence of misreporting of age at death was "extreme" and that when less extreme misreporting of age at death is present, as in the case of, for example, the Canadian data, the method will produce reliable results.

In conclusion, the method of extinct generations is reliant on accurate death registration to produce valid results, and therefore if the death data suffer from defects the data would need to be corrected for these before use.

2.4 Death Distribution Methods

The Death Distribution Methods (DDMs) are defined by Hill (1981: 2-3) as methods that "use mathematical models of population age distributions to relate the age pattern of deaths to the age pattern of the population in such a way that the completeness of death registration can be estimated."

The population data are usually census results, or less often the results of a large-scale survey, and the death data are the deaths that have been registered in the vital registration system or deaths reported by households in a census. Once the death data have been corrected for incomplete registration, they provide a source of information for estimating mortality.

Three main categories of methods can be distinguished. The first category relies on the balancing equation (Brass 1975) to estimate the proportion of deaths missing from the registered death data. Brass's original method assumed a stable population and was extended to non-stable populations by Preston and Hill (1980) and Hill (1987).

The second category of methods recalculates the population based on death data that have been adjusted for the growth experienced by the cohorts comprising the population relative to each other. The original method was proposed by Preston, Coale, Trussell *et al.* (1980) for stable populations and was generalised to non-stable populations by Bennett and Horiuchi (1981, 1984).

The third category of methods estimates the mortality rate of the population and compares the estimated mortality rates with the reported mortality rate to derive an estimate of completeness of reporting. Two techniques in this category that will be reviewed are due to Preston and Lahiri (1991) and Courbage and Fargues (1979).

A key reference used for this review of the DDMs is Moultrie, Dorrington, Hill *et al.* (2013).

2.4.1 Growth Balance Methods

The balancing equation is a simple demographic identity that states that the growth rate of a population in open-ended age intervals is equal to the difference between the entry rate and exit rate (Hill and Queiroz 2010). Mathematically, between times t_1 and t_2, $r(x+,t_1,t_2) = b(x+,t_1,t_2) - d(x+,t_1,t_2)$, where $r(x+,t_1,t_2)$ is the growth rate of the population aged x and older, and $b(x+,t_1,t_2)$ and $d(x+,t_1,t_2)$ are the partial birth (i.e. the rate of turning x in the population) and death rates of the population aged x and older. A simpler version of the identify can be expressed in terms of absolute numbers in the population (instead of growth rates) as $N(x+,t_2) = N(x+,t_1) + B(x+,t_1,t_2) - D(x+,t_1,t_2)$, where $N(x+,t)$ is the population aged x and older at time t, $B(x+,t_1,t_2)$ is the "births" into the open interval $x+$ between times t_1 and t_2, and $D(x+,t_1,t_2)$ is the deaths occurring in the open interval $x+$ between times t_1 and t_2.

The Brass method assumes that the population is stable, i.e. $r(x+,t_1,t_2) = r$ for all ages and time periods and that a constant proportion of deaths, c, are not reported by age. On these assumptions, the balancing equation can be rearranged as $b(x+) = r - \frac{1}{c}.d^*(x+)$ where $d^*(x+) = c.d(x+)$. Preston (1984) points out that $\frac{1}{c}$ could be solved using data for any pair of ages. However, making use of all of the available information to provide a more accurate estimate of the completeness, the parameters are usually estimated by fitting a straight line to $b(x+)$ and $d^*(x+)$.

Preston and Hill (1980) used the second identity described above to derive a death distribution method that uses population data at two different points in time. Ignoring migration and assuming, without loss of generality, that population data are available at instances five years apart, $N(x,t+5) = N(x-5,t) - D(x-5,t,t+5)$, where $D(x-5,t,t+5)$ represents the deaths occurring to those aged between $x-5$ and x at over the period t and $t+5$. Defining $D^*(x,t,t+5) = c.D(x,t,t+5)$ as the recorded deaths and $N^*(x-5,x,t) = k_t.N(x-5,x,t)$ as the population recorded in a census at time t, which may be under- or over-counted relative to the true population by a factor k_t, the above equation can be rewritten in terms of the reported quantities as

$\frac{N^*(x-5,t_1)}{N^*(x,t_2)} = \frac{k_1}{k_2} + k_1 . \frac{D^*(x-5,t_1,t_2)}{cN^*(x,t_2)}$. Assuming that the under- or over-count of the censuses and the incompleteness of death reporting are invariant with age, then the parameters k_1, k_2 and c can be estimated through regression (once the predictor and the response of the equation have been calculated for several ages using the reported data). The Preston and Hill method therefore provides estimates of the coverage of two consecutive censuses and avoids the assumption of stability made in the Brass technique. The Preston and Hill method was found to be "unstable" in practice because of the distorting effect of age misreporting on the number of reported cohort deaths (Hill 1987: 9).

Hill (1987) proposed the General Growth Balance (GGB) method, which relies on the balancing equation (like Brass's method, thus avoiding the problem of instability in Preston and Hill's method), and allows for changing census coverage (like Preston and Hill's method). The GGB method dispenses with the assumption of stability made in Brass's method and estimates age-specific growth rates using two consecutive censuses. Since $r(x+) = \frac{\ln\left(\frac{N(x+,t_2)}{N(x+,t_1)}\right)}{t_2 - t_1}$, it can be shown that $r(x+,t_1,t_2) = r^*(x+,t_1,t_2) + \frac{\ln\left(\frac{k_1}{k_2}\right)}{t_2 - t_1}$, allowing the balancing equation to be rearranged as $b^*(x+,t_1,t_2) - r^*(x+,t_1,t_2) = \alpha + \beta . d^*(x+,t_1,t_2)$, where $\alpha = \frac{\ln(k_1 / k_2)}{t_2 - t_1}$ and $\beta = \frac{(k_1 k_2)^{\frac{1}{2}}}{c}$.

The reported growth and partial birth and death rates can be calculated from the data, allowing the parameters α and β to be estimated by regression. On the assumption that the better enumerated of the two censuses has a coverage of 100 per cent, the coverage of the other census can be calculated from the estimated parameters by setting the larger of the parameters k_1 or k_2 equal to unity and solving for the remaining parameter (Dorrington 2013b). The estimates of completeness produced by the GGB method are relative to the census with the relatively higher enumeration.

Migration can be allowed for by adding the partial migration rate to the equation above, which becomes $b^*(x+,t_1,t_2) - r^*(x+,t_1,t_2) + i^*(x+,t_1,t_2) = \alpha + \beta.d^*(x+,t_1,t_2)$, where $i^*(x+,t_1,t_2)$ is the partial (net) migration rate for those aged $x+$ between times t_1 and t_2 (Dorrington 2013b).

Since the predictors and responses of the GGB equation are arbitrary (i.e. the term $b^*(x+,t_1,t_2) - r^*(x+,t_1,t_2)$ could be used as the predictor instead of $d^*(x+,t_1,t_2)$), Bhat (2002) recommended estimating the parameters using orthogonal regression, which is a procedure that minimises the vertical and horizontal distances in the regression, as opposed to normal linear regression which minimises only the vertical distance alone.

2.4.2 Methods of reconstructing the population from deaths

The methods discussed in this section are similar in that the death data are used to provide an estimate of the population, either through the method of extinct generations or through the use of population models. Comparing the estimated population to the reported population provides an estimate of completeness of reporting.

2.4.2.1 The method of extinct generations

Assuming that completeness of death reporting does not vary with age, and over years of reporting, then completeness of reporting of deaths relative to the population can be estimated using the method of extinct generations as $\frac{N^*(x,t)}{N(x,t)} = \frac{\sum_{i=0}^{\omega-x} D^*(x+i,t+i)}{N(x,t)} = c$.

If the completeness of reporting of deaths does vary over the years of reporting, then $\frac{N^*(x,t)}{N(x,t)} = \frac{\sum_{i=0}^{\omega-x} D^*(x+i,t+i)}{N(x,t)} = \frac{\sum_{i=0}^{\omega-x} D(x+i,t+i).c_{t+i}}{\sum_{i=0}^{\omega-x} D(x+i,t+i)} = c^*$ where c_t is the completeness of reporting in year t and c^* is a weighted average of the completeness of reporting, where the weights are the deaths occurring to the cohort at each time t. This estimate of completeness will be biased towards the level of completeness of reporting of the deaths at the ages with the most deaths, since these ages will have more weight in the average.

If the full time series of deaths is not available, then the unconstrained[10] NEG methods could theoretically be applied to project future deaths. However, the unconstrained NEG methods tend to underestimate the population, as discussed in section 2.3.2.8, and, therefore, completeness of reporting estimated using these methods may be somewhat understated. Furthermore, if the completeness of reporting of deaths is changing over time, then the cohort (and survivor) ratios calculated using the reported death data will be biased. Since population estimates are calculated by multiplying the cohort ratios together, the population estimate determined using biased cohort ratios may be unreliable.

2.4.2.2 Population models

Since in general the time series of deaths experienced by each cohort is not yet available, the methods discussed here rely on relating the current deaths experienced in the population to the deaths that will be experienced in each cohort under consideration.

In a stable population[11], the future deaths to be experienced by a cohort can be projected using the current deaths in the population and the growth rate of the population (Preston, Coale, Trussell *et al.* 1980). Using the notation above, the relationship can be expressed as $D(x,t) = D(x,t-k).e^{rk}$, where r is the growth rate of the stable population. The size of a cohort at time t is equal to the sum of the future deaths and, therefore, $N(x,t) = \sum_{i=0}^{\omega-x} D(x+i,t).e^{ri}$. Assuming that the completeness of death reporting does not vary by age, then c may be estimated as $\frac{N^*(x,t)}{N(x,t)}$, where $N^*(x,t) = \sum_{i=0}^{\omega-x} D^*(x+i,t).e^{ri}$.

---

[10] Since constraining the re-estimated population to the reported population will produce an estimated completeness of reporting of 100%, using the constrained NEG methods to estimate the population for the purpose of calculating completeness of reporting is a futile exercise.

[11] Since it has been shown Horiuchi and Coale (1982) that populations at the older ages may be well approximated by the assumption of a stable population, it is worth considering the use of this method when estimating completeness of reporting for the aged. The method does not require the use of age specific growth rates, the accuracy of which may be distorted by age misreporting and therefore the estimates of completeness produced by this method may be more accurate than those produced with more complicated methods if the assumption of stability is not severely violated.

The Preston and Coale method, which only applies to a stable population, was generalised to apply to any closed population by what has become known as the Synthetic Extinct Generations[12] method (SEG) (Bennett and Horiuchi 1981, 1984).

The SEG method relies on the insight that one cohort in a population is related to another cohort using the rate at which the cohorts have grown relative to each other. Preston and Bennett (1983) provide an example of two cohorts, one aged 16 in year $t-1$ and the other in year t. If r_{16} is the rate of growth at which the cohort aged 16 in year t grew from the cohort aged 16 in year $t-1$ then it follows simply that $N(16,t)=N(16,t-1)e^{r_{16}}$. Similarly, by rearranging the previous equation, those aged 17 in year t can be calculated as $N(17,t)=N(16,t)e^{-r_{16}}p_{16}$, using the growth rate and p_{16}, the probability of survival of those aged 16 in the current year, t. The current period mortality rates can be used (i.e. the mortality rate of those aged 16 in the prior year does not need to be used) since any difference in mortality experience of the two cohorts is already accounted for in the growth rate (Horiuchi and Preston 1988). Similarly, it is possible to approximate the population aged 16 in year t by using data on those aged 17 in year t and the deaths occurring at age 16 in year $t-1$, as $N(16,t)\approx N(17,t)e^{r_{16}}+d(16,t-1)e^{\frac{r_{16}}{2}}$. The continuous time identity on which this formula is based is given in Bennett and Horiuchi (1981) as $N(x,t)=\int_{x}^{\omega}D(x,t)e^{\int_{x}^{y}r(z.t)dz}dy$. If $D(x,t)=D^{*}(x,t)*\frac{1}{c}$ for all ages x, an estimate of the completeness of reporting, relative to the original population, can be derived as $c=\frac{N^{*}(x,t)}{N(x,t)}$ where $N^{*}(x,t)=\int_{x}^{\omega}D^{*}(x,t)e^{\int_{x}^{y}r(z.t)dz}dy$.

To apply the method, one must discretise the integral in the previous equations. Bennett and Horiuchi (1981: eq 6) suggested the following discrete approximation to the integral, which assumes that data are available in five year age groups, $N^{*}(x,t)=N^{*}(x+5,t)e^{5r(x,x+5,t_1,t_2)}+D^{*}(x,x+5,t_1,t_2)e^{2.5r(x,x+5,t_1,t_2)}$ where $N^{*}(x,t)$ is the

[12] Dorrington (2013c) points out that this description is misleading since the method does not rely on the time series of future deaths as in the extinct generations method, but uses the history of the population, as contained in the age-specific growth rates, to relate the current period deaths to the deaths to be experienced by each cohort.

reported population aged x at time t, $r(x, x+5, t_1, t_2)$ is the growth rate of the population aged from x to $x+5$ over the period t_1 to t_2 and $D^*(x, x+5, t_1, t_2)$ is the reported deaths of those aged between x and $x+5$ over the period t_1 to t_2. $N^*(x,t)$ can be approximated through $\sqrt{N^*(x-5, x, t_1) N^*(x, x+5, t_2)} * (t_2 - t_1)/5$, where $N^*(x, x+5, t_i)$ is the population aged from x to $x+5$ at time t_i.

The approximations suggested thus far are applicable to discrete counts of the population and of deaths in five-year age bands. An approximation is still needed for the highest ages which are generally grouped together in an open interval at a high age such as 80 and older or 90 and older. Bennett and Horiuchi (1981: eq 7) suggest an approximation based on stable populations for the open interval $x+$ as

$$N^*(x+, t) = D^*(x+, t_1, t_2) * \left(e^{r(x+,t)e(x+,t_1,t_2)} - \left(\frac{r(x+,t)e(x,t_1,t_2)^2}{6} \right) \right)$$ where $e(x, t_1, t_2)$ is the life expectancy at age x between times t_1 to t_2. Dorrington (2013c) suggests various methods for estimating the life expectancy in the open interval.

An important extension to the SEG method is the allowance for differential completeness of coverage of the two censuses used to calculate the age-specific growth rates. The extension was originally proposed in Bennett and Horiuchi (1981: footnote 10) and is elaborated on in Dorrington (2013c).

Defining all terms as in the derivation of the GGB method, the equation below shows that the addition of a term δ to the calculated growth rates produces the growth rates corrected for change in census coverage over the period t_1 to t_2:

$$r(x+, t_1, t_2) = \frac{\ln\left(\frac{N(x+, t_2)}{N(x+, t_1)} \right)}{t_2 - t_1}$$

$$= \frac{\ln\left(\frac{N^*(x+, t_2)/k_2}{N^*(x+, t_1)/k_1} \right)}{t_2 - t_1}$$

$$= r^*(x+, t_1, t_2) + \delta$$

The value of δ can be taken from the estimate of $\ln(\frac{k_2}{k_1})$ from the GGB method, or solved iteratively as the value that produces the most level sequence of ratios $\frac{N^*(x, x+5, t)}{N(x, x+5, t)}$ (Dorrington 2013c).

The effect of migration can be allowed for, by deducting the net migration rate from the estimated growth rate (Dorrington 2013c).

2.4.3 Other techniques for estimating completeness of reporting

2.4.3.1 Courbage and Fargues

Courbage and Fargues (1979) noted that the age distribution of reported deaths can be used to identify a level of a model life table. Comparison of the mortality rates implied by the model life table to the mortality rates calculated using reported deaths provides an estimate of the under-reporting. The method assumes that the completeness of reporting of deaths is invariant with age and that a model life table could be found that represents the mortality of the population. On these assumptions, an 'index of concentration' is calculated as $i(\alpha,\beta)=\frac{D(\beta+)}{D(\alpha+)}=\frac{D^*(\beta+)}{D^*(\alpha+)}$ where $\alpha<\beta$, $D(\alpha+)$ is the deaths in the open interval $\alpha+$ and $D^*(\alpha+)$ is the reported deaths in the interval $\alpha+$ ($D(\beta+)$ and $D^*(\beta+)$ are defined similarly). Since in the absence of extreme distortions in mortality such as HIV/AIDS, lower levels of mortality result in a higher proportion of deaths at older ages and thus a higher index of concentration, it should be possible to use this index to select a model life table with the correct level of mortality. The mortality rate derived from the model life table is then compared to the reported level of mortality to derive a correction for under reporting of deaths as $c=\frac{m^*_{\alpha+}}{m_{\alpha+}}$, where $m_{\alpha+}$ represents the mortality rate at ages $\alpha+$.

2.4.3.2 Preston and Lahiri's "Shortcut" method

Preston and Lahiri (1991) proposed a method that replaces the assumption of completeness of reporting that is invariant with age (used in the SEG) with the assumption that the mean age at death can be measured reliably. The method is based on a formula expressing the rate of change of the mean age of the population over time in terms of birth and death rates and the difference between the mean age of deaths and the mean age of the population at time *t*, namely,

$\frac{dA_P(t)}{dt}=1-b(0,t)A_P(t)-d(0+,t)*(A_D(t)-A_P(t))$ where $A_P(t)$ is the mean age of the population at time t, $A_D(t)$ is the mean age at death of the population at time t, $b(0,t)$ is the birth rate of the population at time t and $d(0+,t)$ is the death rate of the population at time t. Since the growth rate of a closed population is the difference between the birth rate and the death rate, i.e., $r(0+,t) = b(0,t) - d(0+,t)$, the previous equation can be rearranged as $\frac{dA_P(t)}{dt} = 1 - r(0+,t)A_P(t) - d(0+,t)A_D(t)$. Isolating the death rate term, the equation becomes $d(0+,t) = \frac{1 - r(0+,t)A_P(t) - \frac{dA_P(t)}{dt}}{A_D(t)}$. The death rate of the population can therefore be estimated from estimates of the growth rate, the mean age of the population and the mean age at death, and the rate of change of the mean age of the population. The resulting death rate of the population is then compared to the death rate calculated from the registered deaths, providing an estimate of completeness of death reporting.

To implement the method, childhood deaths (which often have a different completeness of reporting than adult deaths) are excluded and the formula above is rewritten in terms of quantities applicable to the open interval 5+ or 10+. The mean age of the population is calculated by assigning the population within five-year age bands, x to $x+5$, to the age at the middle of that age band, $x+2.5$. The mean age of the population in the open interval is calculated using an approximation, based on the assumption that the population above age x is stable.

Since Preston and Lahiri's method does not make the assumption that completeness of reporting of deaths is invariant with age, in cases where it is suspected that this assumption is violated more accurate estimates may be provided compared to estimates from one of the other methods discussed above, provided that the mean age of the population and the mean at age at death can be measured sufficiently accurately.

2.4.4 Performance of Death Distribution Methods

A number of studies (Dorrington, Timæus and Moultrie 2008; Hill 2001; Hill and Choi 2004; Hill, You and Choi 2009; Murray, Rajaratnam, Marcus *et al.* 2010) have recently assessed the performance of the GGB and SEG methods under various scenarios, with similar conclusions regarding the optimal choice of a method.

Early work by Hill (2001) tested the GGB and SEG methods (the SEG method tested did not include the correction for change in census coverage) in the presence of changing census coverage and age misreporting. With no data errors, both the GGB and SEG methods performed well. Age misreporting gave rise to larger errors in the GGB than the SEG method whereas change in coverage resulted in smaller errors for the GGB method than the SEG method. Given this, a combined method was also tested – the GGB method was used to estimate the change in census coverage and the population numbers adjusted accordingly, before applying the SEG. The results of the combined procedure were found to be more accurate than applying the GGB method alone. An extended version of Hill's study, reaching the same conclusion, was described in Hill and Choi (2004).

It is notable that calculating growth rates from two sets of census data with different levels of coverage will lead to erroneous growth rates and it is, therefore, unsurprising that the SEG method (without correction) performed poorly in this instance. However, the version of the SEG method (SEG+delta) originally suggested in Bennett and Horiuchi (1981), that caters for this problem was discussed in Dorrington, Timæus and Moultrie (2008) who extended the simulation study[13] performed by Hill and Choi (2004). The adjustment to the growth rates, δ, was derived so as to minimise the mean absolute difference of the age-specific estimates of completeness from the median estimate of completeness. The SEG+delta method was found to be slightly more accurate than the GGB+SEG method in the majority of the scenarios tested, indicating that the earlier conclusions of Hill and Choi (2004) – that the GGB+SEG method was more accurate than the SEG method - were attributable to using the SEG method without allowing for changing census coverage.

Hill, You and Choi (2009) then extended the work of Hill and Choi (2004) by including the SEG+delta[14] method in their analysis. They concluded that all of the methods perform well for the data errors for which they are designed, as well as in the cases of age-misreporting in the population and death data and age varying census coverage. The root mean squared errors (RMSE) were lowest for the GGB+SEG method, but only by a small margin, leading them to conclude that the GGB+SEG method is optimal. Noting that their results disagree with the results of Dorrington, Timæus and Moultrie (2008), who recommend the SEG+delta method, they hypothesise that the different recommendations are due to their use of the RMSE criteria and note that the SEG+delta method would be optimal in their study using the criterion of best performance in the most scenarios (used by Dorrington, Timæus and Moultrie (2008)). However, Dorrington, Timæus and Moultrie (2008) note that the

[13] A second contribution of the study was to include scenarios representative of an African population experiencing an HIV epidemic. The scenarios were created by allowing for an adult HIV prevalence rate of 17% together with the same patterns of data errors as used in .

[14] The SEG+delta method that was tested was slightly different from the version tested by Dorrington, Timæus and Moultrie (2008). Instead of minimising the mean absolute difference of age specific estimates of completeness from the median estimate of completeness, Hill, You and Choi (2009) adjusted the coverage of the first census relative to the second census until a regression of the completeness estimates on age had a slope of zero.

SEG+delta method had a lower RMSE than the GGB+SEG method in the scenarios that they tested.

Thus, in conclusion, both Dorrington, Timæus and Moultrie (2008) and Hill, You and Choi (2009) agree that the SEG+delta method is the best performing of the methods in the most scenarios tested, but they disagree which method produces the lowest RMSE. On this basis, the approach taken in this research is to use the SEG+delta method and also apply the GGB method to check if the estimates of completeness from the GGB method are significantly different (since it has been shown that the GGB method sometimes outperforms the SEG+delta method).

Murray, Rajaratnam, Marcus *et al.* (2010) performed a simulation study to measure the performance of the GGB, SEG and GGB+SEG methods and to find the optimal age ranges or "age trims" in which the DDMs should be applied. The methods were applied to three populations – a micro-simulation of a population over 150 years, data from the US counties from 1990-2000 and data from high income countries from 1950-2000. The optimal age-trims for the methods were found to be 40-70 for the GGB method, 55-80 for the SEG method and 50-70 for the GGB+SEG method. All three methods, with the selected age-trims, performed well in the high income countries, with slightly better performance of the SEG method. Age heaping had little effect on the median relative errors of the methods and did not affect the optimal choice of age-trims for the methods. Different levels of migration affected the optimal choice of age trim for the GGB method only. Defining "uncertainty intervals" as confidence intervals generated from a normal distribution that capture the true value of the completeness of reporting within the simulations 95 per cent of the time, they concluded that confidence intervals for the three methods are in the range of 20 per cent-25 per cent of the best estimate of completeness, which, if correct, implies that the DDMs are not particularly accurate.

However, this study can be criticised on a number of grounds. Notably, the study asserts that the DDMs assume no net migration; however, migration can be incorporated into the DDMs (Dorrington 2013b, c; Hill and Queiroz 2010). Applying the DDMs without allowing for migration, when in fact migration has occurred, would likely produce uncertainty intervals that are too wide and age ranges that are only optimal when migration is ignored. Also, the migration assumptions used in this study (scenarios were constructed assuming population migration rates of 5, 10 and 25 per thousand) are seemingly unrealistic – the highest migration rate in South Africa in the period 1996-2011 of any age group was 5.6 per thousand (based on the migration data in Moultrie and Dorrington (2015)). Finally, because of the large number of simulations and lack of clearly defined scenarios it is difficult to pinpoint the underlying reasons why certain age-trims perform better than others, and in which circumstances these conclusions will hold.

2.5 Other techniques for correcting population and death data

Besides the method of extinct generations, several other techniques have been proposed to identify and correct demographic data for age misreporting. Most of these techniques rely on model life tables or the assumption of stability. Since mortality rates estimated using demographic data corrected with these methods will be influenced by the choice of model life table, these methods are of limited use for this research.

2.5.1 Techniques based on population projections

On the rather strong assumption that an accurate age distribution of the population is available, then age exaggeration can be identified by projecting this recorded population to the date of a later census. Mortality could be projected using a model life table or reported deaths (corrected for underreporting) could be used. If mortality and migration have been accounted for correctly, then the projected age-specific population should be similar to the enumerated population. For example, if it is known that age exaggeration begins only after age 70, then the population aged 60-69 can be projected forward for ten years, to identify age exaggeration at the ages 70-79. If the projected and recorded populations are not similar, then this may be indicative of age misreporting (or undercount of one census relative to the other). Projections at the younger ages for short periods will not be affected by errors in the mortality estimates and at the older ages migration can be assumed to be negligible. However, a major disadvantage of population projections is that differences in census coverage will produce spurious results unless these changes are allowed for.

Dechter and Preston (1990) showed how a transition matrix of age misreporting probabilities could be estimated once population projections between two censuses have been performed. Ratios of the expected population to the actual population are calculated, with a ratio less than one indicating age exaggeration. The reported population and deaths are then multiplied by an initial transition matrix (they used the transition matrix derived empirically through a study performed by Ortega and Garcia (1986)) and the ratios of expected to actual are recalculated. The elements of the transition matrix are then adjusted, and the process repeated until the ratios of the expected to actual population are close to unity. A similar example of estimating a reporting transition matrix is in Hill, Preston and Rosenwaike (2000: Footnote 3).

Related to population projections is the technique proposed by Demeny and Shorter (1968) to correct two census age distributions for age misreporting. The technique works by estimating the pattern of age misreporting on the assumptions that: mortality rates in the population are known, the pattern of age misreporting is the same in both censuses, the total counts of both censuses are complete and the censuses are five years apart. Compared to population projections, the advantage of the Demeny-Shorter technique is that an accurate age distribution is not required as an input into the projections. The technique was extended by Das Gupta (1975) (who showed that the technique implicitly assumes a constant age structure of the population and relaxed this assumption), Ntozi (1978) (who extended the method to work with three censuses) , Feeney (1997) (who extended the technique to the case when the censuses are not five years apart) , and Lee (1982) (who, amongst other refinements, relaxed the assumption that mortality rates in the population are known).

A similar technique was suggested by Pullum (1991), who used a Generalised Linear Model (GLM) to model the surviving members of a cohort as a combination of an age-specific misreporting effect, a cohort effect and a sample size effect (respectively a_i, k_t, c). The probability of survival for each age, chosen from a model life table, enters the model as an offset in the GLM. The population at age i at time t is modelled as $\ln(N_{i,t}) = a_i + k_t + c + \ln(S_{i,t})$. Once the age misreporting effect, a_i, has been estimated, the corrected age distribution of the population can be produced.

2.5.2 Techniques based on model age structures of the population

The age structure of a population can be inspected by assuming stability or, more generally, by modelling the population using variable growth rates as discussed in Preston and Coale (1982). In this setting, the rate of change of one cohort relative to another at age x and time t is defined as $\frac{1}{N(x)}\frac{dN(x)}{dx} = -u(x) - r(x)$ where

$r(x) = \lim_{\Delta t \to 0} \frac{N(x,t+\Delta t) - N(x,t)}{N(x,t)\Delta t}$ (Preston and Coale 1982). The number alive in this population at time t at age x can be expressed as $B_o e^{\int_0^x -r(a)da} l(x)$. The proportion of this population at ages x can then be expressed as

$c(x) = \frac{e^{-\int_0^x r(a)da} l(x)}{\int_0^\omega e^{-\int_0^x r(a)da} l(x)dx} = be^{-\int_0^x r(a)da} l(x)$, where $b = \frac{1}{\int_0^\omega e^{-\int_0^x r(a)da} l(x)dx}$, $l(x)$ is the current period life table function (which is assumed to be known) and ω is the oldest age in the population. Bhat (1995) discretised the above formula as $c(x) = \frac{e^{-5\sum_{i=0}^{x-5} {}_5r_i - 2.5\,{}_5r_x} l(x)}{\sum_{x=0}^{\omega} e^{-5\sum_{i=0}^{x-5} {}_5r_i - 2.5\,{}_5r_x} l(x)}$, where ${}_5r_x$ is the growth rate in the five year interval beginning at age x. Once the implied age distribution has been calculated, the enumerated population can be redistributed in proportion to the implied age distribution of the population. Bennett and Horiuchi (1984: footnote 7) suggest a similar method for correcting the age distribution of the death data by using the formula for the age distribution of deaths given in Preston and Coale (1982), namely $\frac{D(x)}{\int_0^\omega D(x)dx} = \frac{d(x)e^{-\int_0^x r(a)da}}{\int_0^\omega d(x)e^{-\int_0^x r(a)da} dx}$, where $d(x)$ is the number of deaths at age x from the life table with a radix of one. Similarly to the procedure in Dechter and Preston (1990), a transition matrix of age misreporting can be estimated from the reported and corrected age distributions. An iterative algorithm for this procedure is given by Bhat (1990).

Besides assuming that the period life table is known, another difficulty in applying these models is that accurate estimates of the age-specific growth rates are required. Estimating the age-specific growth rates from the population data means that distortions in the population data will be reproduced in the estimated age distribution. If the extent of misreporting of age has not changed between censuses, the estimated growth rates will not be affected (Bhat 1990; Preston, Elo, Rosenwaike *et al.* 1996: 204), but this is difficult to ascertain. Bhat (1990: 152) notes, however, that since the growth rates are used in a cumulative form in the above equations, errors in the growth rates are "not as significant as they may seem".

Feeney (1990) proposed a graphical technique for inspecting the age distribution of the population, based on the assumption of a stable population. Since the age distribution of a population obscures the oldest ages due to the scale of numbers of the younger and older ages, Feeney proposes that the age distribution be 'untilted' by transforming the population data by a factor $\frac{{}_5N_x e^{rx}}{{}_{x+2.5}p_0}$, where r is the growth rate of the population over the past 80-100 years and ${}_{x+2.5}p_0$ is the proportion of survivors from birth to age $x+2.5$ (on the assumed mortality schedule), to remove the distorting effect of population growth and mortality from the data. Age exaggeration can be identified in the untilted age distribution by looking for data points that are significantly higher than the rest of the untilted data.

2.5.3 Techniques for correcting age heaping

Moving averages are a simple method for correcting data for age heaping. Moving average schemes are discussed in Feeney (1979) and Saxena and Gogte (1985).

Heitjan and Rubin (1990) apply multiple imputation (Rubin 1976) to the problem of age heaping. Their method assumes that a series of rules that define the way in which data has become heaped can be specified. For example, in the data in Heitjan and Rubin (1990), it is assumed that age heaping is a result of rounding, and therefore the interval specified for a child reported to be six months of age is the range three to nine months. Uniform distributions are assumed over the ranges defined in the rules, and the data are resampled a number of times from these ranges. The resampled datasets are then averaged to produce a final dataset on which inference can be conducted or alternatively, inference can be performed on each dataset separately and the results aggregated to produce a distribution of the inferred quantity.

Camarda, Eilers and Gampe (2008) used a Composite Link Model (CLM) to smooth heaped numerical data. The method assumes that unobserved smooth data has been heaped by a transition matrix, and both the smooth data and the transition matrix are estimated statistically using the CLM. The smoothness of the age distribution is estimated statistically and therefore not much judgement is required when using this model, however, the corrected data are not guaranteed to be demographically reasonable.

The techniques described in the preceding paragraphs do not appear to be useful for this research, since they focus on age heaping, whereas the South African death data suffer from year of birth heaping (Machemedze 2009) as well as age heaping. Smoothing these data would require a complicated moving average scheme to average the counts between and within cohorts. Similarly, the CLM model would need to be extended substantially to smooth the counts between and within cohorts. Finally, since the cause of the heaping in the data is not well understood, it is not possible to specify accurately the rules defining the heaping, to apply the multiple imputation method.

2.6 The question of exponentially increasing mortality

The earliest mortality law[15] still in common use is the Gompertz (1825) law, which describes the force of mortality as an exponential function of age, $\mu_x = e^{a+bx}$ where μ_x is the force of mortality (or hazard rate) at age x. Gompertz's law was modified by Makeham (1860, 1867) to include a term for background mortality that does not depend on age, i.e. $\mu_x = e^{a+bx} + c$. Both of these laws assume that mortality rates increases exponentially with age.

Perks (1932) and Beard (1971) argue that the exponentially increasing mortality rates produced by these laws rise too quickly relative to empirically observed mortality rates at the older ages, resulting in mortality rates that are too high[16]. A similar observation is in Thatcher (1992), who investigated mortality rates over the period 1911-1990 in England and Wales and found that the force of mortality plotted on a log scale did not lie on a straight line, as predicted by the Gompertz law, but appeared rather to "bend downwards" (Thatcher 1992: 423), indicating that mortality does not increase exponentially. In a broader study with similar conclusions, Thatcher, Kannisto and Vaupel (1998) used the high quality data from thirteen countries contained in the Kannisto-Thatcher database to test empirically several mortality laws, using extinct generations estimates of mortality, with survivors in more recent periods estimated through the Survivor Ratio method. They showed that the laws which predict exponentially increasing mortality result in rates that are too high at the older ages. Supporting these conclusions is Thatcher (1999), who similarly found that the logistic model (i.e. a mortality law that does not assume exponentially increasing mortality) fits a variety of historical datasets (from the early medieval period to the English Life Table 14) well, with a low average relative error, despite the very different periods considered in the study.

[15] Parametric representations of mortality rates as a function of age are often referred to as mortality laws in the literature, in the sense of "an observed regularity or pattern" (Thatcher 1990: 135) and "a fundamental law of nature" (Kirkwood 2015: 4) and this review will follow the convention.

[16] Kirkwood (2015) proposes that Gompertz was aware of this. Similarly, Olshansky and Carnes (1997) quote Gompertz (1871) as saying that his law was only applicable from age 20 to age 60 and they conclude that the Gompertz law was never intended to be used at the oldest ages, thus dismissing the conclusion of other researchers who rejected the Gompertz law after finding examples of mortality rates that do not rise exponentially in old age.

On the other hand, a study of mortality in the United States by Gavrilov and Gavrilova (2011) fitted the Gompertz law to mortality data for ten cohorts born in the period 1886-1895 on the age range 88-106. They found that the Gompertz law outperformed the logistic model in eight of the ten cohorts considered. The findings of studies which reject the Gompertz law at advanced ages were attributed by Gavrilov and Gavrilova to poor data or to the statistical methods used to estimate mortality. In particular, they speculate that the results of Thatcher, Kannisto and Vaupel (1998) was due to age exaggeration in the data used in that study. However, given the extensive scrutiny applied to the data in Thatcher et al's study, it seems hasty to dismiss their results as an artefact of imperfect data. Therefore, the balance of the research suggests that mortality for developed countries with accurate data should be modelled assuming that mortality does not increase exponentially with age.

Several mortality laws which do not assume that mortality increases exponentially have been proposed. Perks suggested embedding the Makeham formula within a logistic function, giving a function of the form $\mu_x = \frac{ae^{bx}}{1+ce^{bx}} + d$. At older ages the force of mortality increases at a decreasing rate, eventually reaching an asymptote (which is never observed in practice). Setting c equal to a in the previous equation (and dropping the background mortality term, which is of little consequence at older ages (Beard 1961)) reduces the number of parameters to be estimated without substantially worsening the fit and produces the Kannisto law (Thatcher, Kannisto and Vaupel 1998), namely, $\mu_x = \frac{ae^{bx}}{1+ae^{bx}}$. Coale and Kisker (1990) proposed a method of extrapolating the central rate of mortality at age 85, m_{85}, to the age of 120 in order to close out a life table when the data are of dubious quality at older ages. The method extrapolates the observed mortality rate using the rate of change of mortality, k_x, such that $m_x = m_{84} e^{\sum_{i=85}^{x} k_i}$. In the Coale and Kisker method, instead of assuming that central rates of mortality increase at a constant rate with age (as is the case for Gompertz law), rates are assumed to increase at a decreasing rate. On the assumptions that k_x changes linearly after the age of 85, $k_x = k_{85} + s(x-85)$, and that the force of mortality at age 110 is fixed (at 1 for males and 0.8 for females), the slope of the previous equation, s, can be solved as $s = -(\log(\frac{m_{84}}{m_{110}} + 26k_{85})/325$.

2.6.1 Exponentially increasing mortality in developing countries

A firm conclusion on whether mortality increases exponentially with age in developing countries cannot yet be drawn since high-quality mortality data are generally not available for the majority of developing countries, which has prevented detailed studies of old age mortality rates in these countries. Since population and death data in these countries often suffer from age exaggeration, which leads to mortality rates that appear to decelerate with age, mortality models which allow for a deceleration in mortality may erroneously appear to fit the data better than models which assume accelerating mortality. It could, on the one hand, be hypothesised that old age mortality increases exponentially in developing countries since older people in these countries are more likely to live in rural areas compared to developing countries and may not have access to quality healthcare services and the medical advances promoting low mortality in developed countries. On the other hand, an alternative hypothesis, discussed in detail and ultimately rejected in Coale and Kisker (1986), is that the higher level of mortality in developing countries will have a selective effect on the population surviving to old ages, which may produce rates that increase at a decreasing rate.

While a firm conclusion cannot yet be drawn, Chinese mortality data offer some hints that mortality decelerates at the oldest ages even in developing countries. These mortality data do not suffer from age exaggeration to the same extent as other developing countries because accurate dates of birth are recorded due to the importance of astrology in Asian culture, as discussed previously. For example, Coale and Li (1991) fit the Coale and Kisker model to Chinese mortality data in the period 1980-1982. Data known to be biased by age misreporting of certain minority populations were excluded (from the census counts and presumably from the death data), allowing accurate calculation of mortality rates directly from the data. The Coale and Kisker model, fit to the male mortality rates for ages 80+, was shown to produce exceptionally accurate predictions of the calculated mortality rates, implying that the Chinese population experiences mortality deceleration at older ages. A further example is Yi and Vaupel (2006), who studied Chinese mortality based on the 1990 census. Various models (including the Gompertz and the Kannisto models) were fitted to the Chinese data, for both sexes, for ages 80-96 (with the latter age chosen because of increasing fluctuations in the data from that point) and mortality rates to age 106 were predicted and compared with the observed rates. All of the models fit well until the age of 96. The predicted values from the Kannisto model were the closest to the observed values after the age of 96 while the values from the Gompertz model were too high compared to the observed rates. Re-fitting the Kannisto model to the data from ages 80-105 produced almost identical curves to the fit at ages 80-96, lending support to the contention the Chinese population experiences mortality deceleration at older ages.

More evidence is in Thatcher (1999), who estimated the parameters of the logistic model for a number of historical populations (using life tables from the Roman era, medieval Hungary, Renaissance England, Halley's life table and the English Life Tables 1 and 14) and found that the logistic model fit historical mortality rates (including medieval populations) well over the age range 30-100. The parameter defining the slope of the force of mortality was found to have varied only slightly (with relatively small standard errors), rising from 0.077 in medieval Hungary to around 0.11 in modern England and Wales. This implies that the logistic model may be appropriate for populations with very different levels of mortality (and that the rate of increase in mortality at old ages is not necessarily linked to the level of mortality).

Therefore, when modelling mortality in developing countries, it appears appropriate to allow for the possibility of mortality rates that do not increase exponentially with age, subject to an adequate validation of the quality of age reporting.

2.7 Estimates of South African old age population and mortality

This section reviews estimates of the South African population and mortality at the older ages, and specifically focuses on an application of the method of extinct generations to South African data (Machemedze 2009).

2.7.1 Machemedze's study

A study of South African old age mortality was conducted by Machemedze (2009), and further results of this study were reported in Machemedze and Dorrington (2011). The method of extinct generations was applied to re-estimate the population and derive mortality rates using the death data from 1996-2006. These estimates were compared to the population enumerated in the censuses of 1996 and 2001, and the population estimates produced in the Community Survey of 2007.

A modified Das Gupta method[17] was applied to the reported death data to produce population estimates at the ages 75-110[18], which were then corrected for the effect of incomplete death registration by dividing the estimates by the ratio of the

[17] Similar to the method proposed by Meslé and Vallin (2002).

[18] A different method were applied to re-estimate the population at the date of the community survey since no death data for 2007 were available when the study was performed.

NEG population estimates to the enumerated population at ages 75+ (on the assumption that this ratio represented the extent of the incompleteness of registration for all ages older than 75). It was found that the death data were around 91 per cent complete for males in both censuses and for females 95 per cent complete at the time of the census of 1996 and 92 per cent complete at the time of the census of 2001.

Age exaggeration in the population data was assessed by comparing the enumerated populations with population projections, which indicated that age exaggeration began from age 75. The NEG estimates were higher at younger age groups (from 75-84) than the enumerated populations and lower at the older age groups, (from 85+). However, it was observed that in both censuses and for both sexes, too few had been counted at ages 97-99, after which too many had been enumerated. The results of the same exercise for the Community Survey indicated age exaggeration from ages 95+. The NEG estimates were somewhat smoother than the census data, which is likely due to the approximate year of birth cohorts used in the study, which had to be derived by averaging deaths in consecutive ages and years. Although age heaping was not identified, year of birth heaping was identified in both the population and the death data

Crude mortality rates were calculated for the periods 1996-2001 and 2001-2007 by dividing the number of deaths occurring in these periods by the number of person-years lived (Machemedze and Dorrington 2011). It was found that old age mortality was underestimated when calculated using the population data as the denominator, as compared to the rates using the NEG estimates as the denominator and that mortality was more underestimated in the period 2001-2007 than in the period 1996-2001, indicating worsening age exaggeration in the population data over time. The ungraduated rates increased monotonically with age, but the rate of increase of the raw mortality rates with age appeared to lessen significantly with age, especially in the period 1996-2001, leading to a levelling off of the rates at advanced ages. The rates were compared to those in Dorrington, Moultrie and Timæus (2004) for the period 1996-2001, as well as to those underlying the United Nations and US Census Bureau projections for both periods and it was found that the raw rates were lower than the rates from the other three sources.

2.7.1.1 Conclusions and limitations

Machemedze showed that South African population and death data at the older ages suffer from year of birth heaping and age exaggeration. The effect of heaping on the performance of the NEG methods was not assessed in Machemedze's study and it is shown later that mortality rates produced by applying the NEG methods to death data suffering from age heaping will be biased. Thus, the NEG methods need to be modified to address these problems in the death data.

Incomplete reporting of deaths was addressed by constraining the NEG estimates to equal the aggregate enumerated population at ages 75+. Since the unconstrained NEG methods tend to underestimate the population (Andreev 2004; Thatcher, Kannisto and Andreev 2002), it is not clear if the adjustment for incompleteness of reporting derived by Machemedze is, in fact, an accurate estimate of the incompleteness of reporting of deaths. Another source of possible inaccuracy is the changing completeness of reporting with time which may bias the NEG methods as discussed in section 2.4.2.1.

Two limitations of the study mentioned by Machemedze are that race was not used since it was missing on a large number of the death certificates and the accuracy of the NEG estimates was constrained by the lack of death data classified by both year of death and year of birth, which were approximated before the NEG methods were applied.

The present study aims to address these limitations. The NEG methods applied in this study were modified to quantify the effect of age heaping explicitly and produce unbiased estimates of mortality, and, before the NEG methods were applied, the death data were corrected for incompleteness of reporting using the DDMs. The DDMs were also used to correct the death data for incompleteness of reporting of race. Finally, more detailed death data classified by both year of death and year of birth were available for this study.

2.7.2 Estimates of South African mortality

Life expectancy at age 70 is used in this section to give an indication of the level and trend in old age mortality, since data on life expectancy in South Africa covering most of the 20th century are available (whereas life tables are not available in some periods).

One of the earliest available estimates is the life expectancy of White males in Johannesburg in 1910, included in the abridged life tables of Van Tonder and Van Eeden (1975), which also include abridged versions of the South African Life Tables (SALT) covering the period 1921-1970. Although Van Tonder and Van Eeden (1975) show that their abridged life tables produce similar life expectancies at birth to the official SALT, they do not consider the accuracy of the mortality at advanced ages. Another early study is Sadie (1951) which presents abridged life tables for various population groups in 1946. A later study by Sadie (1988) produced estimates and projections of life expectancy covering the period 1938 to 2002, for both sexes and the four population groups.

Since African mortality was not estimated in the SALTs, an important source is the detailed study of the mortality of the African population in the period 1984-1986 performed by Dorrington (1998), who also produced national life tables for the periods 1984-1986 and 1989-1991.

The latest published study examining the mortality of the population groups is Dorrington, Moultrie and Timæus (2004), in which mortality at the provincial level and for each of the population groups using population data from the censuses of 1996 and 2001. Other[19] estimates of South African mortality are those produced by the United Nations (2013) (UNPD[20]) and the AIDS model of the Actuarial Society of South Africa (ASSA) (Actuarial Society of South Africa 2009). Where mortality rates were only available in quinquennial age groups or with a low open age interval, for the comparisons below, the rates were interpolated and extrapolated to the age of 120, using a Gompertz curve. The life expectancies thus produced may underestimate the true life expectancy if the mortality rates of the population increase at a lower rate than predicted by the Gompertz curve. On the other hand, if the mortality rates are affected by age exaggeration, then the life expectancies will be overstated. However, neither effect is likely to be significant at age 70.

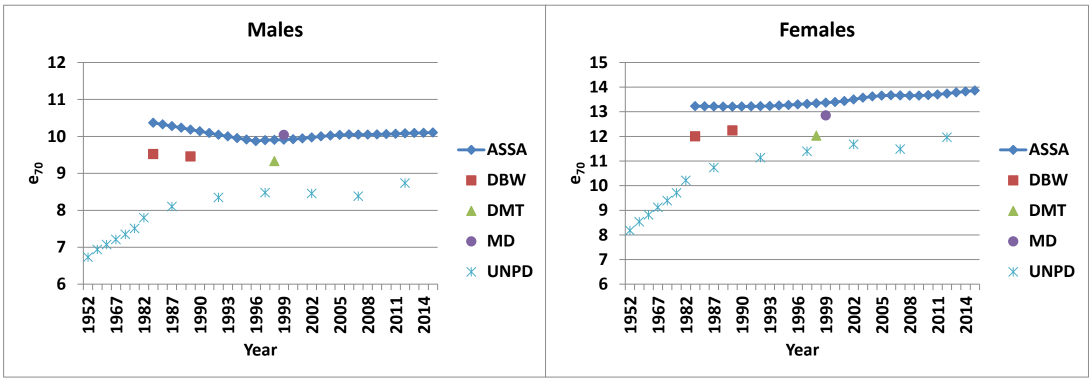
National life expectancy for both sexes at age 70 is shown in Figure 2.1. Most striking is the lack of consistency between the ASSA and UNPD estimates, which differ by approximately 1.5 years since 1987. The other estimates of Dorrington, Bradshaw and Wegner (1999), Dorrington, Moultrie and Timæus (2004) and Machemedze and Dorrington (2011) are roughly consistent with each other. An upwards trend since 1985 can be discerned for females, while for males life expectancy appears to have remained constant over the period.

---

[19] Sources which do not include e_{70} were not included in the review.

[20] Although the 2012 revision of the UN World Population Prospects is somewhat out of date, having been updated twice since, by the 2015 and 2017 revisions (United Nations 2015, 2017), inspection suggests that the comparisons remain largely the same.

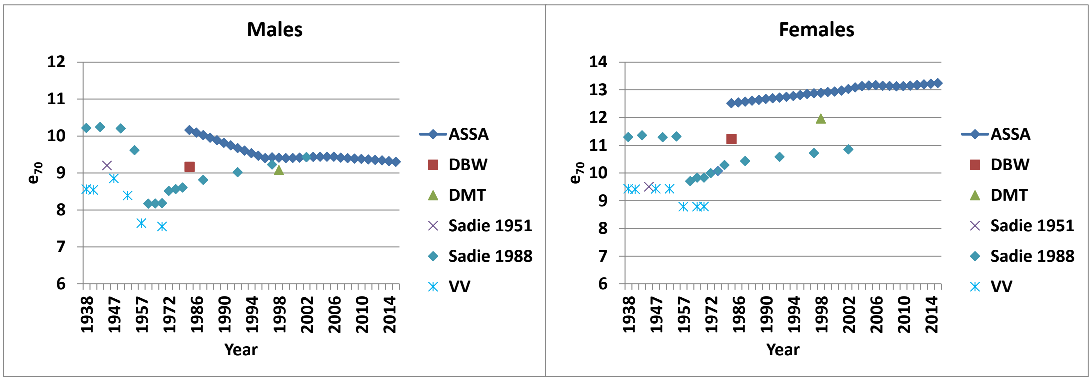
Figure 2.1: Estimates of e_{70}, National



ASSA - Actuarial Society of South Africa, UNPD – United Nations Population Division, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), MD - Machemedze and Dorrington (2011)

Figures 2.2 to 2.5 show estimates of life expectancy for each of the population groups. The ASSA model predicts the highest life expectancy for all population groups and sexes, in almost all of the years considered. For Africans, the estimates of Sadie (1988), Dorrington, Bradshaw and Wegner (1999) and Dorrington, Moultrie and Timæus (2004) are somewhat consistent with each other, while the estimates of Van Tonder and Van Eeden (1975) in the years since 1950 appear too low. The ASSA model and the projections of Sadie appear to assume that mortality at the older ages is improving slowly for both sexes and all population groups (except for the ASSA model estimates for African males).

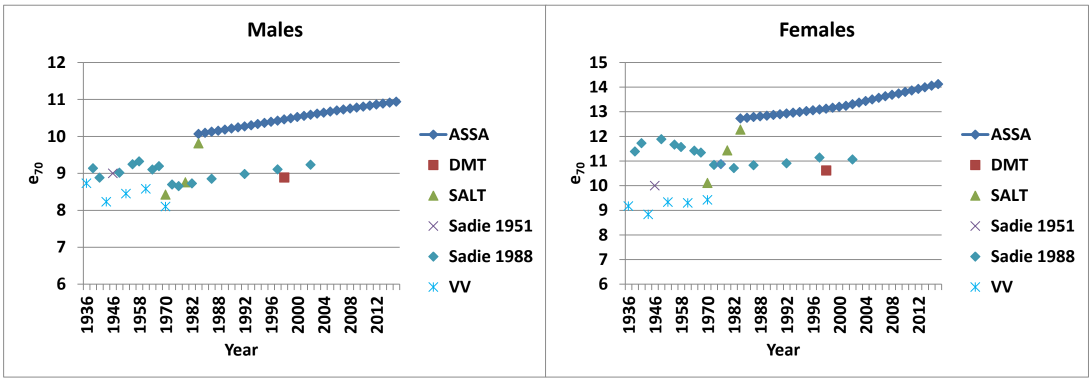
Figure 2.2: Estimates of e_{70}, African



ASSA - Actuarial Society of South Africa, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

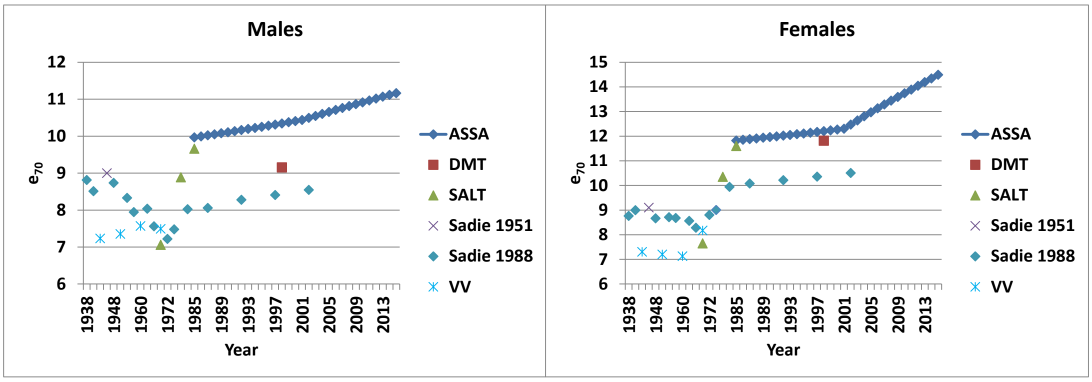
The sharp upwards trend in life expectancy for Coloureds, Indians and Whites implied by the SALT over the period 1970-1985 is surprisingly inconsistent with the estimates of Sadie (1951) and Van Tonder and Van Eeden (1975), which imply that life expectancy changed only marginally over the extended period 1921-1970, as well as with the estimates of Sadie (1988) and Dorrington, Moultrie and Timæus (2004), which are in almost all cases lower than the estimates in the SALT, while the ASSA estimates are somewhat higher.

Figure 2.3: Estimates of e_{70}, Coloured



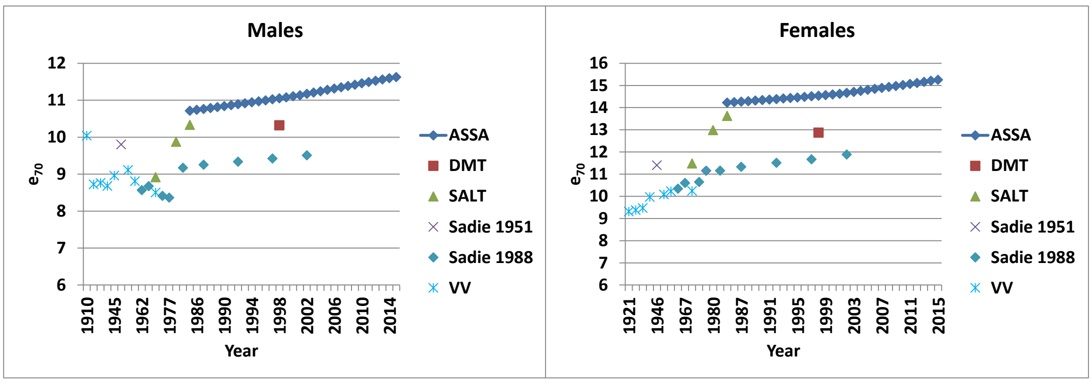
ASSA - Actuarial Society of South Africa, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), MD - Machemedze and Dorrington (2011), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

Figure 2.4: Estimates of e_{70}, Indian



ASSA - Actuarial Society of South Africa, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

Figure 2.5: Estimates of e_{70}, White



ASSA - Actuarial Society of South Africa, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

In summary, the level of life expectancy at age 70 over the past century and even during the more recent years is subject to considerable uncertainty as shown by the inconsistency between the different sources, as well as the striking inconsistency between the official life tables and later estimates.

2.7.3 Estimates of the population at old ages

Recent censuses of the South African population have been conducted in 1996, 2001 and 2011 and in lieu of a census in 2007, a large-scale survey was performed. These recent censuses covered the entire population of the Republic of South Africa. Before these, censuses were performed in 1985 and 1991 that excluded the so-called "homelands" of Transkei, Bophuthatswana, Venda and Ciskei. Population numbers in these years were collated from various sources as described in Section 3.2.1. Earlier censuses have not been considered, since this research focuses on producing estimates the population from 1985 using the NEG methods.

Official mid-year population estimates are produced by Statistics South Africa (Stats SA) for the population as a whole and by population group (Stats SA 2015) on an annual basis. Noting the disagreement between the mid-year population estimates and the results of the census conducted in 2011, in particular for ages younger than 40, Dorrington (2013a) produced an alternative set of mid-year estimates (AltMYE) with an age distribution consistent with the 2011 census. A back projection of the 2011 census population aged 70 and older to the date of the 2001 census revealed a significant over count of the elderly population in 2011 census, possibly due to age exaggeration. The age exaggeration was corrected in the AltMYE by projecting the 2001 census population forward to 2011, and redistributing those enumerated at ages 70 and older in 2011 in proportion to the projected population. A close correspondence between the AltMYE and the census 2001 age distribution was found for the national estimates, and for Africans, Coloureds and Indians.

Other estimates of the South African population include those produced by the ASSA AIDS model (Actuarial Society of South Africa 2009), the United Nations[21] (United Nations 2013) and the United States Census Bureau (US Census Bureau 2015) (USCB).

Figure 2.6 shows estimates of the South African population aged 70+ for males

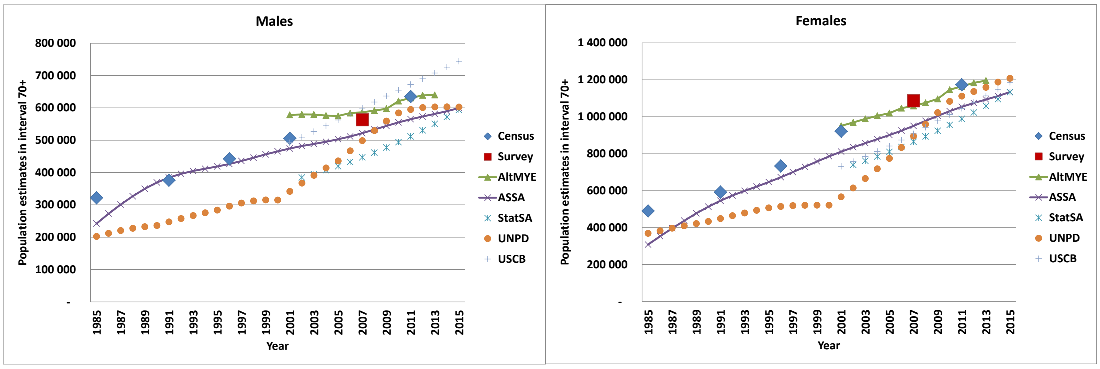
---

[21] Although the 2012 revision of the UN World Population Prospects is somewhat out of date, having been updated twice since, by the 2015 and 2017 revisions (United Nations 2015, 2017), inspection suggests that the comparisons remain largely the same.

and females. All sources agree that the population aged 70+ has increased since 1985 but none of the estimates agree well with all of the censuses and in the recent years, there is considerable disagreement amongst the estimates regarding the size of the population aged 70+. The AltMYE estimates are by design highly consistent with the census 2011 data for those aged 70+ and remain consistent with the census 2001 data for females, but not for males. The UNPD estimates proceed in a relatively smooth fashion but experience implausibly rapid growth after the year 2000 and a subsequent drop off, which are features that are not consistent with the estimates from other sources. The ASSA estimates are highly consistent with the census data of 1996 but are lower than the community survey and census 2011 data for both sexes. The mid-year estimates of Stats SA appear to underestimate the population for both sexes.

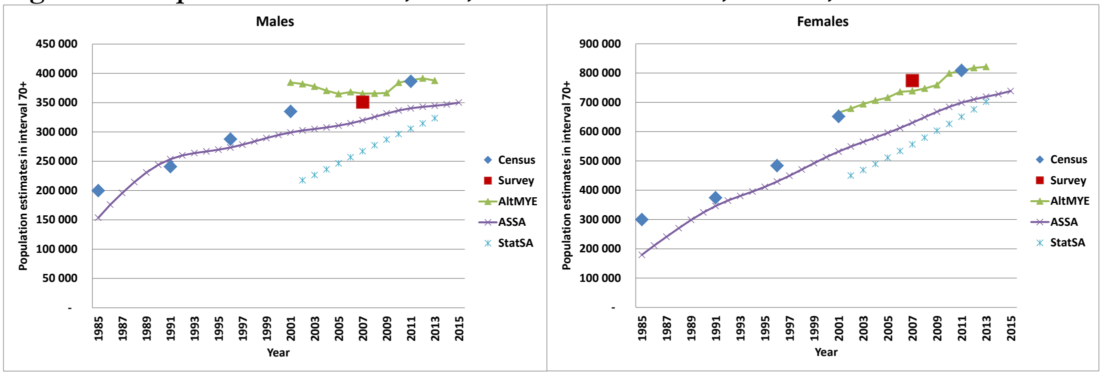
Figures 2.7 to 2.10 show estimates of the population aged 70+ for the population groups. All sets of estimates point to the growth of the elderly population for both sexes. There is a general disagreement of the estimates for Africans, Coloured Females and White Males. The AltMYE are consistent with the census 2001 and 2011 numbers except for Whites and African males, which may indicate that 2011 census over-counted those aged 70+ in these specific groups, relative to the 2001 census. The Stats SA estimates are significantly too low, except for Indians.

Figure 2.6: Population estimates, 70+, Males and Females, National, 1985-2015



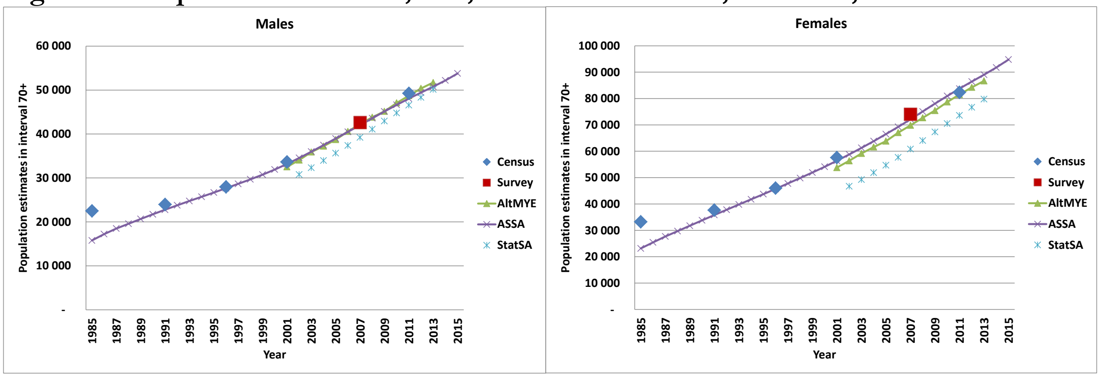
AltMYE – Alternative mid-year estimates, Dorrington (2013); ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015); UNPD – United Nations Population Division (2013); USCB – United States Census Bureau (2015)

Figure 2.7: Population estimates, 70+, Males and Females, African, 1985-2015



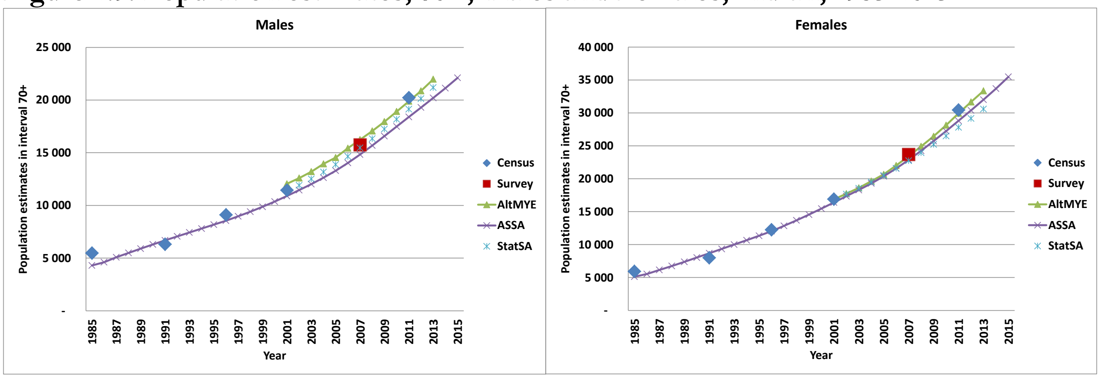
AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Figure 2.8: Population estimates, 70+, Males and Females, Coloured, 1985-2015



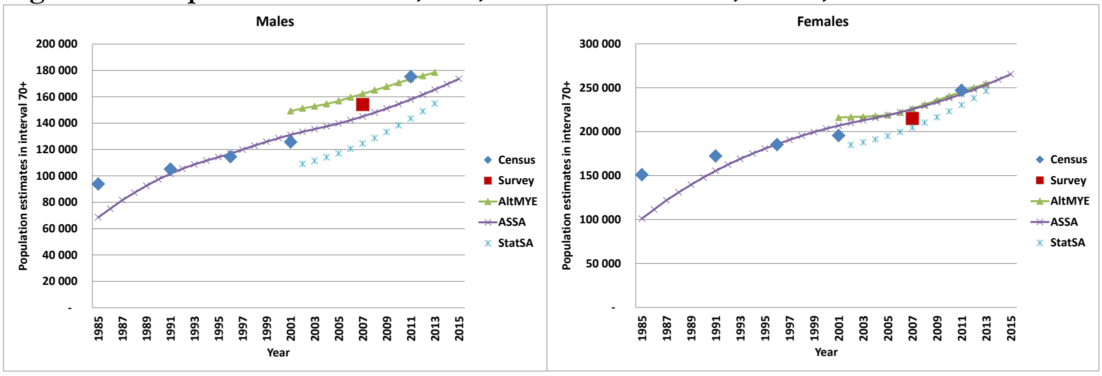
AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Figure 2.9: Population estimates, 70+, Males and Females, Indian, 1985-2015



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Figure 2.10: Population estimates, 70+, Males and Females, White, 1985-2015



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Of the estimates considered in this section, the AltMYE are the most consistent with both the 2001 and 2011 censuses, and should, therefore, be the preferred set of population estimates. To the extent that the 2011 census counts for Whites of both sexes and African males are found to be inaccurate, the AltMYE estimates are also inaccurate, however, no better alternative set of estimates for the population groups is currently available.

2.8 Conclusion

This chapter has examined the evidence of misreporting in population and death data and the techniques used to correct these data when suffering from misreporting.

Since death data are not completely reported in South Africa (Dorrington, Moultrie and Timæus 2004; Stats SA 2014), it is necessary to correct the death data before applying the NEG methods (or risk biased results, as discussed in Section 2.4.2.1). This review has identified that the SEG+delta method represents a best-practice technique for correcting death data for incomplete reporting. The "shortcut" technique of Preston and Lahiri (1991), which is potentially more robust to incompleteness of reporting that varies with age, has not previously been applied to South African data and this research will apply the technique and assess its results compared to the SEG+delta method.

The South African death data suffer from age and year of birth heaping (Bradshaw, Schneider, Laubscher *et al.* 2002; Machemedze 2009), but whether the NEG methods will produce unbiased population estimates and mortality rates despite the heaping has not yet been investigated in the literature. This research aims to investigate the effect of heaping on the results of the NEG methods and whether the methods can be adapted to produce unbiased results in the presence of heaping.

As discussed in Section 2.7.1.1, Machemedze's study did not apply the DDMs to the death data before applying the NEG methods, which may produce biased results if completeness of reporting is changing over time. This study will apply the DDMs before the NEG methods and, thus, avoid this source of potential bias in the results. Furthermore, the availability of death data containing year of birth information and covering more years allows for a more complete investigation than previously possible., as will applying the methods for each of the population groups (African, Coloured, Indian and White) separately.

The review of estimates of the population aged 70+ and life expectancy at age 70 has shown that there is major disagreement between different sources as to the level and progression over time of the population and life expectancy at the older ages, leading to significant uncertainty regarding the most accurate set of estimates. This research aims to reduce the level of uncertainty by applying best practice techniques adapted to the South African data to produce new estimates of the population and life expectancy.

Finally, the South African data, if not adversely affected by age exaggeration, may provide insight into the question of whether the rate of increase of mortality rates with age declines at the oldest ages in developing countries, as appears to be the case in developed countries with high quality data.

3 DEATH DISTRIBUTION METHODS

3.1 Introduction

Estimates of the completeness of reporting of deaths and of population group are derived in this chapter using the Death Distribution Methods (DDMs). The chapter is divided into two main parts – the first part discusses the methodology that was applied to derive the estimates and the second part presents the results.

An outline of the steps followed to produce estimates of completeness of reporting is as follows:

1. Death and population data covering the period 1984-2013 were collated from various sources.
2. Estimates of completeness in the intercensal periods 1985-1991, 1991-1996, 1996-2001 and 2001-2011 were derived using the SEG+delta method.
3. The estimate of national completeness of reporting was set to the weighted average of the estimated completeness for the population groups, for the periods 1996-2001 and 2001-2011.
4. The estimates of completeness were validated by considering if the mortality rates implied by the estimates are reasonable, by comparing the estimates to those produced using Preston and Lahiri's "shortcut" method and by comparing estimates of ${}_{15}q_{50}$ to those produced using orphanhood data.
5. Curves were fit through the intercensal estimates to produce annual estimates of completeness relative to the population in 2011.
6. The annual estimates were adjusted to take into account under-reporting of recent deaths due to late reported deaths, a fall in the level of reporting compared to the population register and an increase in the reporting of population group.

The chapter concludes by comparing the estimates of completeness to those produced in other studies.

3.2 Methodology

3.2.1 Data

Data containing records on all registered deaths for the years 1984-2013 was provided by Statistics South Africa (Stats SA) to the South African Medical Research Council

(SAMRC), which in turn provided the data for this research[22]. The deaths in this dataset are split by sex, year of death and age last birthday at death, as well as by population group for the years for which this was available (1984-1990 and 1999-2013). Records that were missing sex or age were excluded from further analysis since the proportion of these records was small (0.00072 per cent for 1984-1996 and 0.8 per cent for the period 1997-2013) and the application of DDMs to the data excluding these records corrects for the missing records, on the assumption that the absence of sex or age is independent of age.

To apply the DDMs at a national level, the record level data were aggregated by sex, year of death and age last birthday at death. Where possible, the data were also aggregated by sex, year of death, age last birthday at death and population group, to apply the DDMs for each population group. Deaths were apportioned into the intercensal periods of 1985-1991, 1991-1996, 1996-2001 and 2001-2011, on the assumption that deaths occurred uniformly within each calendar year into which they were apportioned.

Over the period 1999-2013, 23.8 per cent of records were missing information on the population group of the deceased (the four population groups currently used by Stats SA are Africans, Coloureds, Indians and Whites). Instead of imputing the population group, the DDMs were used to estimate the proportion of deaths by population group that were registered with population group recorded (i.e. the death is both registered and has the population group of the deceased recorded) and correcting reported deaths for under-reporting corrects for both under-reporting of deaths and under-reporting of the population group of the deceased. This is based on the assumption that reporting of deaths and population group is constant with respect to age; the latter assumption being somewhat validated by the data which show that the proportion of those missing race is almost constant with age.

Population group information was missing from nearly all of the records in the years 1991-1998. To apply the DDMs in the intercensal period 1996-2001[23], an estimate was made of the number of deaths that would have been registered in the period 1996-1998 had registration and identification of population group been at the same level of

[22] The dataset was specially provided and contains data that are not publicly available – deaths occurring in the years 1984-1995 and population group information covering the years 1984-2013.

completeness in the period 1999-2001. The deaths were estimated by projecting backwards the average deaths in each population group recorded over the period 1999-2001 using the population growth rates, for each population group, implied by the population counted in the 1996 and 2001 censuses.

For this study, population data were collated from several sources. For 1985, the census numbers for Coloureds, Indians and Whites were those made available online by Stats SA[24]. The numbers for Africans, including those living in the so-called "homelands" (Transkei, Bophuthatswana, Venda and Ciskei or "TBVC") were taken from Dorrington (1998: Appendix 1). For 1991, the RSA census numbers were taken from the report on the 1991 census by the Central Statistical Service (1992a: page 119) and combined with the estimates of the population of the TBVC countries as estimated by Jan Sadie (Central Statistical Service 1992b: page 22). For ease of reference, these data are provided in Appendix B.1. The numbers by sex, population group and five year age group enumerated in the 1996, 2001 and 2011 censuses (full census, not 10 per cent sample) were downloaded from the Stats SA website[25].

Of the total census count in 1996, 1.2 per cent were recorded as being of unknown age, and 0.9 per cent were recorded as being of unknown population group. To estimate the growth rates of the population as accurately as possible, the data missing population group were therefore allocated in proportion to the numbers recorded by population group, for each five-year age band separately and similarly, the data missing age information were allocated in proportion to the numbers recorded in each age band, for each population group separately.

Although the 2007 Community Survey provides data that could be used in the DDMs, Dorrington (2013c) notes that data from small surveys rarely produce satisfactory results. An initial application of the DDMs to the periods 2001-2007 and 2007-2011 produced implausible results in comparison to the results from the period 2001-2011, mainly because of distortions in the age structure of the older population in

[23] The DDMs could have been applied to the reported data (i.e. without projecting the deaths backwards). However, the resulting estimate of completeness would have been too low for the deaths reported in 1999-2001, although it would have been correct for the period 1996-2001 on average. Since we wish to produce annual estimates of completeness of reporting, the deaths were projected backwards to avoid this issue.

[24] http://interactive.statssa.gov.za:8282/webview/

[25] http://interactive.statssa.gov.za/superweb/login.do

the 2007 numbers. The data from the Community Survey were therefore not used in producing the estimates of completeness by single years[26].

Migration data for the single years 1985-1996 were taken from the Actuarial Society of South Africa (ASSA) AIDS model (Actuarial Society of South Africa 2009) and aggregated into the periods 1985-1991 and 1991-1996.

Migration data for the intercensal periods in 1996-2011 were provided by Professor Dorrington (personal communication, 29 June 2016). Although there is some uncertainty about the migration figures, it is unlikely that any errors would impact the results significantly (since migration is small relative to the population and the reported deaths).

3.2.2 Death distribution methods

On the basis of the discussion in section 2.4.4, the SEG+delta method is used in this study as the primary method, with the results of the GGB method (which are reported in Appendix B.2) used mainly as a check on the reasonableness of the results of the SEG method.

The DDMs were applied using the Excel workbook templates accompanying Dorrington (2013b, 2013c). The workbooks calculate life expectancy in the open interval for use in the SEG method using mortality rates derived from death data corrected for incompleteness of reporting and graduated using Brass' logit relational model. Various standards, including the model life tables for a population experiencing an AIDS epidemic from Timæus (2004), were tested for the graduation but achieving a good fit was problematic because of the pattern of mortality at the younger ages due to AIDS and also since the mortality rates calculated from the corrected data appeared to be too low at the older ages in comparison to the life tables, possibly as a result of age exaggeration in the data. The life tables produced by the ASSA model, taken at the mid-point of each intercensal period, provided the closest fit at the younger ages and were therefore used as the standard. The parameters for the Brass logit relational model were determined using data in the age range 45-84, providing enough data points for the graduation while hopefully avoiding the worst of the age exaggeration in the data.

Some judgement is necessary when choosing the age ranges to be used to determine completeness. The GGB method was fit to data in the age range 5-84, unless

[26] Since mortality changed significantly in the period 2001-2011, the estimates of completeness would perhaps have been slightly more accurate had data been available to split the period into five year intervals; however it is unlikely that this would have resulted in significantly different estimates since completeness for the period is quite high.

both a poor fit was observed in the diagnostic plots and changing the age range produced a significant change in the estimated completeness, in which case the age range was narrowed. When applying the SEG method, delta was estimated using the age range 5-84, unless the resulting estimates of completeness were inconsistent with those produced by the GGB method, in which case narrower age intervals were used if they improved the match. The SEG diagnostics were inspected for indications of differential completeness of registration of deaths with age and completeness was calculated excluding those age groups showing a departure from the average level of completeness. The estimates of completeness at ages 25-64 were generally used for this calculation.

The DDMs were applied at a national level in the four intercensal periods 1985-1991, 1991-1996, 1996-2001 and 2001-2011. The DDMs were applied to population group data in the periods 1996-2001 and 2001-2011[27]. Deaths for Coloureds, Indians and Whites were assumed to be completely reported in the period 1984-1990 (this is the assumption underlying the official South African Life Tables (Bah 2005)); completeness for Africans in this period was derived as described in the next section.

Three consistency checks[28] on the results of the SEG method were performed.

Firstly, the mortality rates produced by correcting the death data using the estimated completeness of reporting were calculated and then checked for consistency with other sources.

Secondly, as suggested by Preston and Lahiri (1991), alternative estimates of completeness derived using their "shortcut" method were compared to the SEG estimates, since the "shortcut" estimates do not depend on the assumption of constant completeness of reporting with age. The average age of the population and the average age at death at ages 30+ were calculated directly from the census and death data. The growth rate of the population was corrected using the estimate of delta calculated within the SEG method.

Lastly, an independent check of the mortality rates implied by the estimates of completeness was performed by comparing ${}_{15}q_{50}$ calculated using the corrected death

[27] Since death data classified by population group are not available for the years 1991-1996, it is not possible to calculate the coverage of the censuses in 1985 and 1991 relative to the 2011 census for each population group because this relies on chaining the estimated coverage of the censuses together. Therefore an estimate of completeness calculated for the population groups in the period 1985-1991 could not be adjusted to be relative to the coverage of the 2011 census. For this reason, the simplifying assumptions described next in the main text were made.

[28] The GGB method is not used as a consistency check since the SEG method was fit paying regard to the estimates produced by the GGB method and is therefore not an "independent" check.

data to ${}_{15}q_{50}$ calculated using orphanhood data from the censuses provided by Professor Dorrington (personal communication, 29 July 2016).

The estimated true number of deaths produced using the national estimates of completeness did not reconcile exactly to the aggregate of the true estimated deaths using the population group estimates, and thus the national estimates were adjusted to match the aggregate estimate of completeness, for the reasons discussed in section 3.3.5.

3.2.3 Estimating annual completeness of registration of deaths

3.2.3.1 Derivation of annual estimates

The estimates of completeness provided by the DDMs are relative to the population data used in the calculations. The estimates of census coverage from the DDMs indicate that the censuses prior to 2011 were undercounted relative to the 2011 census (except for Indians), implying that, relative to the 2011 census, the estimates of completeness are an overestimate of the level of completeness of registration of deaths by the extent of the undercount relative to the 2011 census. To produce estimates of completeness relative to the 2011 census in each of the intercensal periods, each census before 2011 was corrected to the extent of the undercount calculated by the application of the DDMs in each intercensal period after that date. The 2001 census was corrected for undercount relative to the 2011 census, and an estimate of completeness was produced for the period 2001-2011. Then, the 1996 census was corrected for undercount relative to the corrected 2001 census, an estimate of completeness was produced, and so forth.

To derive estimates of national completeness in each year, a curve was fit through the estimates on the assumption that the completeness for the year in the middle of an intercensal period is equal to the average level of completeness for that period. Denoting an estimate of completeness of reporting at time t by c_t, it is assumed that

$c_t = \frac{L}{1+e^{-k(t-x_0)}}$. The parameters of this logistic curve were found by minimising the absolute percentage error. Estimates of completeness for the individual years 1988-2009 were derived using the fitted curve. For the years 2009-2014, it is assumed that the trend defined by the fitted curve continues and therefore, the curve is used to extrapolate the estimates for these years. For the years 1985-1987, however, the curve produces estimates of completeness as low as 40 per cent, which appear to be implausibly low when compared to the estimate in 1988. For this reason, a minimum level of completeness equal to the value in 1988 is assumed. The validity of this assumption was considered by comparison with the national estimates of completeness implied by

Dorrington (1998), and is discussed in more detail in Footnote 35 (which appears on page 87).

To derive annual estimates of completeness for each of the population groups, logistic curves were fit to the estimates of completeness for Indians and Whites and constant completeness was assumed for Coloureds (since the estimated completeness for Coloureds was similar in both intercensal periods). The deaths remaining after subtracting the combined estimated true number of deaths for Coloureds, Indians and Whites from the estimated true number of deaths nationally, were attributed to the African population group. This relies on the assumptions that Coloured, Indian and White deaths are completely reported and the estimates of completeness derived using the DDMs are estimates of completeness of reporting of the population group of the deaths. Estimates of completeness for Africans were calculated as the ratio of the reported African deaths to the estimated true number of African deaths.

Since the estimates of completeness of reporting of male and female deaths were reasonably close to one another and there is no reason to expect the estimates of completeness to be systematically different for males and females, they were assumed to be the same in each year and a single curve was fit to the intercensal estimates for both sexes, for each population group.

3.2.3.2 Other adjustments

Three adjustments were then made to the annual estimates of completeness for the most recent years: for an observed drop in completeness of reporting in recent years, for late reported deaths and, more than counterbalancing these adjustments for the population groups, for an increase in the reporting of the population group of the deaths. These adjustments for late reporting and falling completeness were also applied to the annual estimates of completeness for Coloureds, Indians and Whites (none of the adjustments were applied for Africans, since annual estimates for Africans were derived so as to reconcile with the national figures discussed in the previous section). Separate adjustments for each population group were not derived since the data required are not available. However, the NEG model applied later allows the reasonability of these adjustments to be tested.

Since it has been suggested that completeness of reporting has fallen somewhat in recent years (Dorrington, Bradshaw, Laubscher *et al.* 2015), estimated completeness in the years 2012 and 2013 was reduced by 1.5 per cent (Professor Dorrington, personal communication, 29 June 2016). This was derived based on a comparison of the vital registration data recorded by Stats SA with the deaths recorded on the population register, over time.

Some death notifications are processed by Stats SA after the death data relating to a particular year of death have already been reported on and the numbers of deaths originally reported are augmented by these late processed deaths in subsequent Stats SA reports. Data on the number of late reported deaths since 2007, classified by the year of occurrence of the death and the year in which the death was reported, were received from Stats SA. These data were used to project the number of late reported deaths expected to be reported in future years, that relate to years of death up to and including 2013. To project the expected late reported deaths, it was assumed that the proportion of late reported deaths to deaths that have already been reported will remain constant in the future at the same level as was observed in 2013[29]. The expected late reported deaths were calculated as a percentage of the reported deaths, and these percentages were used to adjust the annual estimates of completeness downwards.

Lastly, the proportion of deaths reported without population group has fallen in recent years, implying that completeness of reporting of population group has risen in the years since 2011. A final adjustment to the estimates was therefore made to avoid using estimates of annual completeness that are too low.

The increase in reporting was estimated to be 13 per cent[30]. However, applying a blanket adjustment of 13 per cent to the data produced an implausible time series of deaths, especially for Indians, for whom the level of reporting of population group appears to have fallen, and therefore adjustments for each population group were derived separately. The deaths that would have been expected in each population group in the years 2011-2013, had no change in completeness of reporting of population group occurred, were projected on the assumption that the number of deaths occurring in each year was increasing exponentially (the growth rate was found using least squares regression). The percentage change in deaths actually reported compared to the projected deaths was calculated and the annual estimates of completeness of reporting were increased by this percentage in the years 2011-2013.

---

[29] This is similar to the assumption that is made in the actuarial technique for projecting late reported insurance claims, known as the chain ladder method.

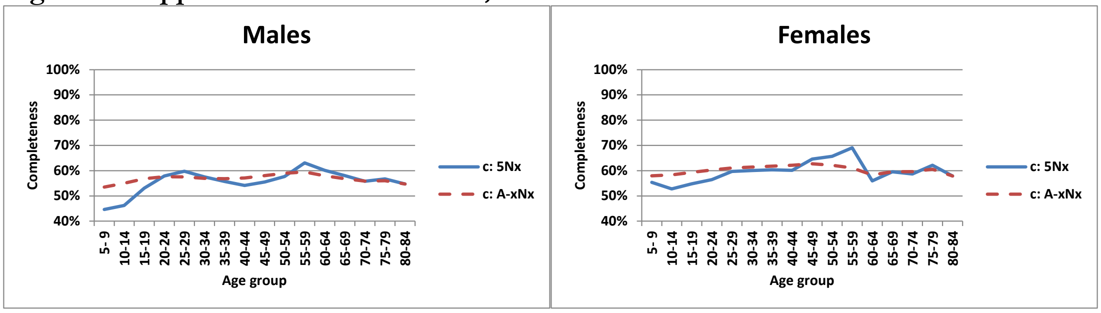
[30] The increase was calculated as $\frac{1 - p_{2011,2013}^{unknown}}{1 - p_{2000,2010}^{unknown}}$, where $p_{2000,2010}^{unknown}$ is the average proportion of deaths reported without population group in the years between 2000 and 2010. Other combinations of years were tested for the denominator of this fraction but all produced similar results.

3.3 Results

3.3.1 1985-1991 - National

The SEG diagnostics (which show two series, the estimated population divided by the reported population at each age shown in the series "$c:\ {}_5N_x$" and the sum of the estimated population from age x to maximum age A divided by the corresponding reported population shown in the series "$c:\ {}_{A-x}N_x$") showed a drop in completeness of reporting for ages 5-24 as shown in Figure 3.1, and the estimates of completeness were therefore calculated on the age range 25-84, producing results of 57 per cent and 61 per cent for males and females respectively, for the population as a whole.

Figure 3.1: Application of SEG method, 1985-1991

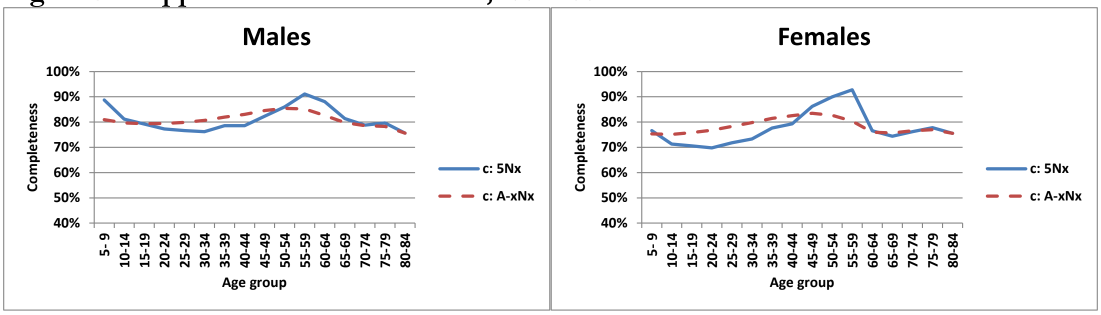


The SEG diagnostic plots (as well as the GGB plots shown in Appendix B.2) show suspicious peaks in estimated completeness at ages 50+ which indicate that the population data are probably somewhat deficient. However, the consistency between the estimates for males and females, and the consistency of the SEG estimates with the GGB estimates (53 per cent and 62 per cent for males and females respectively) encourages confidence in the results.

3.3.2 1991-1996 - National

The SEG diagnostic for males and females showed a lower level of completeness at the open intervals beginning at the younger ages until around age 25, as illustrated in Figure 3.2. The estimates of completeness determined on the ages 25-84 were 80 per cent and 78 per cent for males and females respectively.

Figure 3.2: Application of SEG method, 1991-1996



The SEG method appears to work reasonably well in this period; however, the population data again appear deficient. Some support for the estimates is provided by the close consistency of the estimates for males and the consistency of the SEG estimates with the GGB estimates (79 per cent and 75 per cent for males and females).

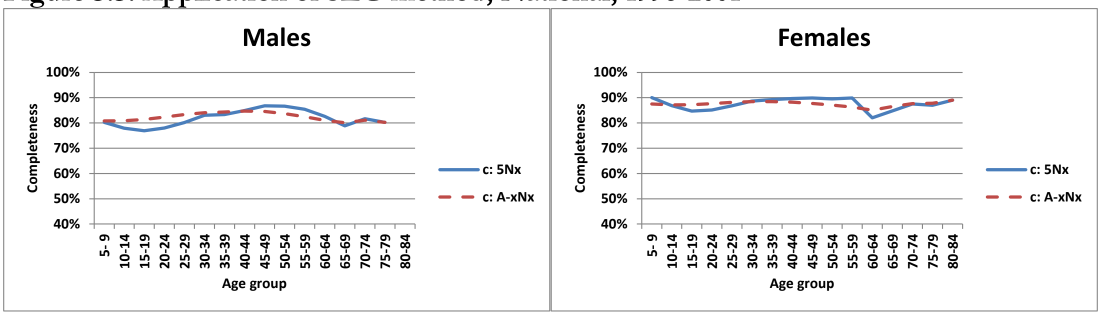
A slightly better fit could have been achieved by not allowing for migration when estimating the completeness of reporting, which implies that the estimates of migration from the ASSA model may not be entirely consistent with the population data used in this period. However, the impact on the estimates of completeness is not significant and, given that substantial migration undoubtedly occurred in this period and the lack of more reliable alternative estimates of migration, the estimates of completeness allowing for migration are used.

3.3.3 1996-2001

3.3.3.1 National

An initial application of the SEG method to the data for males and females showed lower than average registration of deaths in the open intervals beginning at ages 5-29, as illustrated in Figure 3.3, and these ages were therefore excluded when calculating the estimate of completeness. The age group 80-84 was excluded when calculating delta for males to achieve better consistency with the GGB estimates. Since completeness appears to fall at both the younger and older ages, the estimates were derived on the age range 30-69. The SEG method produced estimates of completeness of 85 per cent and 90 per cent for males and females respectively.

Figure 3.3: Application of SEG method, National, 1996-2001



3.3.3.2 Population groups

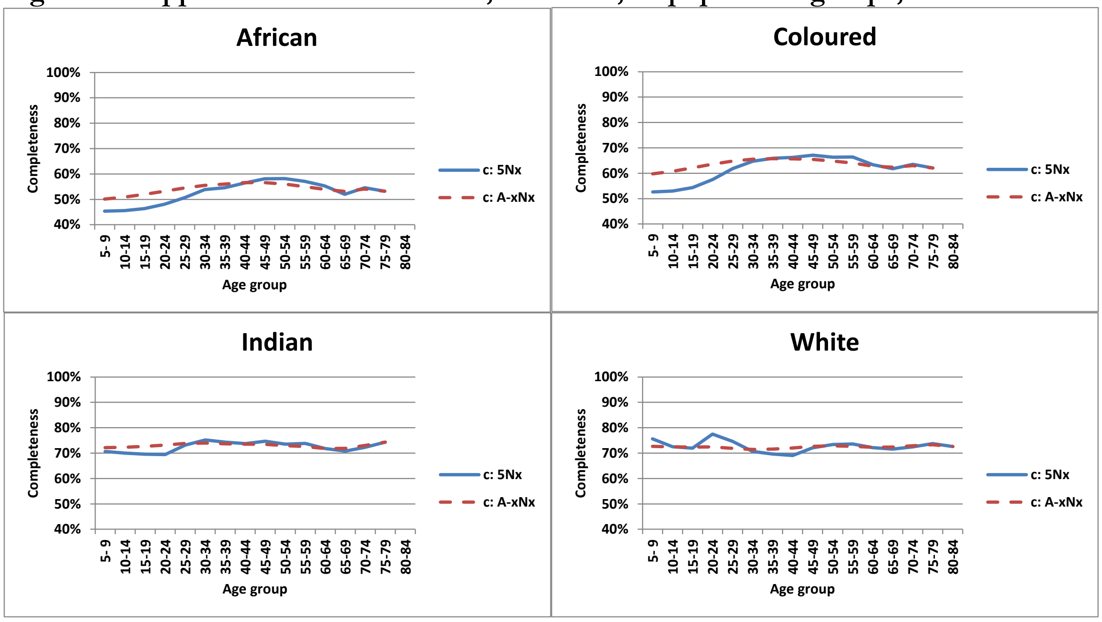
The SEG method produced the estimates of completeness shown in Table 3.1 and Figure 3.4 and Figure 3.5. When determining these estimates, delta was computed on the age range 5-84 for all of the groups except for African, Indian and Coloured males and Coloured females, for which an age range of 5-79 was used to ensure better consistency with the GGB estimates.

A drop in estimated completeness of reporting can be observed in the SEG diagnostic at the younger open intervals, and the estimated completeness is more volatile at the older open intervals, for all population groups and both sexes and therefore the estimates of completeness were calculated on the age range 25-64 in all cases. The estimates were relatively stable over the age ranges selected for the calculation.

Table 3.1: Estimates of completeness and age ranges used when applying the SEG method, 1996-2001, all population groups

	Males		Females	
	Completeness	Age range	Completeness	Age range
African	56%	25-64	64%	25-64
Coloured	66%	30-64	69%	25-64
Indian	74%	25-64	82%	25-64
White	72%	25-64	75%	25-64

Figure 3.4: Application of SEG method, 1996-2001, all population groups, Males

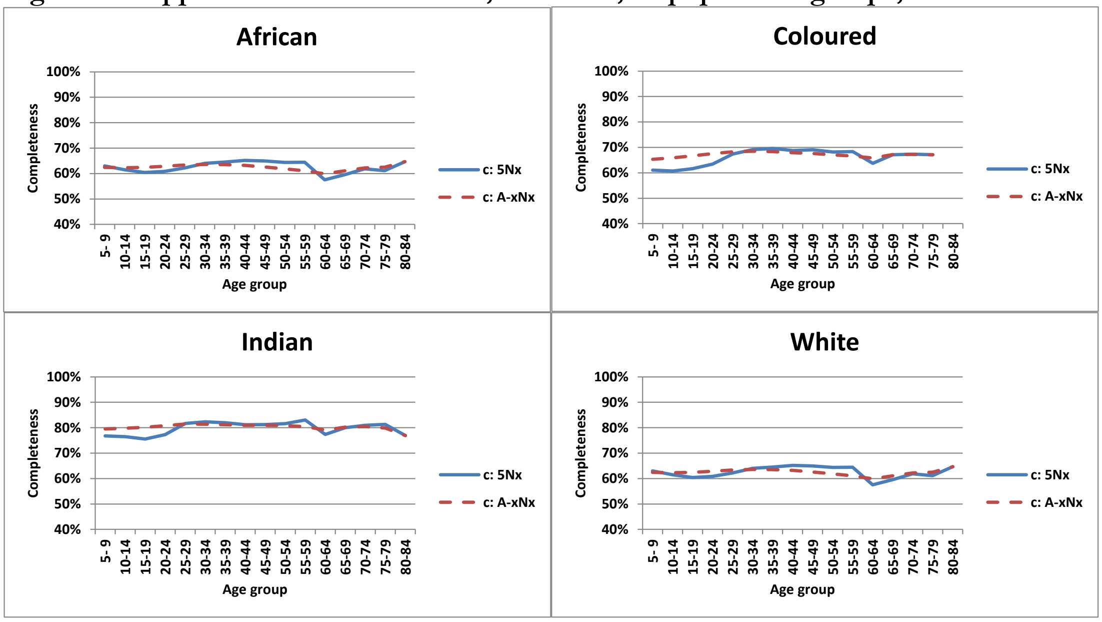


When estimating life expectancy for the SEG method for Indian females, it was found that the ASSA mortality tables do not appear to fit the experience at the oldest ages very well and produce mortality rates that seem too low (and estimates of life expectancy that are too high) in comparison to the ungraduated rates. This is problematic since the SEG calculation for Indian females was very sensitive to the choice of mortality table used as the standard, because of the apparently high growth rate of the Indian female population in the open age interval 85+ between the years

1996 and 2001. To avoid this problem, a life table from the General family of United Nations model life tables for developing countries (United Nations 1982) was used to graduate the mortality rates, since it produces a better fit to the data.

The estimates of completeness for males and females were close to each other for Coloureds and Whites, but are more than 5 per cent apart for Africans and Indians. The incompleteness of reporting for Coloureds, Indians and Whites is mostly due to population group not being recorded on death certificates (i.e. population group was reported as unknown), rather than the deaths not being reported.

Figure 3.5: Application of SEG method, 1996-2001, all population groups, Females

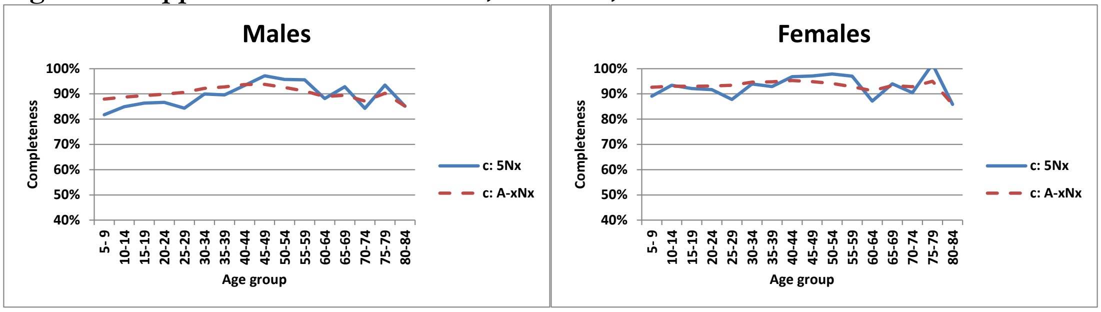


3.3.4 2001-2011

3.3.4.1 National

The diagnostic plots showed decreased levels of estimated completeness in the open intervals with starting ages younger than 35 for males and females, as well as a zigzag pattern at the older ages for both sexes, as illustrated in Figure 3.6. Changing the age range used to calculate delta did not increase the consistency with the GGB estimate and so delta was calculated using the range 5-84. To avoid distorting the estimates of completeness, the estimates were calculated using the age range 30-64 for males and 25-64 for females, producing estimates of 93 per cent and 95 per cent respectively.

Figure 3.6: Application of SEG method, National, 2001-2011



3.3.4.2 Population group

Table 3.2 and Figure 3.7 and Figure 3.8 show the application of the SEG method in the period 2001-2011.

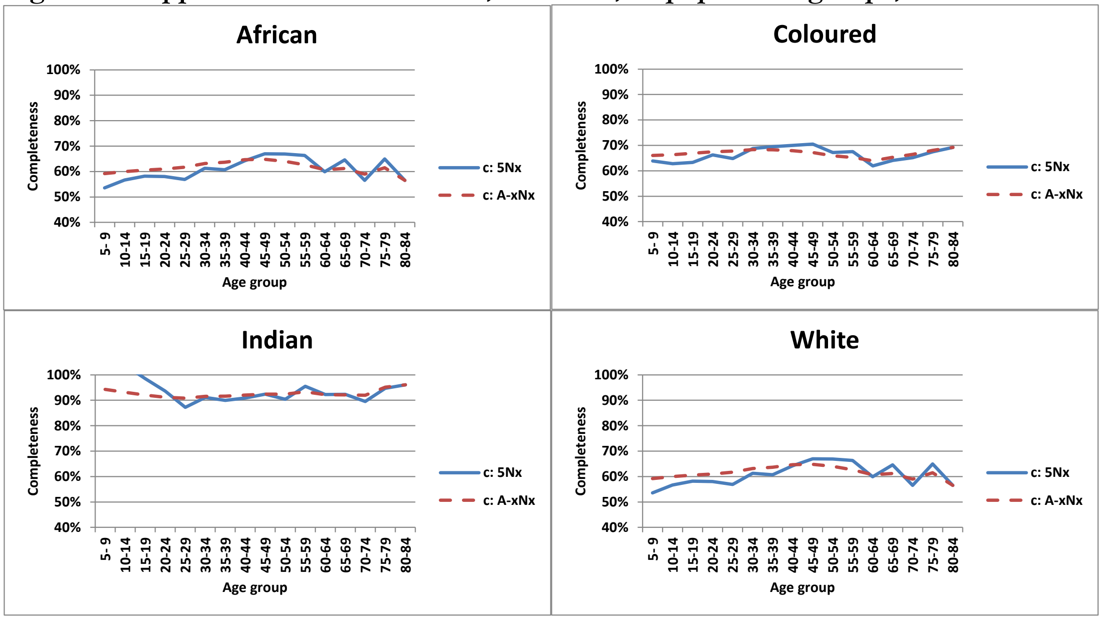
Similar to the estimates in 1996-2001, a different level of completeness of reporting was observed for the open intervals beginning at the younger ages until the age of 24 and the SEG diagnostic was quite volatile at the older ages and displayed a zigzag pattern, so the estimates of completeness were calculated using the age range 25-64. For White males and Coloured and Indian females the estimate of delta was calculated on the age range 5-79 to achieve more consistency with the GGB method and for White females the age range used was 5-74[31]. The estimates of completeness for males were similar to that of females for each of the population groups.

Table 3.2: Estimates of completeness and age ranges, SEG method, 2001-2011, all population groups

	Males		Females	
	Completeness	Age range	Completeness	Age range
African	63%	25-64	64%	25-64
Coloured	68%	25-64	67%	25-64
Indian	91%	25-64	92%	25-64
White	78%	25-64	77%	25-64

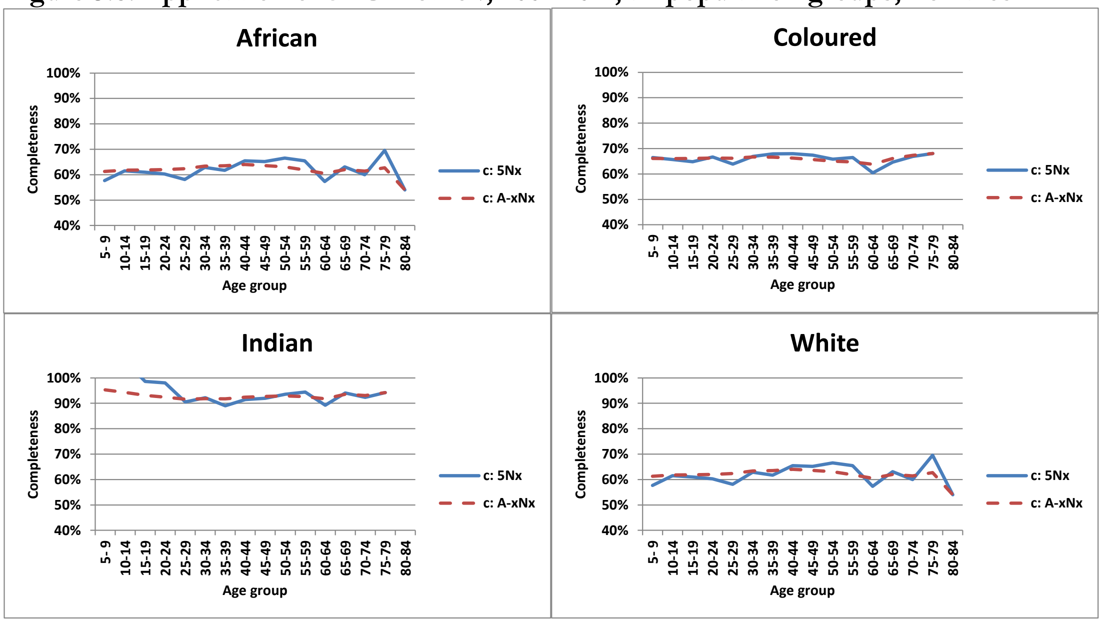
[31] A slightly different procedure was followed for White females to achieve greater consistency between the GGB and SEG methods. After solving for the estimate of delta, the corrected census population for 2001 was used as the input into another SEG calculation, with the estimate of delta set to zero.

Figure 3.7: Application of SEG method, 2001-2011, all population groups, Males



Similar to the finding in the 1996-2001 period, when estimating life expectancy for the SEG method for Indian females, it was found that the ASSA mortality tables again do not appear to fit the experience at the oldest ages and a life table from the General family of United Nations model life tables for developing countries (United Nations 1982) was therefore used to graduate the mortality rates.

Figure 3.8: Application of SEG method, 2001-2011, all population groups, Females



Possible evidence of decreasing completeness of registration with age was found in this period at the ages 80-84 for African males and females, with a 10 per cent and 15 per cent drop in completeness respectively. However, the erratic zigzag pattern shown in the diagnostic indicates that a more likely explanation is that the population data at these ages are deficient (as a result of undercounting or age misreporting in these age groups).

Distortions for Indians and Whites can also be seen at the younger ages, especially for females.

3.3.5 Reconciliation of national and population group results

It would be expected that grossing up the registered deaths (in each intercensal period) using the estimates of completeness at a national level would produce a similar estimate of the true number of deaths to the estimate produced by the sum of the registered deaths in each population group grossed up using the estimates of completeness for each population group. However, as shown in Table 3.3 and Table 3.4, fewer deaths were produced using the national estimates of completeness and the level of national completeness implied by the estimates for each population group are lower than the estimates using the population as a whole.

The implied levels of completeness are lower for both sexes in both periods, which means that either the estimates of completeness calculated for the population as a whole systematically over-estimate the level of completeness, or that the combined estimates of completeness for each population group systematically under-estimate completeness.

Table 3.3: Estimated true number of deaths and implied completeness, 1996-2001

	Males	Females
d_{30+} - National	904 365	719 289
d_{30+} - Population groups	1 027 442	773 439
cx - national	85%	89%
cx - implied	75%	83%

Table 3.4: Estimated true number of deaths and implied completeness, 2001-2011

	Males	Females
d_{30+} - National	2 385 515	2 171 599
d_{30+} - Population groups	2 555 905	2 346 685
cx - national	93%	95%
cx - implied	87%	88%

To test if the estimated true deaths using the population group estimates of completeness are consistent with the census (and migration) data in each intercensal period, the DDMs were applied to the estimated true deaths to assess the level of completeness produced by these deaths. These results are shown in Table 3.5 and Table

3.6 for males and females respectively and indicate that deaths corrected using the population group estimates of completeness are close to 100 per cent complete.

Table 3.5: Estimated completeness of the true number of deaths corrected using population group estimates of completeness, Males

Males	GGB	SEG
1996-2001	97%	101%
2001-2011	98%	102%

Table 3.6: Estimated completeness of the true number of deaths corrected using population group estimates of completeness, Females

Females	GGB	SEG
1996-2001	92%	99%
2001-2011	103%	103%

A comparison of the two sets of corrected deaths is shown for each sex and period in Figure 3.9 below. The ratio of the two sets of deaths is consistently above 100 per cent, with the bulk of the extra deaths predicted at the younger ages, implying that a single estimate of completeness at the national level does not correct for the missing deaths at each age appropriately and that the assumption of constant completeness of reporting by age is violated at the national level. As support for this explanation, Figure 3.10 compares the proportion of deaths by population group at each age. The proportion of deaths from the African population group is at its highest at the ages below 40, and after this age, the proportion of deaths from the White population group begins to rise.

When estimating completeness at a national level, the underlying assumption of the SEG method of constant underreporting by age will be violated since the White population group has a higher level of completeness of reporting than the African population group (as shown above), and proportionately more deaths from the White population group occur at the older ages. This analysis suggests that the estimated true deaths using the population group estimates of completeness are likely to be more robust than those calculated using the national estimates of completeness.

Figure 3.9: True deaths estimated using population group completeness as a ratio of true deaths estimated using national completeness

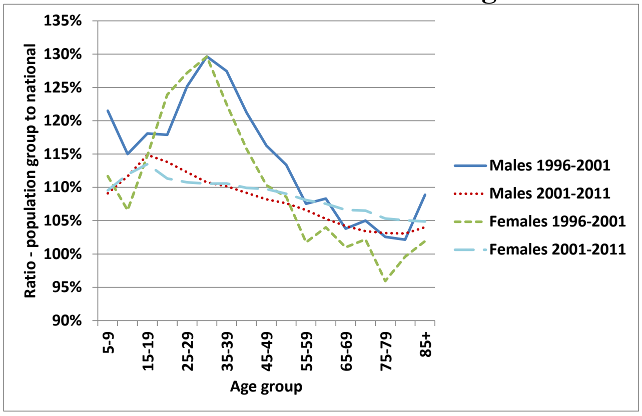
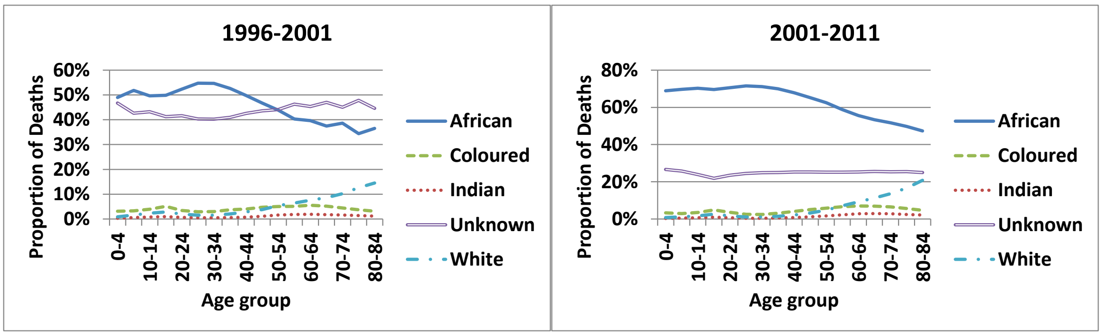


Figure 3.10: Proportion of deaths attributable to each population group by age, 1996-2001 and 2001-2011



On this basis, the estimates of completeness at a national level were recalculated as the proportion of the reported deaths to the estimated true deaths using the population group estimates of completeness at ages 30+. The final set of estimates for each of the intercensal periods 1996-2001 and 2001-2011 are shown in Table 3.7. The adjusted estimate for males in the period 1996-2001 is 13 per cent lower than the estimate in 2001-2011, implying a sharp rise in completeness of reporting in this period. The adjusted estimate for males is significantly lower than the estimate for females by approximately 9 per cent, which is in line with the difference between estimated completeness for African males and females in the period 1996-2001.

Table 3.7: Adjusted estimates of completeness, 1996-2001 and 2001-2011, Males and Females

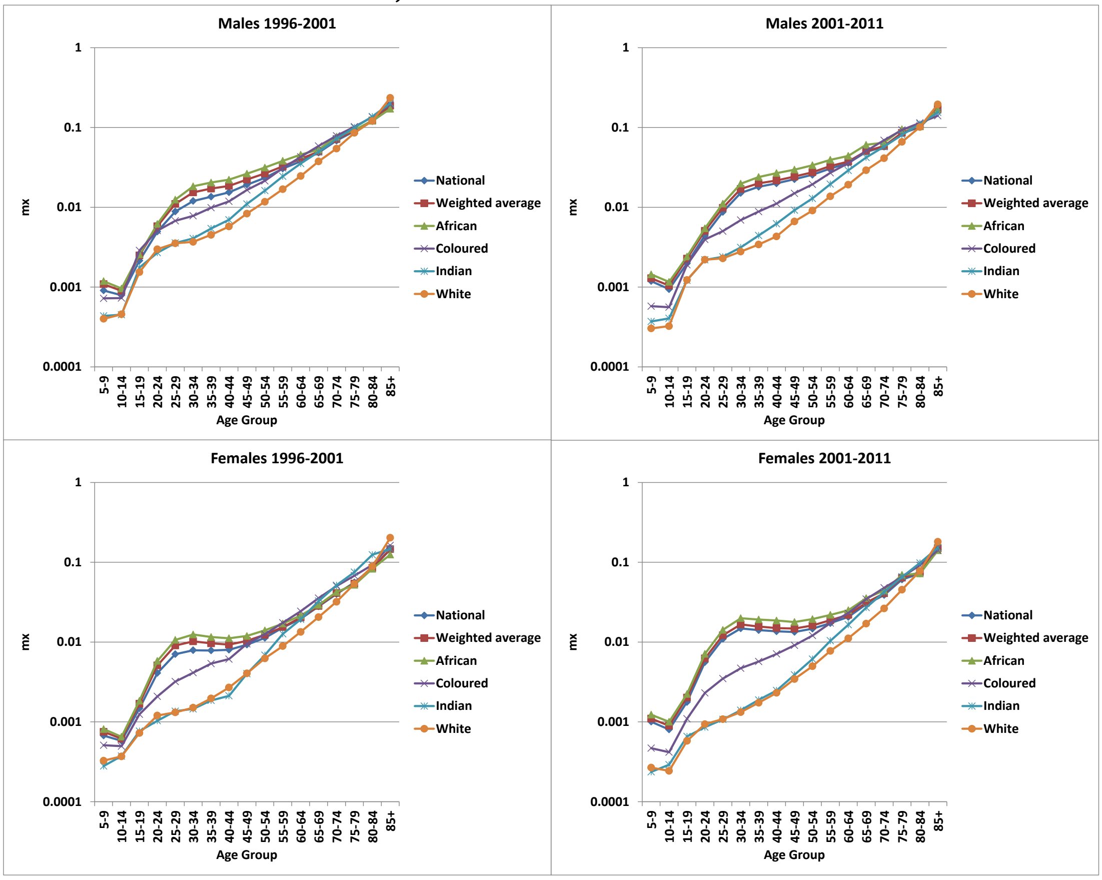
	Males	Females
1996-2001	74%	83%
2001-2011	87%	88%

3.3.6 Validation of estimates of completeness

3.3.6.1 Mortality rates implied by the estimates

Ungraduated mortality rates estimated for the period 1996-2001 and 2001-2011 using the corrected death data are shown in Figure 3.11.

Figure 3.11: Ungraduated mortality rates (log-scale), National and by population group, Males and Females, 1996-2001 and 2001-2011



The ranking of the mortality rates by population group is in line with Dorrington, Moultrie and Timæus (2004) and the ASSA AIDS model (Actuarial Society of South Africa 2009) for both sexes and each period. The rates for Africans are highest, followed by Coloureds, Indians and Whites. In each period, the rates for males are higher than those for females, except for Africans at younger adult ages (20-34) in the period 2001-2011, due to the impact of HIV/AIDS (mortality rates produced by the ASSA model also cross over, but at the ages 20-29).

A comparison of mortality rates between the two periods for the African population group shows, as might be expected because of the impact of HIV/AIDS, an increase in mortality from 1996-2001 to 2001-2011. Mortality for the Indian and White, and to a lesser extent, Coloured, population groups in 2001-2011 was generally lower than in 1996-2001. Thus the comparison shows that the mortality rates produced by applying the corrections for incomplete reporting that were estimated in this section are in line with expectations.

3.3.6.2 Estimates derived using the Preston and Lahiri method

The Preston and Lahiri "shortcut" method does not make the assumption that completeness of reporting of death is constant with age and relies instead on the assumptions that the mean age of the population and the mean age at death are accurate[32] (Preston and Lahiri 1991). Furthermore, the method does not produce diagnostics enabling the evaluation of the results, and has not been well studied in the literature. However, trial analysis performed by Hill, You and Choi (2009) suggests that the shortcut method is affected by errors in a similar way to the SEG method (probably because the method that was tested in that study did not allow for differential coverage of the two censuses[33]) and Palloni, Pinto and Beltrán-Sánchez (2016), who tested a number of different DDMs on simulated populations, report in a working paper that the shortcut method performs relatively poorly[34]. Therefore, the shortcut estimates are used as a comparison to the SEG, instead of being accepted outright. The shortcut estimates from this method are presented in Table 3.8 and Figure 3.12.

The national SEG estimates (before correction to the weighted average) are on average 6 per cent different from the shortcut estimates and are biased upwards (i.e. all of the SEG estimates are higher than the shortcut estimates) compared to 3 per cent after correction without an upwards bias, adding to the evidence that a violation of the assumption of constant completeness of reporting of death affected the national estimates, and indicating that this has been remedied, to some extent, in the corrected estimates. The estimates for the African, Coloured and Indian population groups appear to be biased upwards, whereas the corrected estimates at a national level and for the White population group do not appear to be biased.

The highest average errors are for the Indian and Coloured population groups, followed by the African and White population groups. However, the high average relative error for the Indian population group results mainly from the SEG estimate for

[32] Although it is shown in Chapter 4 that both the population and death data suffer from age exaggeration at the older ages, the bias introduced into the shortcut estimates is probably small since the population numbers at the older ages are small in comparison to the younger ages, and, for Africans and Coloureds, the majority of the deaths occur at the younger ages.

[33] An allowance for differential coverage of the censuses was made when applying the shortcut method in this research, as discussed in Section 3.2.2, by correcting the growth rates using the estimate of delta from the SEG method.

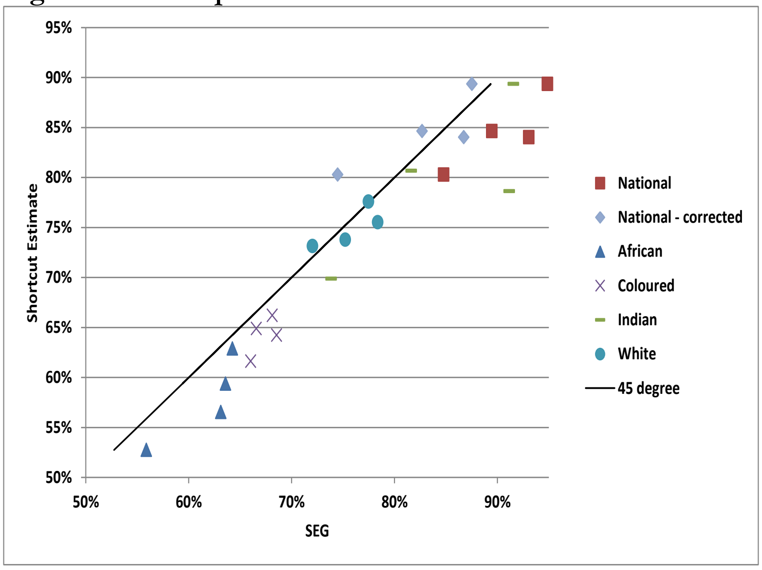
[34] Palloni, Pinto and Beltrán-Sánchez (2016) do not describe how the shortcut method was applied. Also noteworthy is the puzzling difference between the median and mean results of the shortcut method reported in the study.

males in 2001-2011 and excluding this point results in an average relative difference of 2 per cent. Since the SEG and shortcut methods are close for females in 2001-2011 and the SEG estimate for males is consistent with the estimate for females, it does not appear as if there is enough evidence to suggest that the SEG estimate for males in this period should be adjusted.

Table 3.8: "Shortcut" and SEG estimates of completeness, 1996-2001 and 2001-2011, Males and Females

		1996-2001		2001-2011		
Level	Sex	SEG	Shortcut	SEG	Shortcut	Average relative difference
African	Male	56%	53%	63%	57%	4%
African	Female	64%	63%	64%	59%	
Coloured	Male	66%	62%	68%	66%	5%
Coloured	Female	69%	64%	67%	65%	
Indian	Male	74%	70%	91%	79%	6%
Indian	Female	82%	81%	92%	89%	
White	Male	72%	73%	78%	76%	3%
White	Female	75%	74%	77%	78%	
National	Male	85%	80%	93%	84%	6%
National	Female	89%	85%	95%	89%	
National - corrected	Male	74%	80%	87%	84%	3%
National - corrected	Female	83%	85%	88%	89%	

Figure 3.12: Comparison of "shortcut" and SEG estimates of completeness



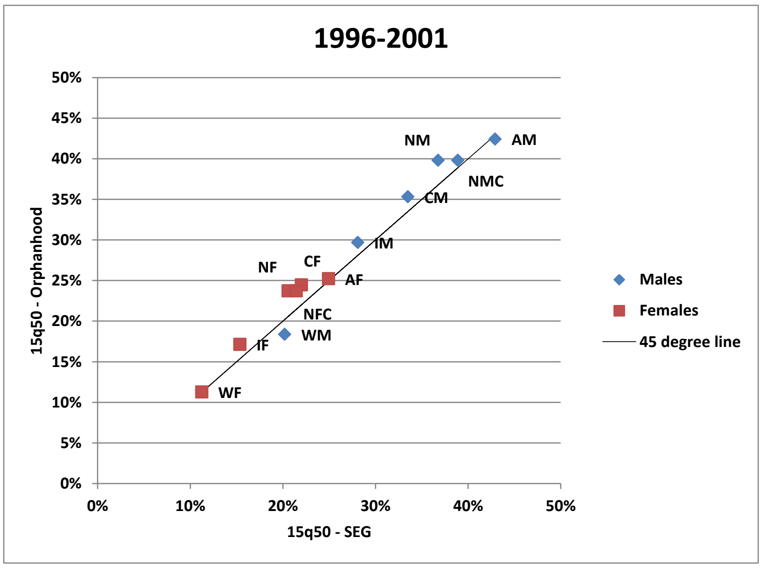
3.3.6.3 Estimates derived using data on survival of parents

Estimates of ${}_{15}q_{50}$ based on the orphanhood data collected in the 1996, 2001 and 2011 censuses were compared to the estimates of ${}_{15}q_{50}$ derived using the corrected deaths and population figures from the SEG method. This comparison is shown in Table 3.9 and Figure 3.13 and Figure 3.14.

Table 3.9: $_{15}q_{50}$ estimated from orphanhood data and from deaths corrected via the SEG method

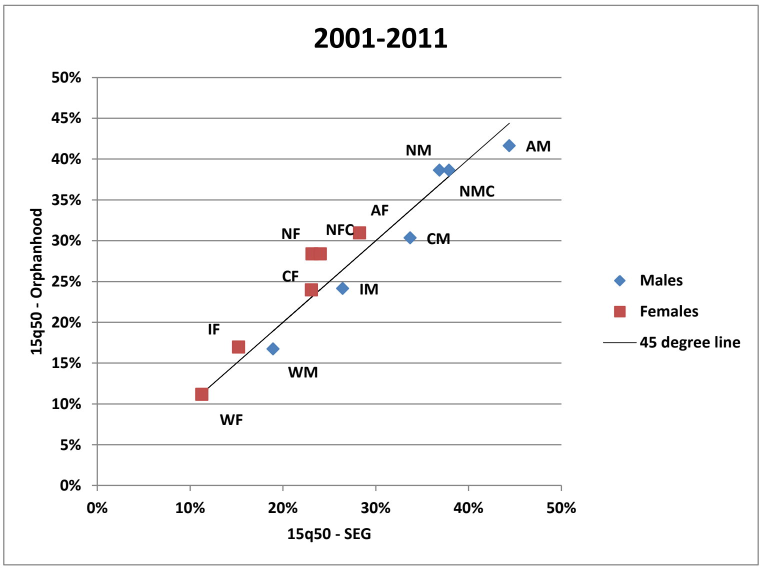
			1996 - 2001		2001 - 2011		
Level	Sex	Abbreviation	SEG	Orphanhood	SEG	Orphanhood	Average relative difference
African	Males	AM	0.41	0.42	0.44	0.42	4%
African	Females	AF	0.25	0.25	0.28	0.31	
Coloured	Males	CM	0.34	0.35	0.34	0.30	7%
Coloured	Females	CF	0.22	0.24	0.23	0.24	
Indian	Males	IM	0.28	0.30	0.26	0.24	9%
Indian	Females	IF	0.15	0.17	0.15	0.17	
White	Males	WM	0.20	0.18	0.19	0.17	6%
White	Females	WF	0.11	0.11	0.11	0.11	
National	Males	NM	0.37	0.40	0.37	0.39	11%
National	Females	NF	0.21	0.24	0.23	0.28	
National – Corrected	Males	NMC	0.39	0.40	0.38	0.39	7%
National – Corrected	Females	NFC	0.21	0.24	0.24	0.28	

Figure 3.13: $_{15}q_{50}$, Orphanhood and SEG estimates, Males and Females, 1996-2001



The mortality estimates from the SEG model are broadly consistent with those produced by the orphanhood method (the Pearson correlation coefficient, R^2, was 99.22 per cent in the period 1996-2001 and 96.77 per cent in the period 2001-2011). Similar to the findings above, the average relative error of the National estimates before correction to the weighted average completeness is higher (and the highest of all of the groupings) than those of the corrected National estimates, showing that the correction applied to the national estimates of completeness ensures greater consistency with the orphanhood data. The SEG estimates for the African population group are the most consistent with the orphanhood data followed by those for the White population group.

Figure 3.14: $_{15}q_{50}$, Orphanhood and SEG estimates, Males and Females, 2001-2011



3.3.6.4 Conclusion

The estimates of completeness produced using the SEG method, and the mortality rates produced using the data corrected with these estimates, are consistent with the other sources examined in this section. Furthermore, no critical problems with the SEG estimates of completeness were revealed in the comparisons.

The correction applied to the national estimates increased the consistency of the estimates of completeness with those of both the shortcut method and those derived from orphanhood data, adding to the arguments presented earlier, that the weighted average completeness should be used at a national level.

The upwards bias of the SEG estimates of completeness compared to the shortcut estimates of completeness for the African, Coloured and Indian population groups indicates that further adjustments to the estimates of completeness for these groups may be warranted. Since the shortcut method itself may have produced inaccurate results due to age exaggeration in the population and death data, the shortcut estimates are not used to adjust the estimates of completeness. Rather, adjustments to the annual estimates of completeness are investigated statistically in the next chapter. For the White population group, the SEG estimates are consistently close to both the shortcut method and the orphanhood mortality estimates, implying that these estimates are reasonably accurate and that further adjustments are likely unnecessary.

3.3.7 Estimates of completeness of reporting relative to Census 2011

By linking the estimates of the relative undercounts of one census to the other (produced during the application of the SEG method), Table 3.10 was produced which shows the undercount of the censuses since 1985 relative to the census in 2011 (the population data provided by Stats SA are corrected for undercount estimated by the Post Enumeration Surveys (PES), therefore, these estimates of undercount are in addition to the PES estimates). Except for the Indian population group, all censuses were undercounted relative to the 2011 census. The 1996 census was undercounted relative to both the 2001 and 2011 censuses except for White males and females, who had slightly better coverage in 1996 than in 2001.

Table 3.10: Census undercount relative to Census 2011

	National Males	National Females	African Males	African Females	Coloured Males	Coloured Females	Indian Males	Indian Females	White Males	White Females
1985	81%	88%								
1991	90%	88%								
1996	92%	94%	86%	91%	92%	94%	100%	101%	91%	92%
2001	97%	99%	93%	97%	97%	99%	102%	103%	89%	90%

Using these estimates of census coverage, the estimates of completeness of reporting of deaths in 1985-2001 were adjusted to be relative to the 2011 census population. These adjusted results are required to estimate the trend in completeness of reporting, independently of changes in census coverage, and are shown in Table 3.11.

Table 3.11: Estimates of completeness of reporting of deaths, 1985-2011, estimates adjusted for census undercount

	National Males	National Females	African Males	African Females	Indian Males	Indian Females	Coloured Males	Coloured Females	White Males	White Females
1985-1991	51%	54%								
1991-1996	74%	73%								
1996-2001	72%	82%	52%	62%	75%	84%	64%	68%	64%	68%
2001-2011	87%	88%	63%	64%	91%	92%	68%	67%	78%	77%

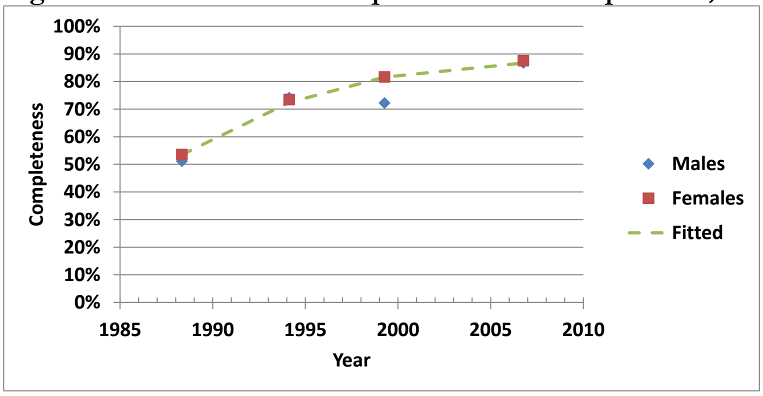
3.3.8 Interpolation and extrapolation of the estimates of completeness

3.3.8.1 National estimates

Since the estimates of completeness of reporting of male and female deaths were reasonably close to one another, and there is no reason to expect the estimates of completeness to be systematically different for males and females, the estimates were combined into one dataset to use for fitting the logistic curve. The equation was estimated as $c_t = 0.89(1+e^{-0.19(t-1986.1)})^{-1}$. The curve was used to produce estimates of the annual completeness for the years 1988-2013. For the years 1985-1987, extrapolating the curve produces unrealistically low levels of completeness and therefore the estimates for the years 1985-1987 were fixed at the fitted estimate in 1988 of 54 per cent, which appears reasonable compared to the average completeness of reporting at a national level implied by Dorrington (1998)[35].

The results of the curve fitting are shown in Figure 3.15 below. The curve appears to fit the data reasonably well, passing within the range defined by the estimates for males and females, except for males in the period 1996-2001.

Figure 3.15: Estimates of completeness and interpolation, National



The annual estimates of completeness in the most recent years were then adjusted to allow for the deaths still outstanding in 2013 and a recent drop in the completeness of reporting.

Data on late reported deaths at ages 30+ received from Stats SA were tabulated into a reporting "triangle", shown in Table 3.12, in order to project the number of deaths that have already occurred, that will be reported late in years subsequent to 2013.

[35] Dorrington (1998) found that reporting for Africans in 1985 was 56% complete for males and 44% for females. Assuming that deaths for the other population groups were completely reported relative to the 1985 census this implies that national completeness of reporting was 67% for males and 58% for females, relative to the 1985 census. Applying the estimated census undercount at a national level in Table 3.10 to these figures implies that relative to the 2011 census, national completeness in 1985 was 55% and 51% for males and females respectively. Alternatively, applying the national undercount in Table 3.10 to the estimated completeness for Africans implies that national completeness was 57% for males and 52% for females; however, these figures are probably a bit too high since the undercount for Africans was likely higher than nationally. These estimates imply that completeness of reporting in 1985 was at a similar level as in 1988 and it seems reasonable to assume that completeness did not fluctuate much in the intervening years (i.e. completeness in these years can be approximated with a constant since there is no evidence of a trend over time in completeness). Therefore, the assumption of completeness at 54% in the years 1984-1988 that is made in the text appears reasonable.

Deaths reported as having occurred in the year of the Stats SA report are shown in column 1 of the table. The following columns add to that the total late registrations recorded in the reports, in subsequent annual reports. For example, 453 804 deaths were

reported as having occurred in 2007 in the Stats SA report on deaths recorded in the year 2007. In the report relating to deaths recorded in 2008, an extra 1 472 deaths were reported as having occurred in 2007, bringing the cumulative total of deaths occurring 2007 to 455 276.

Table 3.12: Deaths occurring in each of the years 2007-2013, and cumulative adjustments made in subsequent reports from Stats SA

		Reporting years elapsed since death							
		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Year of occurrence of death	*2007*	453 804	455 276	455 955	456 126	456 154	456 207	456 850	**457 217**
	2008	448 134	450 317	450 644	450 686	450 753	451 657	**452 253**	
	2009	444 569	449 297	449 511	449 858	450 970	**451 709**		
	2010	423 553	426 469	427 126	428 064	**428 421**			
	2011	404 219	409 235	410 926	**411 273**				
	2012	383 390	391 835	**392 573**					
	2013	370 191	**381 737**						

The proportions of the cumulative total deaths in each elapsed reporting year to the cumulative total deaths in the prior reporting year (called "development factors" in the actuarial literature) were calculated and are shown in Table 3.13. It can be seen that a higher proportion of deaths are reported late in more recent years (for example, 2007 deaths reported late, in 2008, were equal to 0.3 per cent of the deaths reported in 2007, compared to late registrations in 2014 of deaths in 2013 equal to 3.1 per cent of the deaths reported in 2013). The Stats SA reports have not been issued at regular time intervals and recent reports have been issued sooner after the year of death. The increasing proportion of deaths reported late in recent years appears to be the result of an acceleration in the release of data from 23 months to 11 months after the end of the calendar year of death. The most recent proportions, shown in bold in Table 3.13, were used to derive the late reporting adjustments for each year of deaths which shown in Table 3.14. For example, it is expected that the deaths occurring in 2013 that will be reported late in future years will be equal to 3.9 per cent of the deaths reported in the 2013 Stats SA report. These deaths are accounted for by reducing the annual estimate of completeness by 3.9 per cent.

Table 3.13: Individual development factors

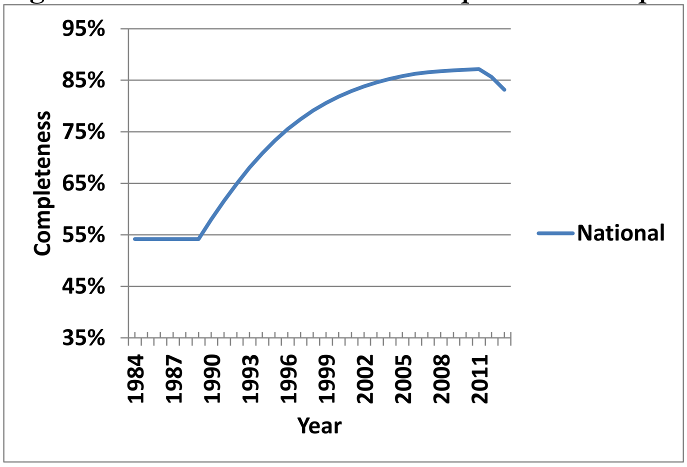
		Reporting years elapsed since death						
		0	*1*	*2*	*3*	*4*	*5*	*6*
Year of occurrence of death	*2007*	1.003	1.001	1.000	1.000	1.000	1.001	**1.001**
	2008	1.005	1.001	1.000	1.000	1.002	**1.001**	
	2009	1.011	1.000	1.001	1.002	**1.002**		
	2010	1.007	1.002	1.002	**1.001**			
	2011	1.012	1.004	**1.001**				
	2012	1.022	**1.002**					
	2013	**1.031**						

Table 3.14: Adjustments made to the estimate of completeness of reporting of deaths for each year since 2007

		Percentage of deaths reported in 2013, still to be reported in future years	Adjustment
Year of occurrence of death	*2007*	100.1%	99.9%
	2008	100.2%	99.8%
	2009	100.4%	99.6%
	2010	100.5%	99.5%
	2011	100.5%	99.5%
	2012	100.7%	99.3%
	2013	103.9%	96.3%

An additional adjustment of 1.5 per cent was applied to the completeness of reporting in the years 2012 and 2013 to correct for what appears to be a drop in completeness observed since 2011 (Professor Dorrington, personal communication, 29 June 2016). The annual estimates of completeness after adjustments are given in Appendix B.2.5 and shown in Figure 3.16.

Figure 3.16: Annual estimates of completeness of reporting of deaths, 1984-2013, national



3.3.8.2 Population group estimates

The estimates of completeness for the Coloured population group are very close together for the two intercensal periods in the period 1996-2011 so completeness was assumed to be constant at the average of the estimates (67 per cent) relating to the period 1996-2011.

For the Indian and White population groups, separate logistic curves were fit to the estimates of completeness for each population group and both sexes (i.e. the four estimates for males and females from each population group shown in Table 3.11 were used to estimate the parameters) to produce annual estimates of completeness. The logistic curves (of the form $c_x = \frac{L}{1+e^{-k(x-x_0)}}$) were estimated as shown in Table 3.15 and in Figure 3.17 and Figure 3.18.

Table 3.15: Fitted logistic curve, Whites and Indians

	Whites	Indians
L	78%	91%
k	2.01	1.20
x0	1999	1998

Figure 3.17: Estimates of completeness and interpolation, White

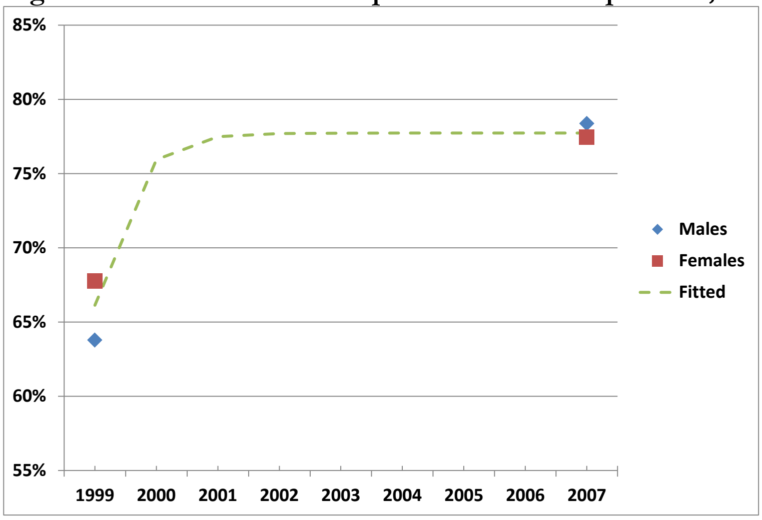
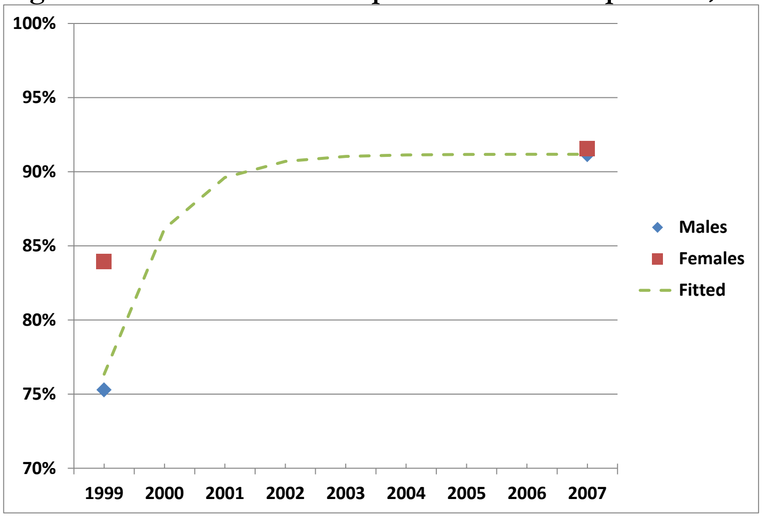


Figure 3.18: Estimates of completeness and interpolation, Indian



The annual estimates of completeness were then adjusted to allow for the deaths still outstanding in 2013, using the same estimates as used for the national estimates (since data to estimate late reporting by population group were not available) and for the recent fall in completeness of reporting.

A final adjustment was made for the marked decrease in the number of unknown deaths that can be observed in the data in 2011-2013, which implies that completeness of reporting of population group increased in these years. The decrease in unknown deaths was estimated to be 13 per cent using $\frac{1-p_{2011,2013}^{unknown}}{1-p_{2000,2010}^{unknown}}$, where $p_{x,y}^{unknown}$ is the proportion of deaths reported without race in the years between x and y. Applying this blanket adjustment to all of the population group data produced an implausible time series of deaths in some groups, especially for the Indian population group, for whom the level of reporting of population group appears to have fallen and, therefore, adjustments for each population group[36] were derived separately.

To derive the adjustment, the deaths that would have been expected in each population group in the years 2011-2013 had no change in completeness of reporting of population group occurred, were projected on the assumption that the number of deaths occurring in each year was increasing exponentially (the growth rate was found by fitting an exponential regression to the deaths in 2000-2010 using least squares regression). The percentage change in deaths actually reported compared to the projected deaths was calculated, and the annual completeness of reporting was increased

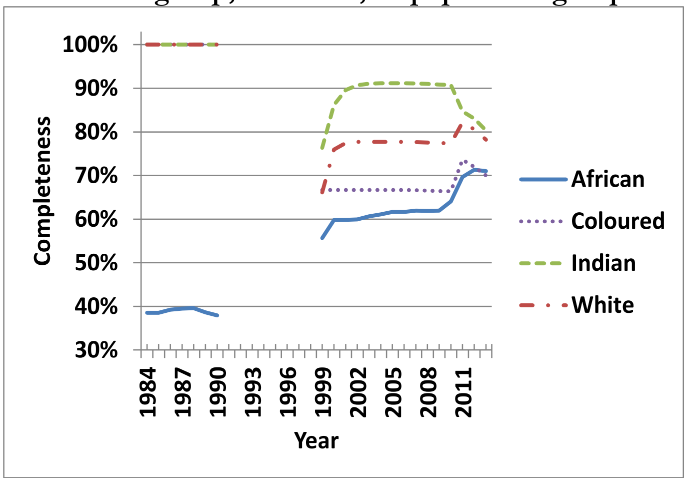
[36] Coloureds, Indians and Whites.

by this percentage in the years 2011-2013. The percentage change in deaths reported with population group is 5 per cent, -6 per cent and 7 per cent for the Coloured, Indian and White population groups respectively.

Two assumptions were made to derive annual estimates for the African population group: that the estimates of completeness for the Coloured, Indian and White population groups are estimates of the completeness of reporting of the race of the deceased, and that reporting of deaths for these population groups is complete[37], and that for the African population group, the completeness of reporting of death, as well as reporting of race, are incomplete. These assumptions imply that the deaths from the African population group can be estimated as the difference between the estimated national deaths and the estimated deaths for the Coloured, Indian and White population groups. The ratio of the reported African deaths to the deaths estimated in the prior step is then taken as the annual estimate of completeness for the African population group. The estimates of completeness for the African population group produced in this manner (1999: 56 per cent and 2007: 62 per cent) lie close to the point estimates from the SEG method, adjusted to be relative to the 2011 census and averaged between males and females (1999: 57 per cent and 2007: 63 per cent).

The annual estimates of completeness after adjustments are given in Appendix B.2.5 and shown in Figure 3.19.

Figure 3.19: Annual estimates of completeness of reporting of deaths and population group, 1984-2013, all population groups



The annual time series of deaths produced by grossing up the deaths using these estimates of annual completeness are shown in Figure 3.20 and Figure 3.21. The level of

[37] This assumption is reasonable, since these populations are highly urbanised.

the series of annual deaths cannot be assessed from these figures, but, as discussed in section 3.3.6.1, the ranking of the mortality rates are sensible for the different population groups, implying that the adjustments made are of the correct relative magnitude. The series of the corrected number of deaths in the African population group conforms closely to the series at a national level, which is sensible since it was assumed that reporting of deaths is incomplete only for this group. The results for the Coloured, Indian and White population groups appear reasonable in most of the years. The drop-off in deaths of Whites in recent years has been corrected and the time series is smooth. For the Coloured population group, the correction applied for increased reporting of population group appears to be too extreme in 2011. For the Indian population group, the drop-off of deaths in 2012 and 2013 has been partially corrected. Further attempts to correct the deaths would need to involve arbitrary adjustments and, therefore, the annual estimates of completeness were accepted[38].

Figure 3.20: Numbers of deaths at ages 30+, 1999-2013, National and African

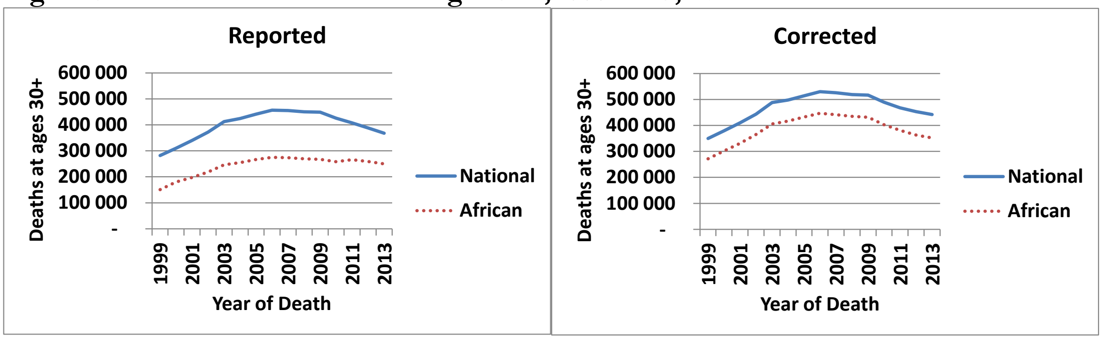
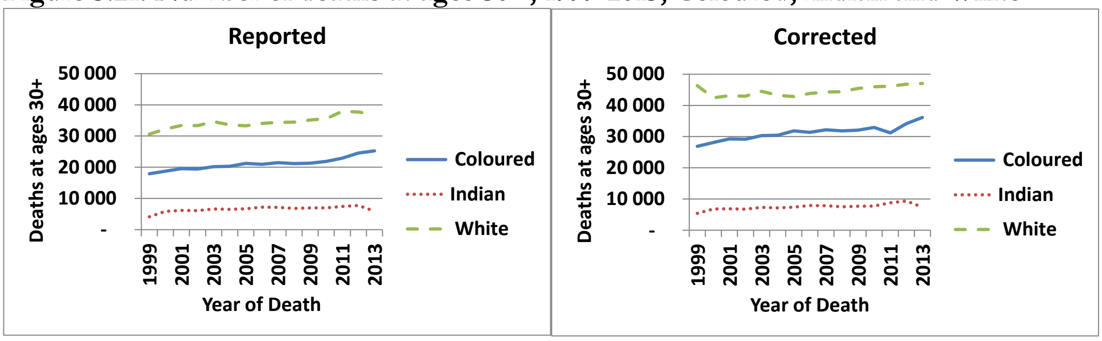


Figure 3.21: Number of deaths at ages 30+, 1999-2013, Coloured, Indian and White



[38] Further adjustments were investigated statistically, as reported in the next chapter.

3.3.9 Comparison of the estimates to other studies

3.3.9.1 National estimates

The estimates presented above are compared with those from other studies in Table 3.16 and Figure 3.22. Table 3.16 presents point estimates of completeness and Figure 3.22 presents annual estimates of completeness (relative to modelled population estimates).

Since the results in Table 3.16 are relative to the enumerated population, it is only possible to compare results that are within the same intercensal period. The estimates for 1996-2001 from other sources lie in a tight range between 84 per cent-89 per cent, which includes the (unadjusted) SEG estimate for males and females. However, the adjusted national results are lower than all of the other estimates, but the various arguments presented above suggest that the adjusted estimates should be preferred over than the estimates calculated using the national data. The rest of the results fall into intercensal periods that are different from the studies conducted previously. Assuming that the 2007 Community Survey numbers did not have a very different coverage from the 2011 census, then the results in the periods 2001-2007 and 2007-2011 should be roughly comparable to the estimates derived above for the period 2001-2011. The estimates from other studies fall in the range 85-94 per cent, which includes both the unadjusted and adjusted SEG estimates.

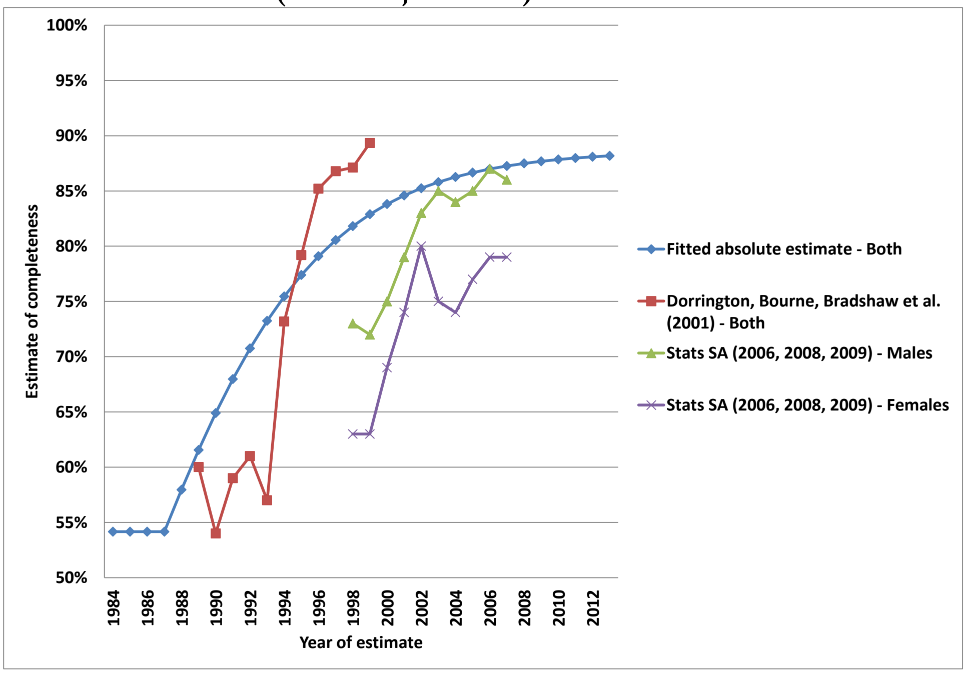
Therefore, the application of the DDMs nationally has produced results consistent with the consensus in the literature and the reasons for the differences between the adjusted national estimates and the consensus in the literature have been explained in section 3.3.5.

The results in Figure 3.22 are comparable to each other over time. Except for the simplifying assumption of a smooth curve, the results derived in this chapter are close to the estimates for males collected from the Stats SA (2006, 2008, 2009) reports on deaths and the estimates in Dorrington, Bourne, Bradshaw *et al.* (2001). Differences between the sets of estimates are likely due to methodology and data: the Stats SA estimates were calculated by applying the Preston and Hill (1980) method using the official mid-year population as the reference population (and therefore depend on the assumption of stability made in the Preston and Hill (1980) method and the quality of the mid-year estimates) and the estimates in Dorrington, Bourne, Bradshaw *et al.* (2001) were calculated by comparing an estimate of registered deaths derived by scaling up those recorded on the population register maintained by the Department of Home Affairs to the number of deaths predicted by the ASSA model (because they didn't have data from the 2001 census needed to apply the DDMs), whereas more up to date vital registration and census data was used to calculate the estimates in this chapter.

Table 3.16: Comparison of estimates of intercensal completeness of reporting of deaths, 1985-2011

Source	Sex	1988	1994	1999	2001	2004	2006	2007	2009
Dorrington, Moultrie and Timæus (2004)	*Males*			84%					
Dorrington, Moultrie and Timæus (2004)	*Females*			87%					
Machemedze (2009)	*Males*			85%		88%			
Machemedze (2009)	*Females*			88%		94%			
Stats SA (2014)	*Both*			89%		93%			94%
Dorrington and Bradshaw(2011)	*Females*				89%	85%	90%		
National	*Male*	57%	80%	85%				93%	
National	*Female*	61%	78%	89%				95%	
National - adjusted	*Male*			74%				87%	
National - adjusted	*Female*			83%				88%	

Figure 3.22: Comparison of annual estimates of completeness of reporting of deaths, 1985-2011 (before adjustments)



3.3.9.2 Population group estimates

The estimates derived for the period 1996-2001 are compared in Table 3.17 to the estimates of completeness for the corresponding period in Dorrington, Moultrie and Timæus (2004). The estimates are close, except for African and Indian males. The majority of the difference in the results is likely to be because Dorrington, Moultrie and

Timæus used a different dataset, since the data used in their study was a sample of the vital registration data scaled up to approximate the full death dataset, whereas this study has utilised the full dataset, including late registrations up to 2013. Other less important reasons for the different results are a different method of calculating completeness (SEG versus GGB), a different method of fitting the GGB method using orthogonal instead of linear regression or different estimates of migration.

It is thus reasonable to conclude that the results of this study are more accurate since the results were derived from a more complete set of death data, using the optimal version of the DDMs according to recent research.

Table 3.17: Comparison of results to Dorrington, Moultrie and Timæus (2004)

	African Males	African Females	Coloured Males	Coloured Females	Indian Males	Indian Females	White Males	White Females
Results	56%	64%	66%	69%	74%	82%	72%	75%
DMT(1999)	64%	67%	70%	70%	65%	83%	77%	79%

3.4 Conclusion

Estimates of the completeness of reporting of death and of population group have been produced in this chapter to be used to correct the death data before the NEG methods were applied to the corrected data.

It has been necessary to make several assumptions to derive annual estimates of completeness – that the completeness for the year in the middle of an intercensal period is equal to the average level of completeness for that period estimated using the DDMs, that the annual estimates of completeness should follow a smooth (logistic) curve, and that the adjustments for late reporting and falling completeness derived at a national level are suitable to apply for each population group. Whether these assumptions produce reasonable estimates of annual completeness will be tested statistically using the NEG-GAM model fit in the next chapter.

4 NEAR EXTINCT GENERATIONS METHODS

4.1 Introduction

The previous chapter has described how the DDMs were used to correct the death data for incompleteness of reporting. If completeness of registration was the same for all ages, the corrected deaths could be used together with estimates of the population from the censuses to calculate mortality rates at the older ages (and, indeed, these results are shown in Figure 3.11 in the previous chapter). However, as it is possible that estimates of mortality produced using census data are severely biased at the older ages due to age exaggeration, this chapter describes how the NEG methods are applied to the death data, corrected using the DDMs for incompleteness of reporting, to produce population estimates corrected for age exaggeration and derive mortality rates.

The chapter begins by describing how the death data were split into cohorts aged x last birthday on the previous 10 October (10 October was chosen to facilitate comparison with the census data from 1996, 2001 and 2011).

The next section discusses how NEG methods were applied to the death data corrected for incompleteness of reporting. This section shows the results of applying the Survivor Ratio method for all population groups combined and argues that a different model is needed for the South African data.

The section following this uses an insight from the actuarial literature to show that the Das Gupta method can be viewed as a regression model (an over-dispersed Poisson generalised linear model). A regression model that incorporates smoothing (a generalised additive model) is proposed. This model allows some of the deficiencies identified in the death data to be modelled and adjusted for explicitly. Finally, the reconstructed population and estimates of mortality and life expectancy based on this new model are presented.

4.2 Splitting the death data into cohorts

Since it is suspected that the deaths may be influenced by year of birth heaping, the assumption of a uniform distribution of deaths over the Lexis diagram is not appropriate. The death data in 2012 and 2013 include information on the date of birth of the decedent, in addition to the date of death, allowing the deaths to be classified into cohorts aged x last birthday on a prior date. The effect of year of birth heaping is captured by using the death data in 2012 and 2013 to calculate the proportional split of deaths aged x last birthday into cohorts aged x last birthday on a prior date. On the assumption that this proportional split of the deaths remains the same for each cohort over time (which is reasonable if the split is influenced by year of birth heaping), the death data prior to 2012 can be split into cohorts.

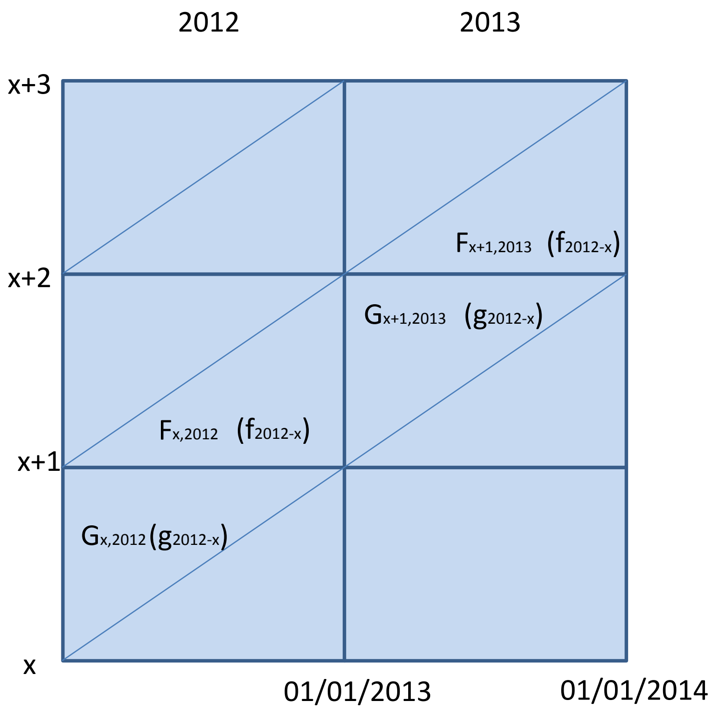
The population that is being re-estimated using the NEG methods will be compared to the censuses of 1996, 2001 and 2011, in which age is reported as age last birthday on 10 October. Therefore, cohorts defined as those aged x last birthday on 10 October of year t will be used in the rest of this study.

For ease of exposition, the simple case of cohorts defined as those aged x last birthday on 1 January of year t is treated first, in the following section. Then, in the subsequent section, the case of cohorts aged x last birthday on 10 October of year t is examined.

4.2.1 Cohorts aged x last birthday on 1 January

A Lexis diagram of the death data in 2012 and 2013 is shown in Figure 4.1.

Figure 4.1: Lexis diagram of deaths occurring in 2012 and 2013, with cohorts defined from 1 January



The data can be grouped as follows:

- $F_{x,2012}$ is the number of deaths occurring in 2012 aged $x+1$ last birthday at death who were also aged x last birthday on 1/1/2012.

- $G_{x,2012}$ is the number of deaths occurring in 2012 aged x last birthday at death who were also aged x last birthday on 1/1/2012.

The groups $F_{x,2012}$ and $G_{x,2012}$ lie on the diagonal $t-x$ of the Lexis diagram. The deaths in 2012 attributable to the cohort aged x last birthday on 1 January 2012 are $G_{x,2012}+F_{x,2012}$. Let f_{t-x} and g_{t-x} represent the proportions of the deaths in 2012 and 2013 aged $x+1$ and x last birthday at death that are attributable to the groups $F_{x,t}$ and $G_{x,t}$ respectively, as shown in Figure 4.1.

Also, let $D(x,t)$ represent the number of deaths in calendar year t aged x last birthday at death. Since it is suspected that these proportions are primarily influenced by year of birth heaping, it is assumed that f_{t-x} and g_{t-x} are unchanging over time and therefore, for each cohort, the proportions can be estimated as

$$f_{2012-x} = \frac{F_{x,2012} + F_{x+1,2013}}{D(x+1,2012) + D(x+2,2013)} \text{ and } g_{2012-x} = \frac{G_{x,2012} + G_{x+1,2013}}{D(x,2012) + D(x+1,2013)}.$$

In years prior to 2012, which do not include data on date of birth, the number of deaths in the groups $F_{x,t}$ and $G_{x,t}$ cannot be calculated directly. However, on the previous assumptions, the cohort deaths in the years prior to 2012 can be approximated as $D'(x+t-2012,t) = g_{2012-x}.D(x+t-2012,t) + f_{2012-x}.D(x+1+t-2012,t)$, where $D'(x,t)$ represents deaths occurring in the year t to those aged x last birthday on 1 January of that year. Also, defining $y = 2012 - x$ then the previous equation can be written succinctly as $D'(t-y,t) = g_y.D(t-y,t) + f_y.D(t-y+1,t)$ for $t \geq y$. Once the deaths by cohort have been projected using the NEG methods, the population aged x last birthday on 1 January of calendar year t can then be reconstructed as

$$P'(t-y,t) = \sum_{i=0}^{\omega} D'(t-y+i,t+i).$$

4.2.2 Cohorts aged x last birthday on 10 October

The cohorts are defined in this section as those aged x last birthday on 10 October of year t. Since the majority of the deaths occur in the year $t+1$, deaths occurring from 10/10/2012 to 09/10/2013 are referred to as deaths occurring in the year 2013, as follows:

- $F_{x,2013}$ is the number of deaths in 2012 aged $x+1$ last birthday on death who were also aged x last birthday on 10/10/2012.

- $G_{x,2013}$ is the number of deaths in 2012 aged x last birthday on death who were also aged x last birthday on 10/10/2012.
- $H_{x,2013}$ is the number of deaths in 2013 aged x last birthday on death who were also aged x last birthday on 10/10/2012.
- $I_{x,2013}$ is the number of deaths in 2013 aged $x+1$ last birthday on death who were also aged x last birthday on 10/10/2012.

The deaths attributable to the cohort aged x last birthday on 10/10/2012 would be calculated as $F_{x,2013} + G_{x,2013} + H_{x,2013} + I_{x,2013}$. These groups are shown in Figure 4.2. The proportions f_{t-x}, g_{t-x}, h_{t-x} and i_{t-x} of the deaths in 2012 and 2013 attributable to the groups $F_{x,2013}$ to $I_{x,2013}$ are calculated on the assumption that the proportions f_{t-x} to i_{t-x} are unchanging over time.

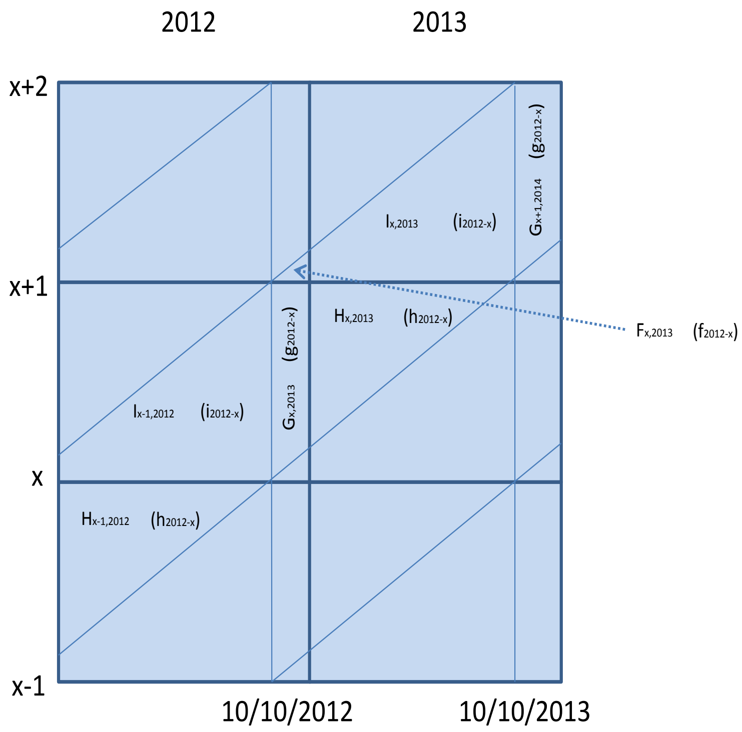
Let $D(x,t)$ represent the number of registered deaths in calendar year t aged x last birthday at death, then the proportion f_{2012-x} can be estimated from the period death data as $f_{2012-x} = \dfrac{F_{x,2013} + F_{x+1,2014}}{D(x+1,2012) + D(x+2,2013)}$ and similarly for the other proportions.

Let $D'(x,t)$ represent deaths occurring in the year t to those aged x last birthday at death on 10 October of that year and defining $y = 2012 - x$, then the cohort deaths in years prior to 2012 can be approximated as

$$D'(t-y,y) = f_y.D(t-y+1,t) + g_y.D(t-y,t) + h_y.D(t-y,t+1) + i_y.D(t-y+1,t+1)$$

and the population can be reconstructed as $P'(t-y,t) = \sum_{i=0}^{\omega} D'(t-y+i,t+i)$, once the future cohort deaths have been projected using the NEG methods.

Figure 4.2: Lexis diagram of deaths occurring in 2012 and 2013, with cohorts defined from 10 October



4.2.2.1 Results

As a crude measure of heaping, the sum of the proportions f_{t-x}, g_{t-x}, h_{t-x} and i_{t-x} for each age group x aged from 70-100 on 10/10/2012, that were estimated from the period death data for the years 2012 and 2013, are shown in Figure 4.3 to Figure 4.5. If the data were free of year of birth heaping, then the sum of these proportions would be close to one while results significantly different from one indicate that year of birth heaping occurs in the data.

The figures show that heaping is most pronounced for the African population group, followed by the Coloured and Indian population groups, with very limited heaping in the White population groups. Many of the highs nationally and for the African and Coloured population groups correspond to year of birth heaping ending in '0' and many of the lows appear to be due to too few being reported in the years adjacent to years ending in '0'. Heaping also occurs on the years 1914 and 1918 for African and Coloured population groups, and nationally.

A different pattern of heaping can be seen for the Indian population group in Figure 4.5 but the heaping appears to be at random and not according to a simple rule, as was the case nationally. Very little heaping occurs for Whites compared to the other population groups, except for around World War I, when too many appear to be recorded in the years 1914 and 1916 and too few are recorded in 1915 and 1917.

Figure 4.3: Heaping in death data, National, 2012 and 2013

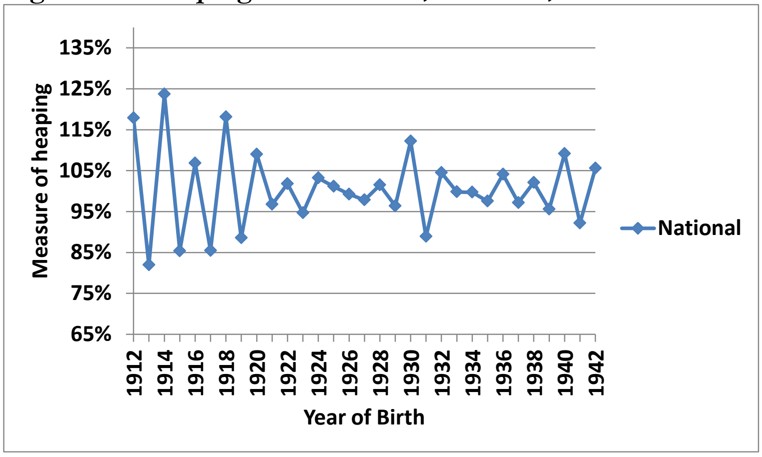


Figure 4.4: Heaping in death data, Africans and Coloureds, 2012 and 2013

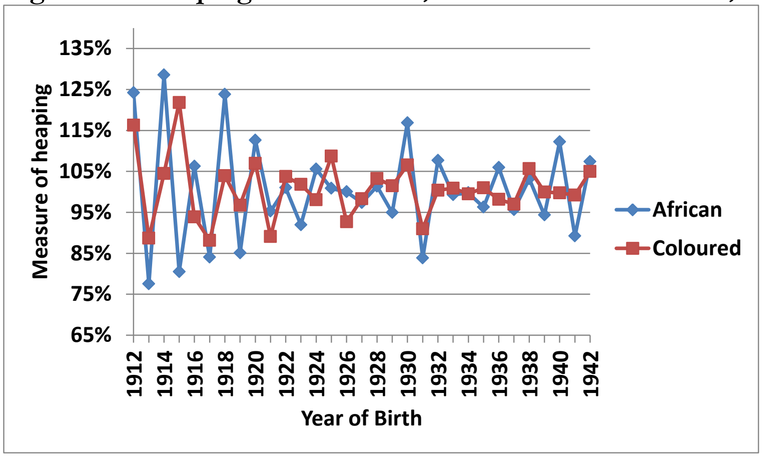
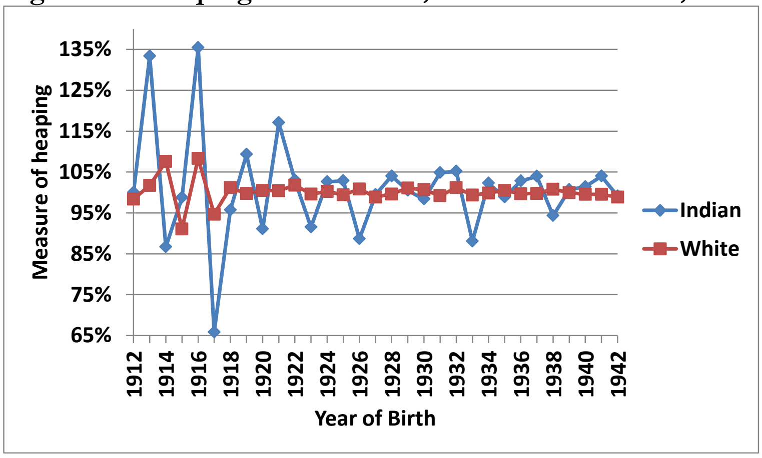


Figure 4.5: Heaping in death data, Indians and Whites, 2012 and 2013

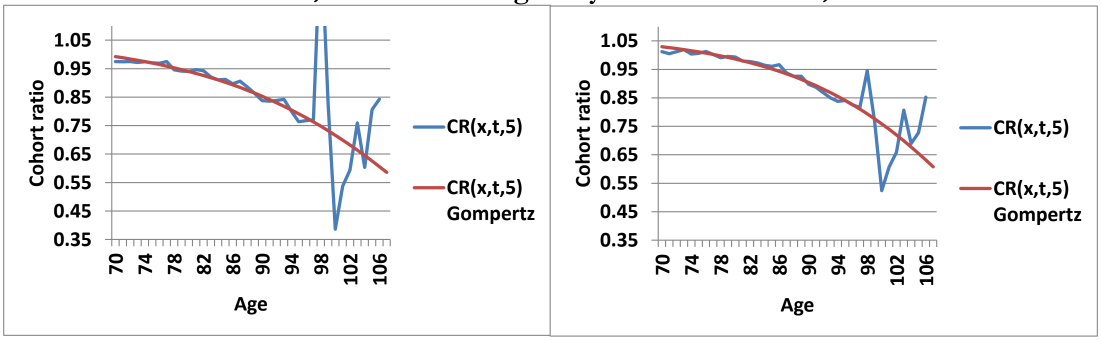


4.3 Application of the Survivor Ratio method to National Data

The results of applying the Survivor Ratio method[39] to the national data are discussed in this section. Population projections suggest that age exaggeration only begins from age 75 in the South African data (Machemedze 2009). To provide a margin for error, and since the age exaggeration in the more recent census data may have worsened, the Survivor Ratio method was applied from age 70 and older.

Cohort ratios were calculated using the previous five years of death data (2008-2012) as suggested by Thatcher, Kannisto and Andreev (2002). The cohort ratios, shown in Figure 4.6 for males and females, were distorted at the ages 97-106 with the ratios firstly rising dramatically and then falling (instead of decreasing steadily with age with minor fluctuations as would be expected).

Figure 4.6: Cohort Ratios and Cohort Ratios smoothed using a Gompertz curve, Males and Females, calculated using five years of death data, 2008-2012



These distortions are caused by severe age heaping in the death data at age 100 (and ages 99 and 101 for males), as can be seen in Figure 4.7. When the period death data are converted into data by cohorts, the excess deaths recorded at age 100 are distributed to the ages adjacent to 100, and these excess deaths distort the cohort ratios from age 98 to age 102. Because of the age heaping in the death data, mortality rates calculated using these data will still suffer from age heaping. To illustrate, raw mortality rates were estimated using the NEG exposures and are shown in Figure 4.8, where it can be seen that the rates at ages 90 and above are very erratic.

Another problem is that the mortality rates flatten at ages 90-100, which can be understood in terms of the relationship between mortality rates and the cohort ratios discussed in section 2.3.2.1, $\hat{q}_{x,t} = \left(1 + \sum_{i=2}^{\omega-x+1} \prod_{j=2}^{i} CR(x+j-1, t+j-1, m)\right)^{-1}$.

If one of the cohort ratios is exaggerated as a result of age heaping, then the denominator of this expression will be too large at all ages younger than the age at which the heaping occurs, producing mortality rates that are too low at these ages. Therefore, the age heaping must be smoothed before the NEG methods can be applied successfully. The flattened mortality rates produced using the NEG methods in the study by Machemedze and Dorrington (2011) are perhaps due to this problem.

Figure 4.7: Total deaths in 2005-2012 at age last birthday at death and at age last birthday on 10 October, 95-109, Males and Females

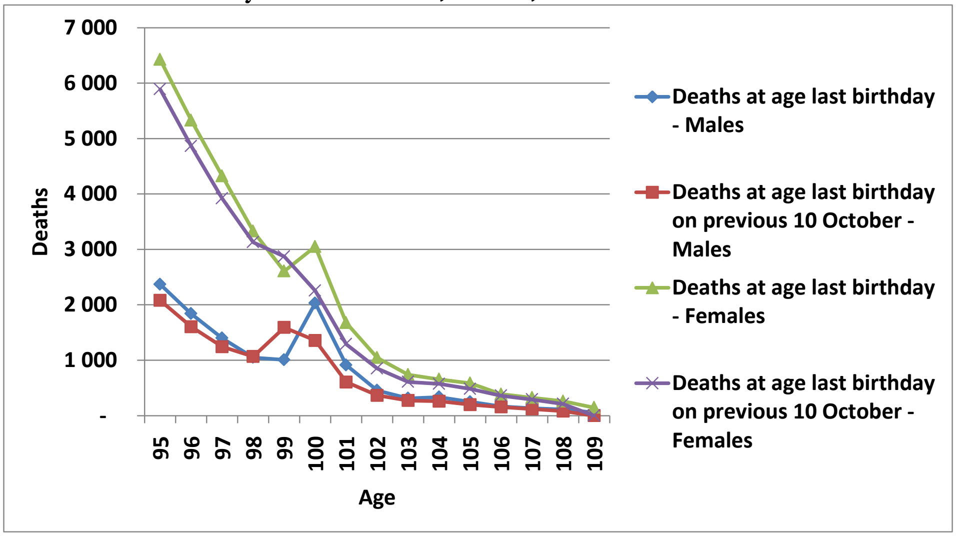
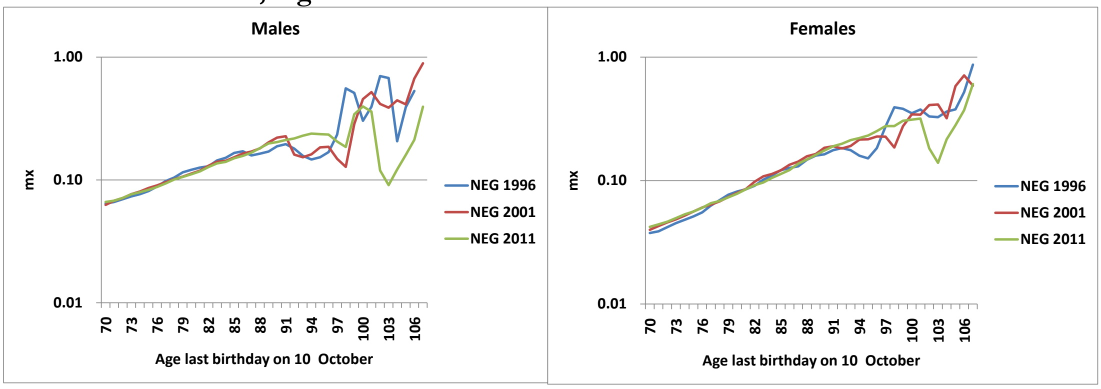


Figure 4.8: m_x implied by the NEG estimates in the years 1996, 2001 and 2011, Males and Females, log scale

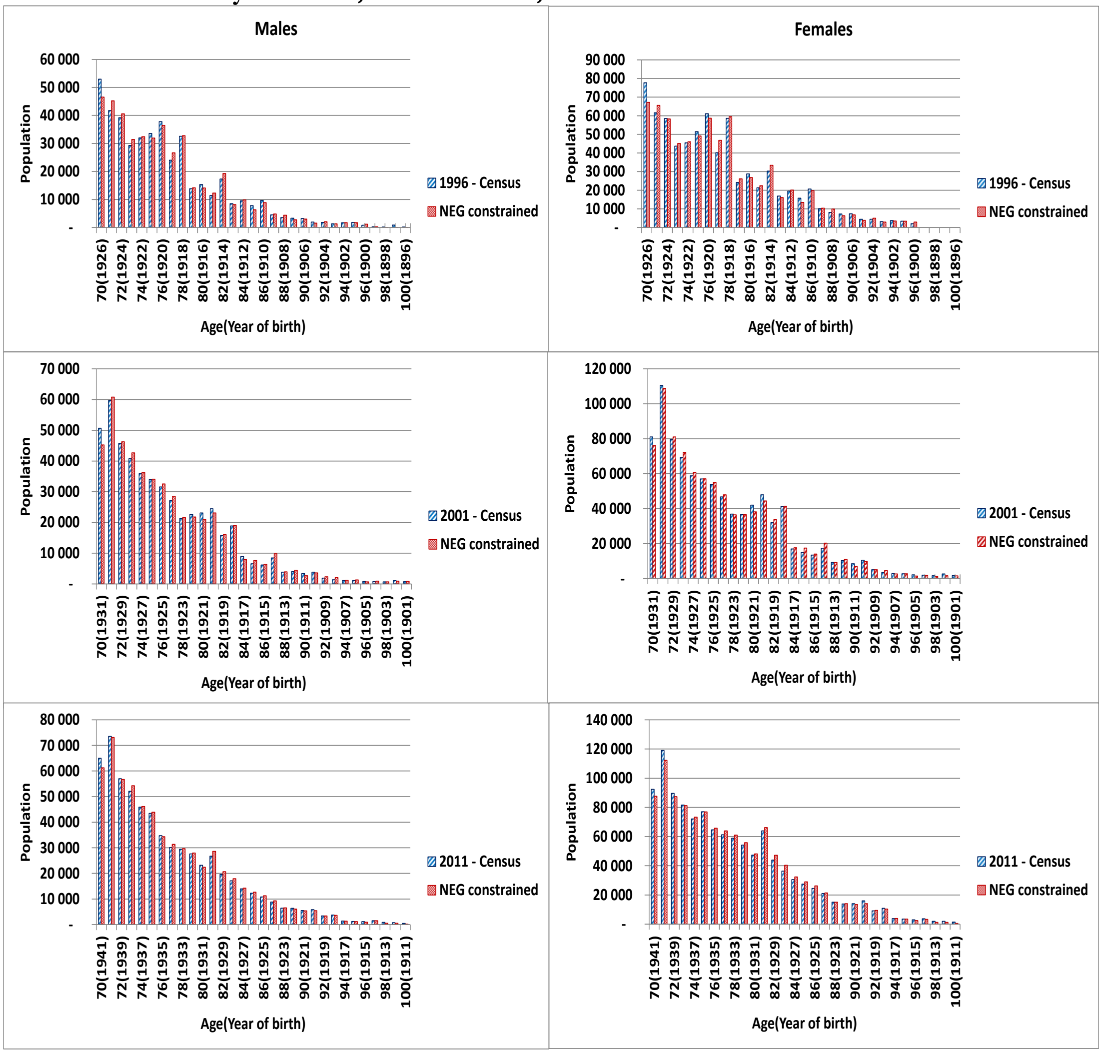


The extent of the age heaping in the death data was assessed by comparing the NEG population estimates by age to the census estimates. To provide a clear comparison of the age structure of the estimates to the census data, the NEG results were constrained to be equal to number of people aged 70+ counted in each of the

[39] The Survivor Ratio method was applied in two steps – the death data were smoothed and then projected, using the formulae in Appendix A.4. The Das Gupta method was also tested on these data and produced very similar results (which are not reported on in the text).

three censuses. These results are shown in Figure 4.9. For both males and females, the NEG estimates and the census counts have very similar age structures that suffer from pronounced year of birth heaping, indicating that the death data is affected by the same form of misreporting as the census data.

Figure 4.9: Comparison of census data and NEG estimates (constrained to match 70+) in the years 1996, 2001 and 2011, Males and Females



As shown above and noted by Bradshaw, Schneider, Laubscher *et al.* (2002) and Machemedze (2009), the heaping is apparent at ages implying a year of birth that ends in a zero. Other years which suffer from heaping are 1914 and 1918, which mark the start and end of World War I. Since death registration in South Africa depends on an identity number containing year of birth, the year of birth heaping must be attributed to the issuance of inaccurate birth certificates (and is not only an artefact of the census enumeration).

A possible solution to the problem of age heaping is to smooth the cohort ratios. The cohort ratios can be expressed in terms of the probability of death at age x during year t, as $CR(x,t,1)=\frac{D_{x+1,t+1}}{D_{x,t}}=\frac{N_{x,t}(1-q_{x,t})q_{x+1,t+1}}{N_{x,t}q_{x,t}}=\frac{(1-q_{x,t})q_{x+1,t+1}}{q_{x,t}}$, allowing a mortality law, such as that proposed by Gompertz (1825), to be fitted to estimates of the cohort ratios. Gompertz curves were fit to the cohort ratios from age 70 up to age 97, the age above which the ratios were distorted, as shown in Figure 4.6. Smoothing using the Gompertz curve can be seen from Figure 4.6 to provide a good fit to the cohort ratios until the age of 97, after which the extrapolated curve provides a series of ratios that is not affected by the heaping at the ages around 100. However, a key disadvantage of smoothing in this way is that mortality calculated using the deaths projected using these ratios will follow the Gompertz mortality law, obscuring the true shape of the mortality rates implied by the data.

This solution will also not solve the bias caused by age heaping. Mortality rates estimated from death data suffering from year of birth heaping will probably be biased, since the age at death of those whose year of birth has been misreported is misstated, and therefore year of birth heaping should also be smoothed before applying the NEG methods. Notably, the Survivor Ratio method does not smooth the year of birth heaping. As shown in Appendix A.4.2, the Survivor Ratio method smooths the death data within each cohort but does not redistribute deaths to different cohorts, and therefore year of birth heaping will not be corrected by the Survivor Ratio method.

In summary, the Survivor Ratio method as applied above does not produce sound estimates of mortality, due to age and year of birth heaping in the national death data. To overcome these difficulties, a new NEG model that solves these problems is proposed in the next section.

4.4 The NEG-GAM Model

4.4.1 Introduction

In order to address the problems with applying the NEG methods to data suffering from digit preference, a statistical variation of the method is proposed in this section to smooth the data. The key advantage of this approach is that the effects of age and year of birth heaping, as well as other deficiencies in the data, can be accommodated explicitly.

In the previous chapters, it has been shown that the Survivor Ratio method can be applied in two steps – smoothing the most recent death data within each cohort and projecting the smoothed deaths. In the following, a statistical model that smooths the death data over different cohorts and ages and projects the future deaths expected for each cohort is proposed as an alternative to the Survivor Ratio and Das Gupta methods.

4.4.2 Generalised Additive Model (GAM)

An algorithm that is analogous to the Das Gupta method is well-known and widely used by actuaries to project outstanding insurance claims arising from each year of an insurance company's operations. This algorithm is referred to as the chain ladder method and is described in detail in Taylor (2012) and Wüthrich and Merz (2008). Since the Das Gupta method is essentially the same deterministic approach as the chain ladder method[40], insights gained from the actuarial literature into the chain ladder method can be applied to formulate a statistical model for projecting the deaths.

The chain ladder method has been recast as a statistical model by Renshaw and Verrall (1998). They show that a generalised linear model (GLM) in which the response is modelled as an over-dispersed Poisson (ODP) random variable, and the predictors are the year of accident of the claims and the time elapsed since the occurrence of the claim, produces the same results as the chain ladder method. The GLM model was further developed by Verrall (1996), who proposed using a generalised additive model (GAM) (Hastie and Tibshirani 1990) that incorporates a smooth function of the year of accident of the claims, instead of a GLM.

In an analogous way, the death data can be modelled using a GAM that incorporates a smooth distribution of the population and smoothed mortality rates that change over time as follows:

$$D_{x,t} \sim ODP(e^{\eta_{t-x,t}})$$
$$\eta_{t-x,t} = f(t-x) + f(x) + f(x)*t + \alpha_{t-x} + \beta_x + \delta_t$$

where $t-x$ is the year of birth of the cohort and x is the age of the cohort. Instead of including these covariates directly in the model, which would result in a linear relationship between the predictor and the covariate (i.e. a GLM), a smooth function $f(.)$ of the covariate is instead used to create a smooth non-linear relationship.

The smooth function can be semi-parametric (for example, splines) or non-parametric (for example, a LOESS regression). Non-parametric smoothers are not suitable for our purposes (since we do not wish to fit the deficiencies in the data) so a less flexible semi-parametric function - b-splines with 3 degrees of freedom – is used to provide more control over the smoothing (Chambers and Hastie 1991: 270)[41]. The first term of the model, $f(t-x)$, models the size of the cohort born in the year $t-x$, and the second term, $f(x)$, models the deferred probability of death, ${}_{t|}q_x$ [42]. The interaction term $f(x)*t$ models the period trend (if any) in mortality over time[43].

α_{t-x}, β_x and δ_t represent optional parameters that control for deficiencies (year of birth heaping, age heaping and variations in the level of completeness, respectively) in the death data. For example, consider heaping on the year of birth 1940 in the national data. The smooth function representing year of birth, $f(t-x)$, is not flexible enough to reproduce the higher than expected number of deaths in this cohort, and the fitted parameters of the model will be somewhat biased as a result. An extra variable α_{1940} is then included in the model so that the deaths in the 1940 cohort will be modelled as a combination of the baseline smooth function of year of birth plus the extra variable, and the fit of the smooth function will not be biased by the heaping in the data. To the extent that the control variables prevent bias in the estimated smooth functions, the estimates of mortality will also be unbiased.

This model is referred to in the rest of this research as the NEG-GAM model.

4.4.3 Comparison to other NEG methods

Compared to the Das Gupta and Survivor Ratio approaches, which use a constant set of cohort ratios to project the deaths, the GAM model explicitly allows for changing mortality rates (which could be improving or worsening at different ages) over time

---

[40] The Das Gupta method projects deaths at age $x+1$ based on the deaths at age x. The chain ladder method is generally applied to cumulative claims data, or in other words the sum of claims paid until reporting period $t+1$ is projected based on the sum of claims paid until reporting period t. The chain ladder method could also be applied to incremental claims data, or in other words the claims paid in reporting period $t+1$ could be projected based on the claims paid in reporting period t. Thus, the chain ladder method applied to incremental claims data is exactly equivalent to the Das Gupta method. Projecting the cumulative deaths at age $x+1$ based on the cumulative deaths at age x is a possible variation on the Das Gupta method that could be more stable at the oldest ages.

[41] As alternative to the b-spline basis, a fully non-parametric model using penalised thin-plate splines with the amount of smoothing set by generalised cross-validation was fit using the mgcv package (Wood 2015). The mortality rates produced by this model were distorted at ages 90-100 in a similar way to the distortions in the mortality rates produced by the Survivor Ratio method. The less flexible b-spline model does not fit the deficiencies in the data as closely and therefore produces more demographically reasonable mortality rates.

[42] ${}_{t|}q_x = {}_tp_x q_{x+t}$

[43] To continue with the analogy to loss reserving, in the same way that actuaries reserving for claims may allow explicitly for claims inflation, for example, as in the model of Zehnwirth (1994), so too a calendar year trend in mortality is allowed for in the NEG-GAM model. Different variations on this model can be specified, for example, a trend in cohort mortality could be specified by adding the term $f(x)*f(x-t)$ to the model.

(although the GAM model does not project the deaths using cohort ratios, the cohort ratios can be derived from the modelled ${}_{t|}q_x$). The GAM model is therefore conceptually similar to both the Mortality Decline method of Andreev (1999) and the Survivor Ratio Advanced method of Terblanche and Wilson (2015), in that changing mortality is allowed for explicitly.

The GAM model is also conceptually similar to the Das Gupta Advanced (DGA) method of Andreev (2004) which allows for changing cohort ratios over time by adding a correction term to the Das Gupta cohort ratios, however, in the DGA method the correction term is parameterised by minimising the difference between the enumerated population and the Das Gupta method whereas the GAM model is fit without reference to the population data. The GAM model implicitly implements the suggestion of Andreev (2004) to model the cohort ratios using splines.

The GAM model bears a close resemblance to the regression model of Siegel and Passel (1976), who model aggregate deaths using a linear regression. The GAM models the deaths in each year directly as a regression on year of birth and age, and mortality estimates are produced as a by-product of this regression.

Compared to the other models, the key advantage of the model proposed above for the South African data is the ability to control for deficiencies in the data, including diagnosing and correcting any incompleteness of reporting remaining after the application of the DDMs. Another advantage of the model is that it produces smooth estimates of mortality directly from the death data, whereas estimates of mortality produced using the other NEG models will generally need to be graduated. The spline used to model mortality is flexible enough to capture various shapes of the mortality curve, allowing the hypothesis of exponentially increasing mortality to be investigated.

A disadvantage of the NEG-GAM model is the risk that real effects in the data are smoothed out by the spline smoothing and not enough corresponding control variables are added to capture the effects. For example, a particular cohort may be larger (or smaller) than the average cohort predicted by the spline, and if a control variable is not added to the model, then that cohort will not be modelled accurately. Similarly, the trend assumed in the model is linear, and if rates have changed in a non-linear manner, then the trend will not be modelled correctly. These risks are partially mitigated by the fitting process, during which residual plots are examined for evidence of a poor fit and control variables are added to the model, as described in Section 4.4.5. However, some risk remains that the model does not represent the data adequately.

4.4.4 Validating the adjusted death data using the NEG-GAM model

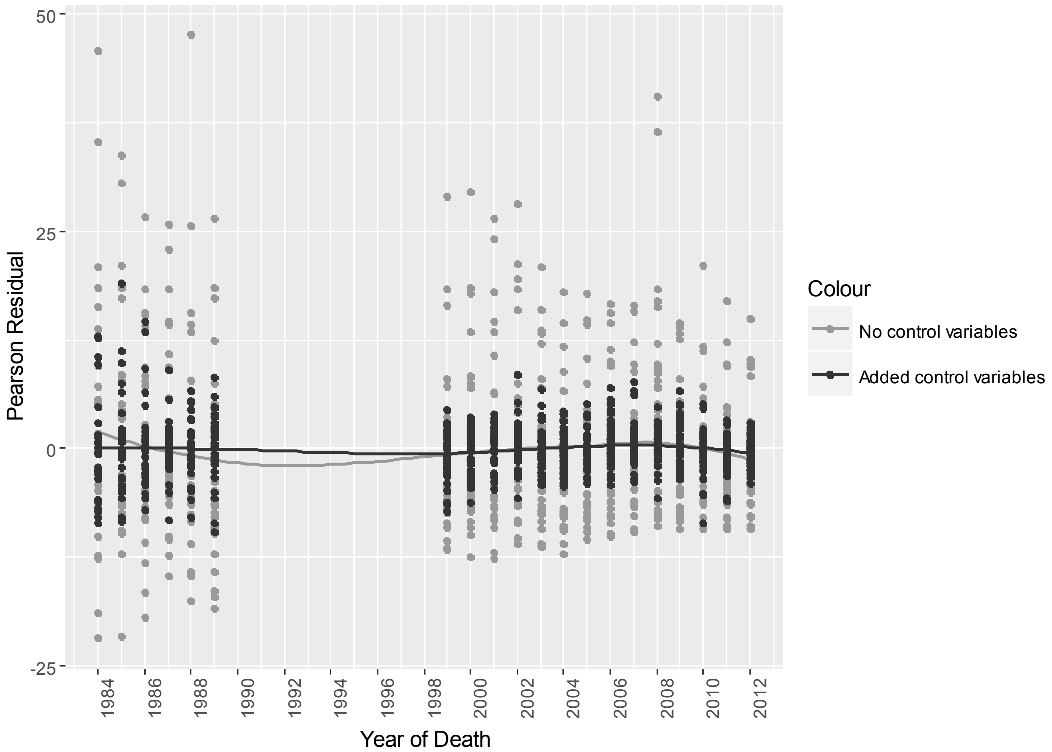
After the death data have been corrected using the annual estimates of completeness, the data should all be completely reported relative to the census 2011 data (i.e. have a constant level of completeness of 100 per cent). The NEG-GAM model takes population growth and trends in mortality into account, so the year in which death occurs should not be a statistically significant predictor of the number of deaths occurring in that year. Plotting the residuals of the model against year of death allows this to be tested. If the plots show that year of death nonetheless remains a significant predictor of the number of deaths in the NEG-GAM model, it can be inferred (barring any unusual events such as a harsh winter or a war) that the deaths in that year are not at 100 per cent completeness. For example, if a group of residuals related to a particular year is much larger or smaller than the average, then this is an indication that the corrected deaths in the particular year are not at the same level of completeness compared to the rest of the corrected deaths. Including the year of death as a control variable in the model allows the (relative) level of completeness to be estimated statistically. In this way, the NEG-GAM model permits the validation of the adjustments made to the death data in a statistically sound manner.

4.4.5 Fitting the model

The GAM model was fitted separately for males and females for each population group and for the population groups as a whole using the glm( ) function in R (R Development Core Team 2016). The models were first fit without controlling for any data deficiencies. The model residuals were then inspected by plotting the residuals and the smoothed average residuals (calculated using LOESS regression (Cleveland, Grosse and Shyu 1992) against year of death, year of birth and age last birthday. Control variables were added to the regression where visual inspection of the residual plots indicated that the model did not fit the data well; in other words when the assumptions of constant completeness of reporting, smooth birth cohorts and smoothed mortality rates appeared to be violated. A poor fit was indicated by groups of residuals that appeared to be much larger or smaller than the average or by the smoothed average residuals deviating significantly from zero. No attempt was made to correct for all of the data deficiencies since this would result in an over-parameterised model. Rather, control variables were added until the residual plots showed that the models fit well and the models reproduced the population in the open interval 70+ in 2011 as closely as possible.

An example of this model fitting is shown in the plots of the Pearson residuals ($\varepsilon_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}$, where ε_i represents the Pearson residual for observation i, y_i is the observed value and $\hat{y}_i$ is the fitted value) against year of death in Figure 4.10. Before adding control variables, the residuals are widely dispersed, indicating a poor model fit, and the smoothed average residuals show a trend by year of death. After controlling for data deficiencies, the residuals lie in a smaller range and the smoothed average residuals show less trend[44]. The improvement in the model fit is also indicated by the change in the residual deviance of the model, from 52 897 before control variables were added to 6 657 after control variables were added. This is a highly significant reduction according to the F-test (p-value = 0.000).

Figure 4.10: Residual plot, African Males, 1984-2012, before and after controlling for data deficiencies



The fitted coefficients of the splines are given in Appendix C.1.1 and the values of the control variables, α_{t-x}, β_x and δ_t, are shown in Appendix C.1.2.

[44] The missing data in the years 1990-1998 do not influence the smoothed average residuals, which are estimated by fitting a LOESS regression of the residuals for the remaining years of death.

The residuals plotted against year of death were generally distorted in the years 1988 and 1989 for deaths reported with population group, and the interpretation of the control variables fit in the model in these years is that completeness of reporting was lower in these years than was assumed when correcting the death data using the annual estimates of completeness derived from the DDMs. In the years 1989 to 1994, completeness at a national level was also lower than implied by the annual estimates of completeness. Since the results of the DDMs are derived from the average completeness in the periods 1985-1989 and 1991-1996, it is possible that the years 1989-1994 had a lower level of completeness than the average in either period. Lower completeness of reporting than assumed was found in 1999, nationally, for males from the African population group and for Coloureds of both sexes. Completeness appeared to be lower than estimated in the previous chapter in the years 2010 to 2012 nationally and for the African population group, 2010 and 2012 for the Indian population group and 2012 for females from the White population group, showing, perhaps, that the adjustments made for late reporting of deaths and the fall in completeness in recent years were not significant enough. The assumptions made when deriving annual estimates of completeness from the results of the DDMs have therefore been tested and, where necessary, modified on a statistical basis when fitting the NEG-GAM model.

Evidence of year of birth heaping on years ending in a zero as well as in 1914 and 1918 was found at a national level, for the African population group and with less pronounced effect for the Coloured population group. Fewer deaths were attributed to the cohorts in the years 1915-1917 than would be expected, for both males and females. For the White population group, fewer deaths than expected were attributed to the years 1915 and 1917-1919, possibly due to low birth rates during the First World War (see Thatcher, Kannisto and Andreev (2002: 13)).

Age heaping appears to have occurred at the ages 98-101, nationally and for Africans, Coloureds and Indians.

4.5 NEG-GAM results

The NEG-GAM models were used to estimate the deaths occurring at each age from 70 to 120 for each of the years of birth represented within the data. The population of each year of birth cohort, at each age, was derived as the sum of the deaths projected for that cohort, occurring at that age and all older ages (i.e. using Vincent's identity). Mortality rates were derived using the model by projecting the deaths using only the smooth functions (i.e. the control variables were not used when predicting the deaths for this purpose). The smoothed deaths and estimates of the population derived from these deaths were used to derive q_x, and m_x was then approximated as $-\ln(1-q_x)$. These mortality rates are for those aged x on 10 October of year t. Rates for those dying aged x last birthday in year t were derived by firstly interpolating between the rates for years $t-1$ and t and then averaging the interpolated rates for ages $x-1$ and x. The final rates for each of the population groups and sexes are shown for single years of age in Appendix C.1.3.

In the following sections, the unconstrained population estimates derived using the NEG-GAM models are compared to the census data and population estimates in the open interval 70+. The smooth age distributions of the population produced by the NEG-GAM method are compared to the results of the censuses of 1996, 2001 and 2011 to quantify the extent of misreporting in the census data. For the sake of comparability, the smooth NEG-GAM estimates were constrained to be equal to the census population aged 70+ for all three censuses and then the NEG-GAM results were compared to the census data[45].

To check if the mortality rates derived using the NEG-GAM are reasonable, the NEG-GAM mortality rates are compared to the mortality rates in the HMD. Coale and Kisker (1986), when comparing mortality rates in developing countries with rates in developed countries, found that mortality rates in developing countries are often lower at the older ages than mortality rates in developed countries. They attributed this finding to age exaggeration in the population data. In South Africa, it could be expected that the true mortality rates at the older ages at a national level, and for the African and Coloured population groups, should be higher than most of the mortality rates captured

[45] The ratios of the smooth NEG-GAM estimates to the census numbers in the interval 70+ are as follows:

Sex	Group	1996	2001	2011
Male	National	104%	100%	97%
Male	African	98%	95%	94%
Male	Coloured	101%	88%	96%
Male	Indian	95%	97%	101%
Male	White	103%	102%	95%
Female	National	106%	97%	92%
Female	African	96%	87%	92%
Female	Coloured	103%	91%	100%
Female	Indian	104%	101%	95%
Female	White	106%	107%	96%

in the HMD, which are for highly developed countries, while mortality rates falling beneath the lowest rates in the HMD would indicate age exaggeration in the death data and would not be an accurate reflection of mortality rates at these ages. To perform this comparison, the lowest and highest values of m_x from the HMD were calculated for each age from 70-109 in the years 2009-2011[46], where the minimum and maximum were estimated over all of the countries appearing in the database. These values and the average value of the mortality rates in the HMD are then compared to the estimated mortality rates using the NEG-GAM.

Lastly, life expectancy at age 70 calculated using the NEG-GAM mortality rates is compared to other sources to check the NEG-GAM rates for reasonability and shed light on the progression of mortality over time.

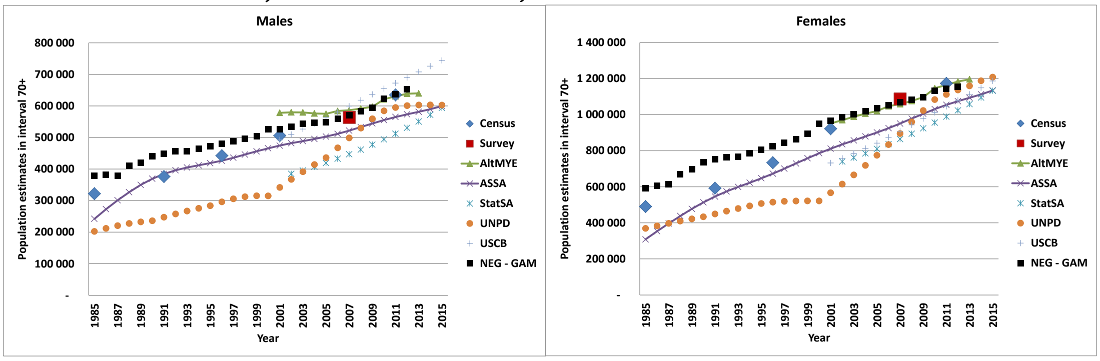
4.5.1 National

National population estimates are shown in Figure 4.11. The estimates follow the trajectory of the censuses and the ASSA model since 1985; for both sexes the estimates are higher than the 1985, 1991 and 1996 censuses, which is not unexpected since the earlier censuses were under-counted compared to the later censuses (as shown in the previous chapter) and are relatively consistent with the estimated population in 2001, 2007 and 2011. The estimates are within 0.2 per cent and 2.5 per cent of the census 2011 estimates for males and females respectively.

For males, the NEG-GAM estimates appear to be more consistent than the AltMYE with the censuses of 2001 and 2011 and for females, the NEG-GAM track the AltMYE closely. Since the age heaping in the death data is modelled explicitly, discontinuities can be observed in the population estimates as the cohorts which suffer from heaping reach age 70.

[46] m_x from the raw HMD data (i.e. not the graduated values from the life tables) for single years of age was used in this calculation. The mortality rates in 2011 estimated using the NEG-GAM model were compared to data from the HMD covering the similar period 2009-2011. The period 2010-2012 could not be used since, at the time of writing, data for 2012 was not available for several countries in the HMD.

Figure 4.11: Population estimates and unconstrained NEG-GAM projections, 70+, National, Males and Females, 1985-2015



AltMYE – Alternative mid-year estimates, Dorrington (2013); ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015); UNPD – United Nations Population Division (2013); USCB – United States Census Bureau (2015)

Figure 4.12 compares the age structure of the population enumerated in the censuses of 1996, 2001 and 2011 with the NEG-GAM estimates. Age exaggeration begins in the census data from the ages of 80, 86 and 77 for males and 78, 84 and 75 for females, in the years 1996, 2001 and 2011 respectively. Since the death data are not immune to age exaggeration, it is likely that age is more exaggerated in the population data. The extent of the age exaggeration appears to have worsened over time.

Estimates of m_x are shown in Figure 4.13 for males and females, in the years 1985, 1996, 2001 and 2011. Mortality is estimated to have improved marginally over the period between the ages of 70 and 79, at the rate of 0.7 per cent and 0.1 per cent for males and females respectively. At the older ages mortality is shown to have worsened, but this is probably a result of more accurate death data in the recent years which suffer from less age exaggeration, producing higher mortality rates.

Figure 4.12: Ratio of NEG-GAM estimates to Census results, Males and Females, National

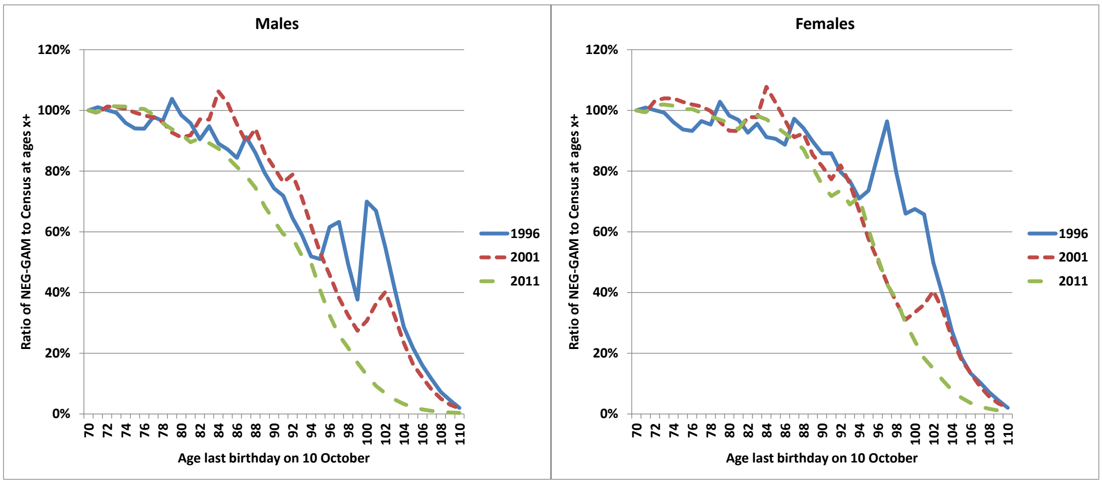
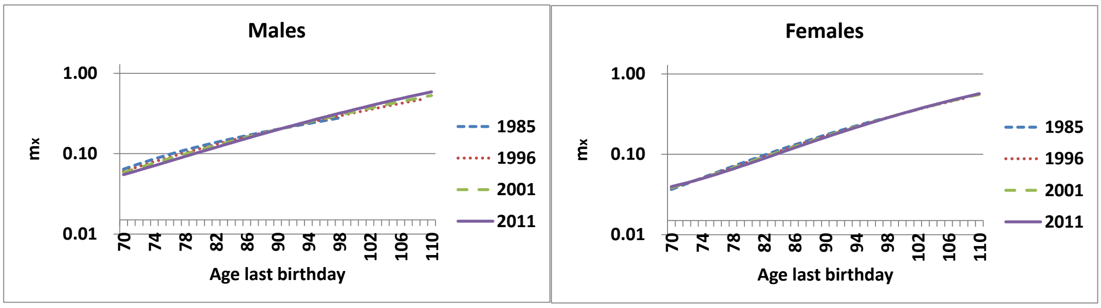
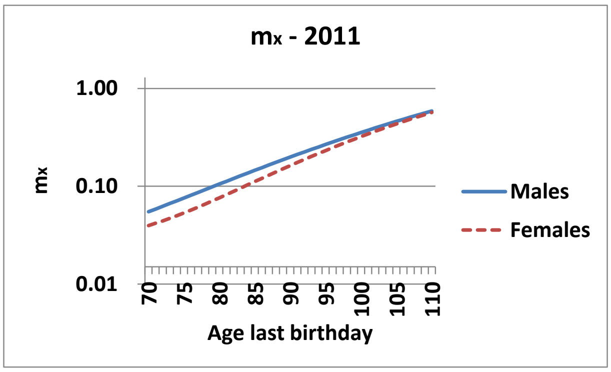


Figure 4.13: m_x, Males and Females, 1985-2011, National estimates, log scale



As shown in Figure 4.14, mortality rates for males remain higher than mortality rates for females at all ages but the female mortality advantage declines at the highest ages, which is not unreasonable (Thatcher (1987) makes a similar finding for the population of England and Wales).

Figure 4.14: m_x, Males and Females, 2011, National estimates, log scale



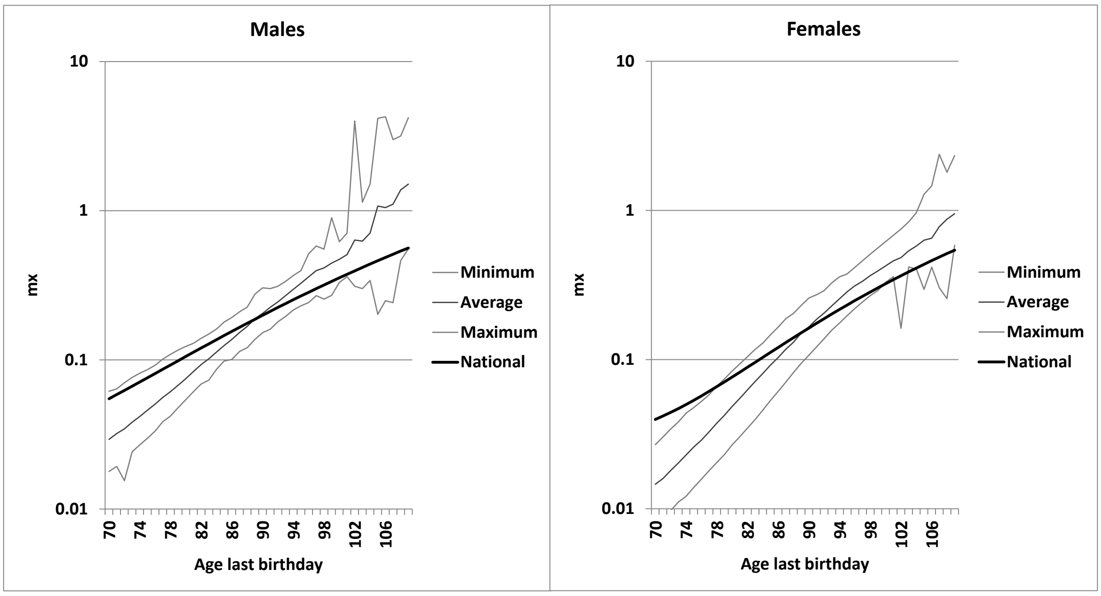
The NEG-GAM mortality rates in 2011 are compared to rates from the HMD in Figure 4.15. For males, the rates are bounded by the highest and lowest mortality rate at each age and cross over the average mortality at age 90. For females, the rates begin higher than any of the rates recorded in the HMD, but by age 78 are bounded by the rates from the HMD. At age 90, the rates cross over the average and remain between the average level and the minimum until around age 100, when the rates are in line with the lowest rates in the HMD. That the rates should start higher than the average and end lower is questionable and probably implies that the death data suffer from age exaggeration[47]. The fall in rates for females at the most advanced ages suggests that these rates probably suffer from more age exaggeration than the rates for males.

On the other hand, in an analysis of the old age mortality rates in China (which has been shown to have relatively accurate population and death data) in 1980, Yi and

[47] The NEG-GAM model is fit only to the death data (i.e. the model is not fit to the population data, which are shown in Figure 4.12 to suffer from age exaggeration relative to the death data), suggesting that the suspiciously low mortality rates may be caused by the exaggeration of age at death.

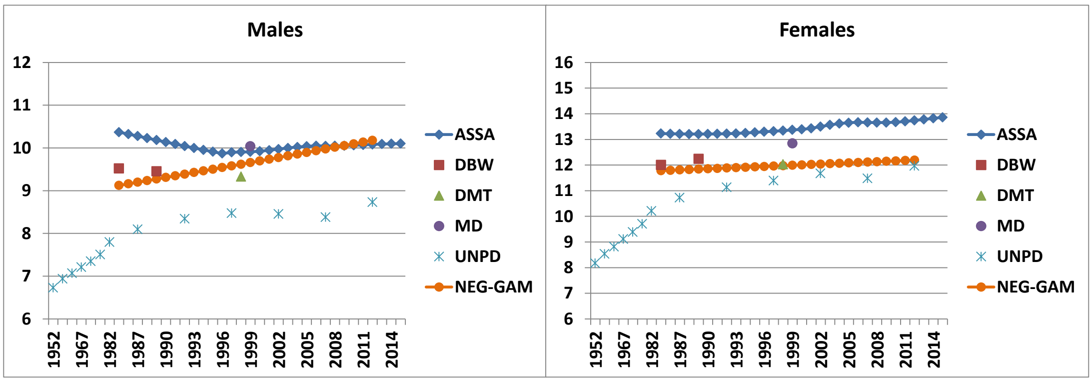
Vaupel (2006) found that the Chinese mortality rates converge to the same level as rates in Japan and Sweden at ages older than 97, which they attribute to the selective effect of poor living conditions, inadequate medical care and various wars on mortality. It is possible that a similar selective effect operates in South Africa but this is unlikely given the evidence of age misreporting in the death data and the cross-over of mortality rates at a relatively younger age (90 compared to 97 in Yi and Vaupel).

Figure 4.15: NEG-GAM mortality rates, National estimates in 2011, Males and Females, versus estimates of m_x from the Human Mortality Database, log scale



Life expectancy at age 70 estimated using the NEG-GAM estimates is compared to other sources in Figure 4.16. The estimates are roughly consistent with the estimates of Dorrington, Bradshaw and Wegner (1999) and Dorrington, Moultrie and Timæus (2004) and are lower than the estimates of Machemedze and Dorrington (2011), probably because the NEG-GAM model has dealt more effectively with the age heaping in the death data than in their study. Given the probable age exaggeration in the death data, the actual values of life expectancy are likely a bit lower than shown in Figure 4.16.

Figure 4.16: Estimates of e_{70}, National

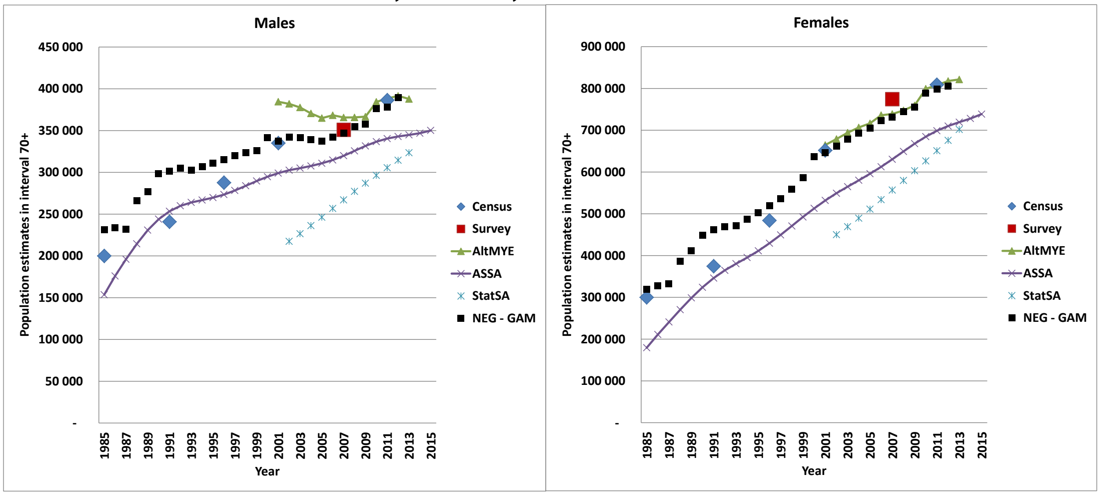


ASSA - Actuarial Society of South Africa, UNPD – United Nations Population Division, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), MD - Machemedze and Dorrington (2011)

4.5.2 Africans

The NEG-GAM population estimates are shown in Figure 4.17 for the African population group. The effect of age heaping is more pronounced in these estimates than at a national level. The estimates follow the trajectories of the censuses and the ASSA model, and, the NEG-GAM estimates are very consistent with the 2001 and 2011 censuses for both sexes and with the Alt-MYE for females.

Figure 4.17: Population estimates and unconstrained NEG-GAM projections, 70+, Males and Females, Africans, 1996-2015.



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Comparing the age distributions for 1996 in Figure 4.18, the NEG-GAM projected more deaths from the age of 100 for males and 95 for females, implying that too few were recorded in the census of 1996 at these ages. In 2001, age exaggeration began at age 74 for males and 77 for females and in 2011, from age 74 and 72.

Mortality rates, shown in Figure 4.19, show little change since 1985 at the younger ages (worsening marginally in the range 70-79 by 0.04 per cent for males and 0.2 per cent for females). At the older ages, mortality rates are shown to have deteriorated significantly since 1985, which is probably a consequence of the death data being less affected by age exaggeration at the highest ages since 1985. As support for this explanation, it can be seen that the mortality rates for females, and to a lesser extent for males, increase with age at an unrealistically slow rate in 1985 compared to the later years, when the rate of increase with age is faster.

Figure 4.18: Ratio of NEG-GAM estimates to Census results, Males and Females, Africans

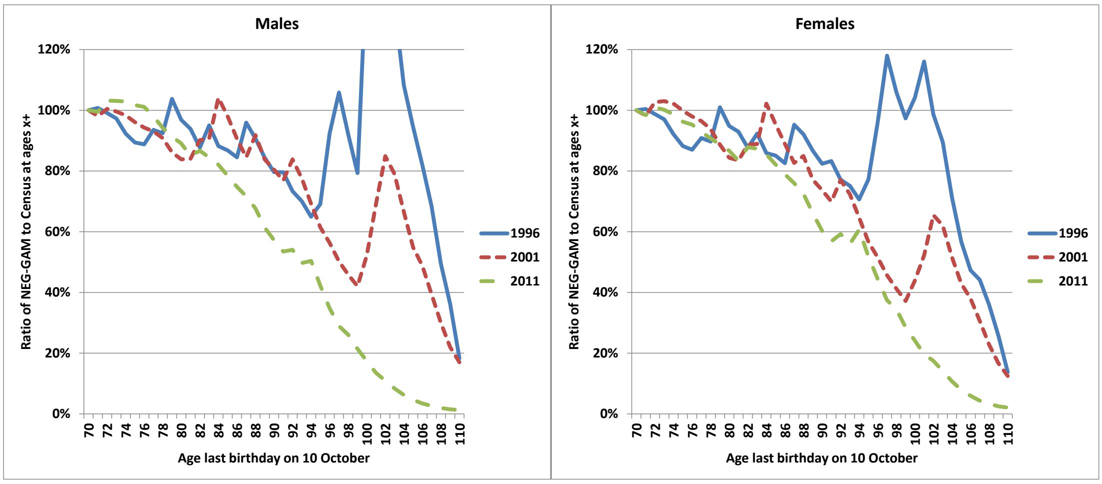
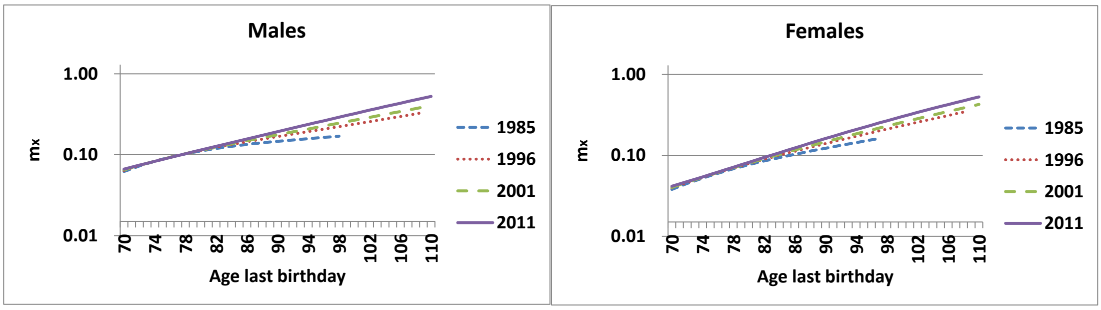


Figure 4.19: m_x, Males and Females, 1985-2011, Africans, 70-110, log scale



Male and female mortality rates in 2011 are shown in Figure 4.20. Rates for males exceed those for females at all ages, with the gap between the rates narrowing with age.

Estimates of life expectancy at age 70 are shown in Figure 4.21. The NEG-GAM estimates for males are consistent with those of Dorrington, Bradshaw and Wegner (1999) and Dorrington, Moultrie and Timæus (2004), as well as the ASSA model in the more recent years. The estimates are somewhat less consistent for females. The projected gains in life expectancy in the estimates of Sadie (1988) appear too optimistic. Given the evidence of age exaggeration in the death data, the true estimates of life expectancy are probably a bit lower than shown in the figure.

Figure 4.20: m_x, Males and Females, 2011, Africans, log scale

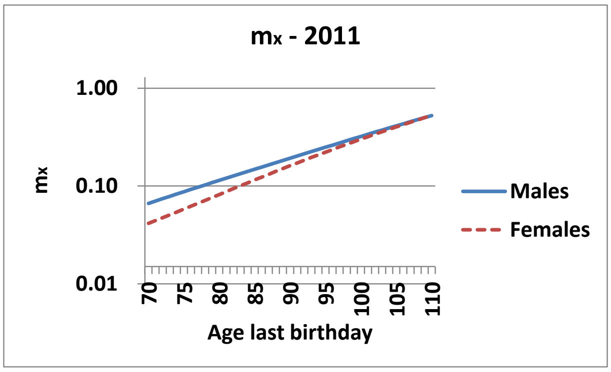
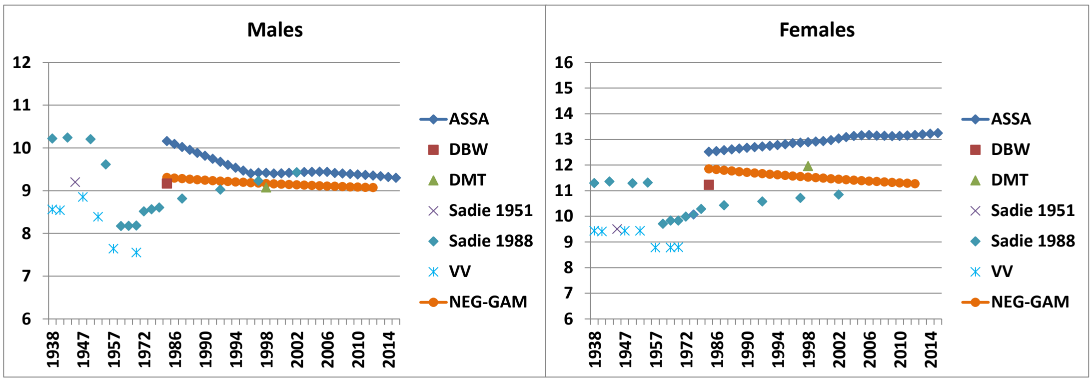


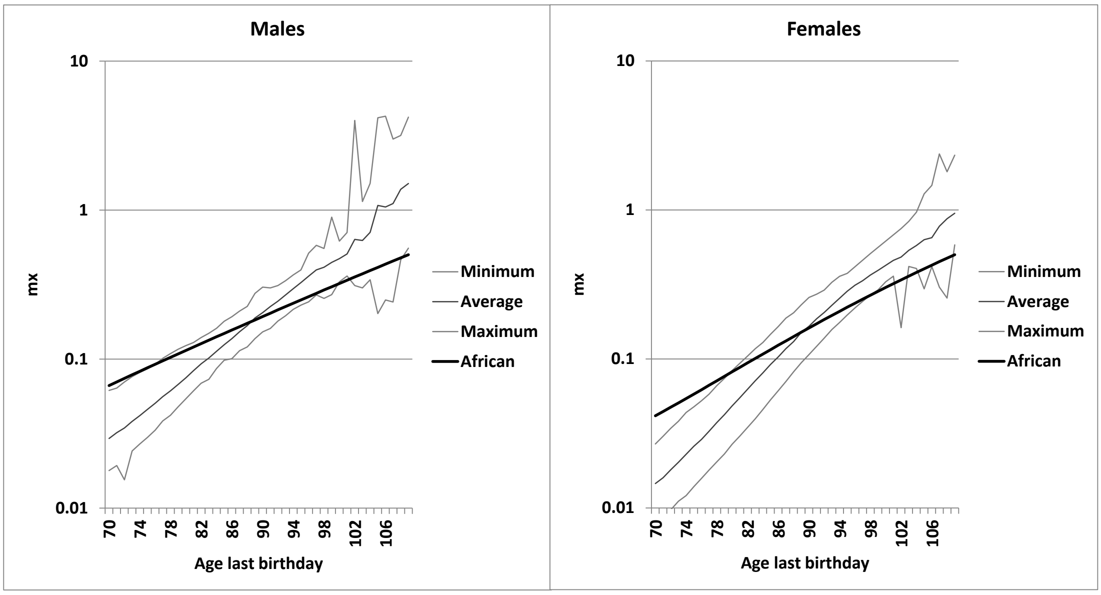
Figure 4.21: Estimates of e_{70}, Africans



ASSA - Actuarial Society of South Africa (2009), DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

As further support for this explanation, Figure 4.22 compares the NEG-GAM mortality rates to the rates from the HMD in 2011. In 2011, the mortality rates were higher than any of the rates in the HMD until ages 77 and 80 for males and females respectively, after which the rates cross over the average at around age 90 (as discussed above, these rates cannot be conclusively dismissed as unreasonable). The flatter mortality rates for 1996 projected by the NEG-GAM would cross over the average rates in the HMD at much earlier ages, pointing again to age exaggeration in the death data (and not the population data, since the NEG-GAM model was fit only to the death data).

Figure 4.22: NEG-GAM mortality rates, Africans in 2011, Males and Females, compared to estimates of m_x from the Human Mortality Database, log scale



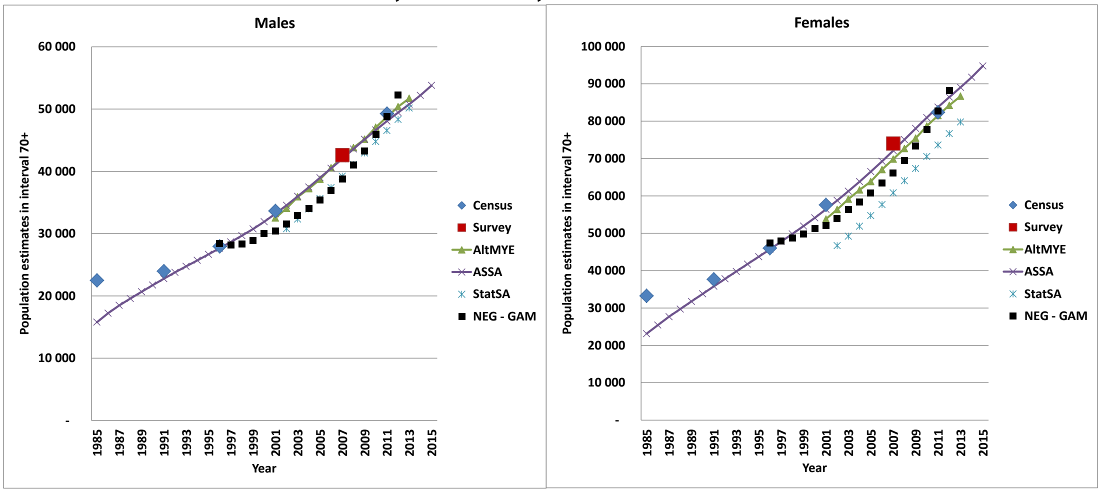
4.5.3 Coloureds

Some difficulties were encountered when fitting the NEG-GAM model to the death data for the Coloured population group. The NEG-GAM model produced estimates of the population aged 70+ in 2011 of 32 862 and 61 171 for males and females respectively, compared to the census counts of 49 269 and 82 348, indicating a problem with either the model fit or the death data. To investigate the possibility that the NEG-GAM model was causing the problem, the Survivor Ratio method (using five cohorts of data) was applied, producing estimates of 38 972 and 73 626, which are still problematically low despite being closer to the census 2011 numbers. The improvement of the Survivor Ratio method over the NEG-GAM model suggests that using only the most recent death data produces improved estimates and indeed, fitting the NEG-GAM model to the death data in the years 1999-2013 (in other words, excluding the data in the years 1984-1990) produced estimates in 2011 of 48 813 and 82 704, showing that including the death data for the earlier period 1984-1990 was distorting the model fit and suggesting that there is a problem with these data.

The assumption that the death data for the Coloured population group was completely reported in the period 1985-1991 was then tested by applying the SEG method to the death data period, which produced estimates of completeness of reporting of 120 per cent for males and 119 per cent for females. These results show that relative to the census estimates of the population in this period, too many deaths were recorded as belonging to the Coloured population group, which lends weight to the argument that these death data are problematic. Since there is no way of deriving suitable adjustments for completeness relative to the 2011 census (since the death data are missing population group in the period 1991-1998), it was decided that the NEG-GAM model should be fit to the data in the period 1999-2013 only. Mortality rates were projected backwards to 1996[48] using the NEG-GAM model. The NEG-GAM model is thus potentially useful as a diagnostic of data quality when some prior information about the population under consideration is available. In this case, census counts for the coloured population are available and the assumption that the death data corresponded to the census counts in the period 1985-1991 was effectively tested using the NEG-GAM model. The conclusion that the earlier data are problematic is supported by the application of the DDMs, which was applied independently of the NEG-GAM model and implies a similar conclusion.

Figure 4.23 shows that the NEG-GAM population estimates are consistent with the censuses of 1996 and 2011, but relatively inconsistent with the census of 2001 and the survey of 2007. The age distribution of the NEG-GAM estimates compared to the censuses is shown in Figure 4.24. The 2011 census showed the most age exaggeration, beginning at age 71 for both males and females. No definite conclusions about age exaggeration can be drawn in the other years.

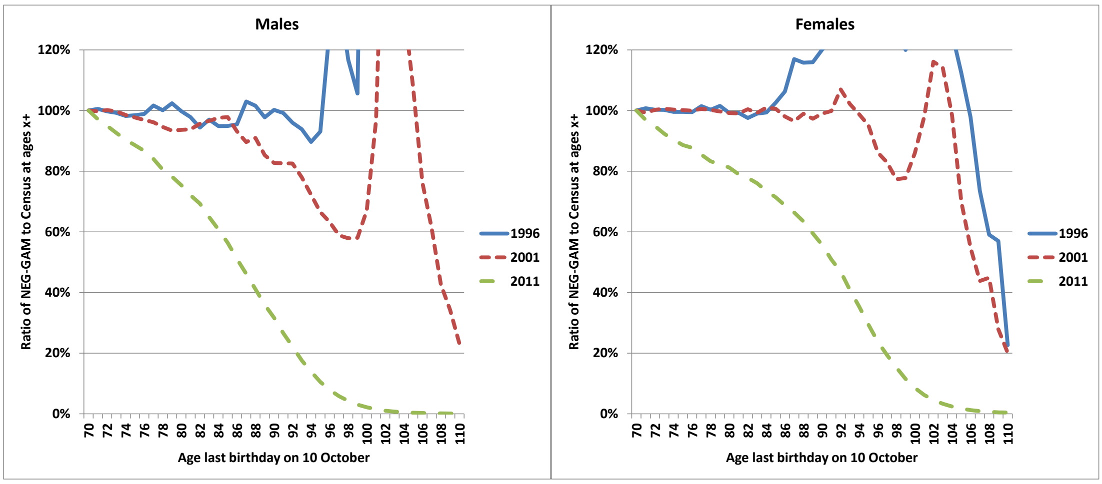
Figure 4.23: Population estimates and unconstrained NEG-GAM projections, 70+, Males and Females, Coloureds, 1996-2015



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

[48] Projecting back to 1985 produced mortality rates that appeared to be too high relative to those produced for Africans.

Figure 4.24: Ratio of NEG-GAM estimates to Census results, Males and Females, Coloureds



Estimates of mortality from the NEG-GAM are shown in Figure 4.25. Mortality rates have improved steadily since 1996 at the ages 70-79, at the rate of 2.7 per cent for males and 2.5 per cent females. Mortality has worsened slightly at the older ages. The rates for males remain higher than those for females at all ages as shown in Figure 4.26.

Figure 4.25: m_x, Males and Females, 1985-2011, Coloureds, log scale

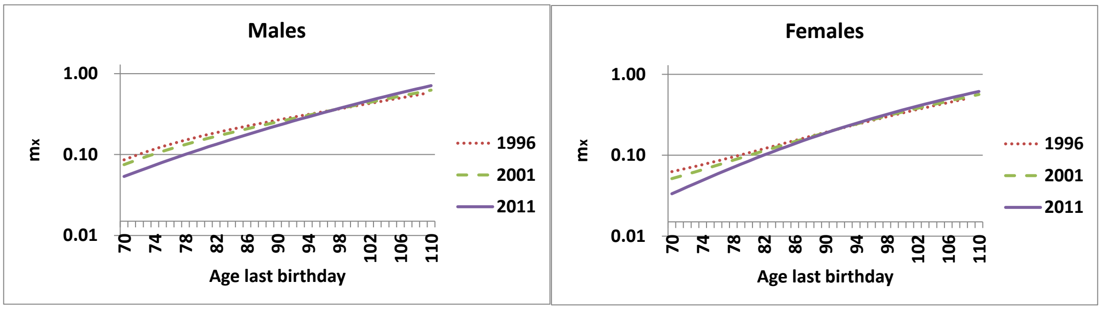
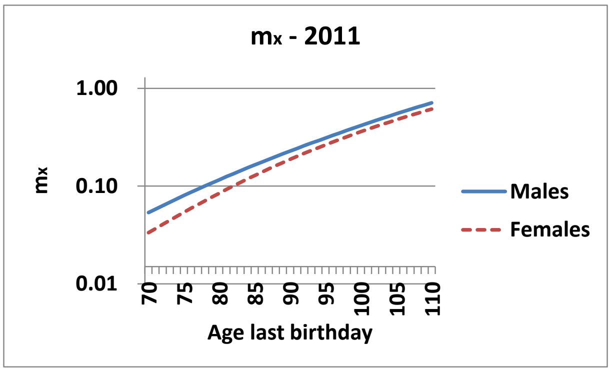


Figure 4.26: m_x, Males and Females, 2011, Coloureds, log scale



Compared to the rates in the HMD shown in Figure 4.27, the NEG-GAM rates cross-over the average mortality rate at ages 94 for both males and females. The rates for both sexes, therefore, appear possibly to suffer from age exaggeration in the death data, but less so than the rates nationally and for Africans.

The NEG-GAM estimates of life expectancy are shown in Figure 4.28. The estimates for males are lower than those of Dorrington, Moultrie and Timæus (2004) for males, but quite consistent with their estimate for females. The projected life expectancy in Sadie (1988) is in line with the NEG-GAM estimates only in 2002. The estimates from the SALT85 appear too high since the NEG-GAM estimates only reach the level of the SALT85 in the most recent years. The rate of mortality improvement is greater in the NEG-GAM estimates than in the ASSA model estimates or those produced by Sadie, which results in a somewhat implausible increase in life expectancy for males over the period 1996-2011. The implausible increase in life expectancy is probably a consequence of fitting the NEG-GAM model to too few years of data (i.e. for Coloureds, the model was fit to deaths occurring since 1999 whereas the other models were fit on deaths occurring since 1984) of relatively poor quality.

Figure 4.27: NEG-GAM mortality rates in 2011, Coloureds, Males and Females, compared to estimates of m_x from the Human Mortality Database, log scale

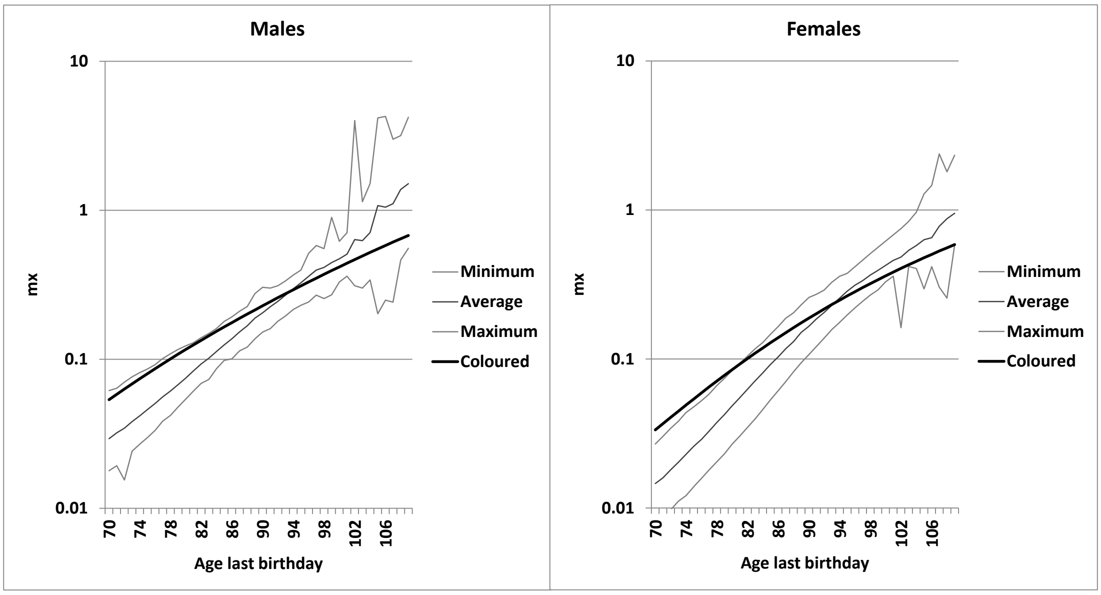
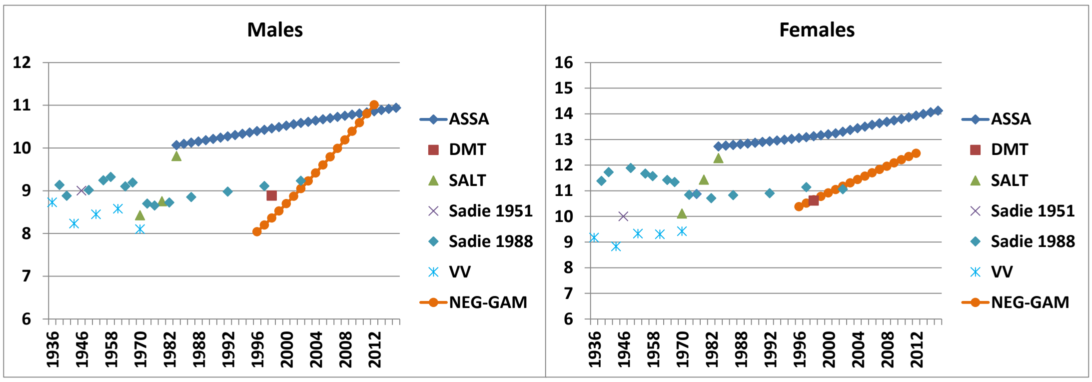


Figure 4.28: Estimates of e_{70}, Coloureds



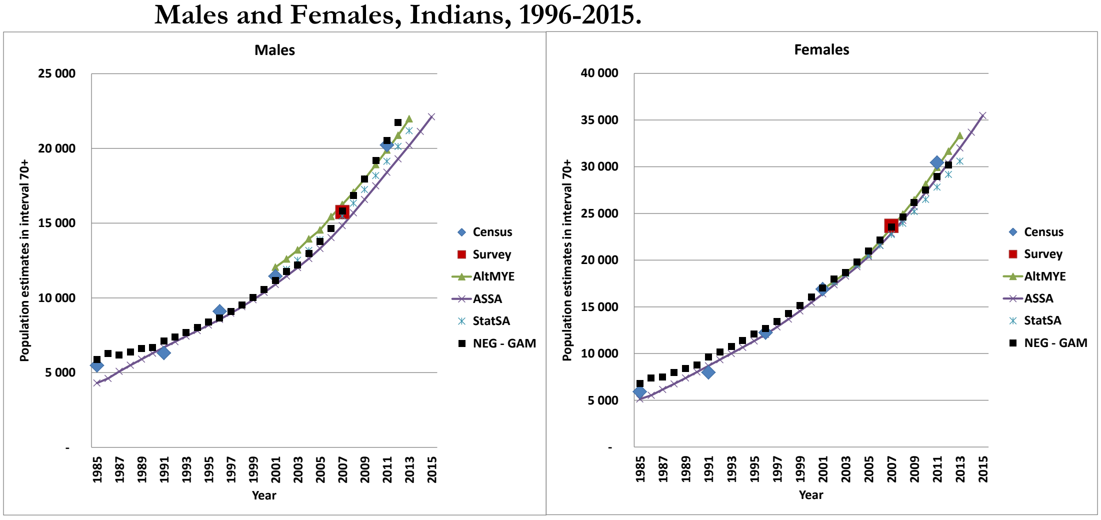
ASSA - Actuarial Society of South Africa (2009), DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

4.5.4 Indians

The NEG-GAM population estimates are shown in Figure 4.29. The estimates track the censuses since 1985 closely. For males, the estimates are in line with the ASSA model until around 2008, after which they are higher and more in line with the AltMYE and Stats SA. For females, the NEG-GAM estimates are in line with all of the other estimates.

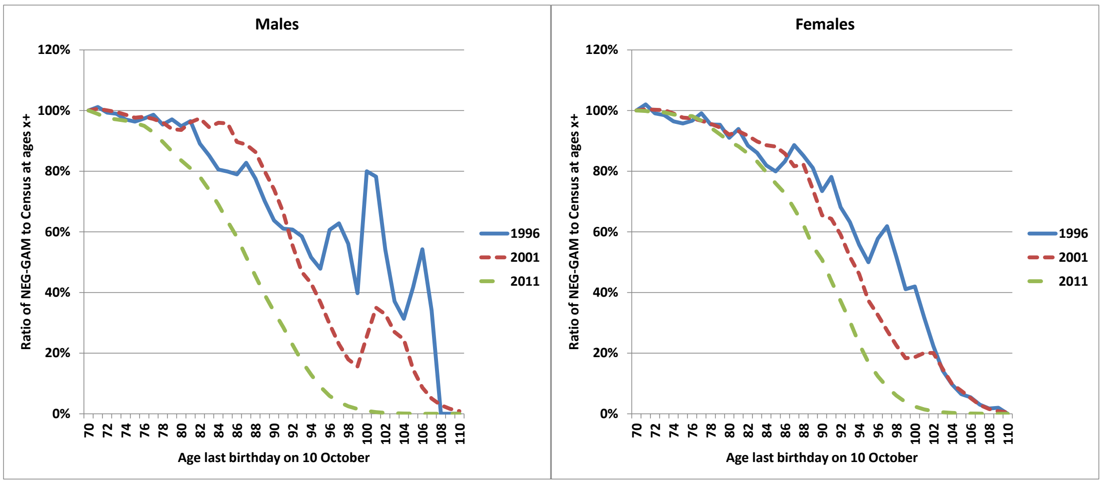
The age distribution of the NEG-GAM estimates compared to the censuses is shown in Figure 4.30. Age exaggeration begins for males at ages 72, 74 and 71 in the censuses of 1996, 2001 and 2011 and correspondingly for females at 72, 74 and 72.

Figure 4.29: Population estimates and unconstrained NEG-GAM projections, 70+, Males and Females, Indians, 1996-2015.



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Figure 4.30: Ratio of NEG-GAM estimates to Census results, Males and Females, Indians



Mortality rates for males and females, shown in Figure 4.31, have decreased at ages 70-79 since 1985, at the rate of 1.8 per cent for males and 1.2 per cent for females. At the older ages, rates appear to have worsened, but this is probably due to the greater volume of the death data at the older ages in the most recent years. It can be seen that mortality rates for males in 1985 increase with age unrealistically slowly, but in the other years appear more reasonable. Rates for males remain higher than females over the age range 70-110 as shown in Figure 4.32.

Figure 4.31: m_x**, Males and Females, 1985-2011, Indians, log scale**

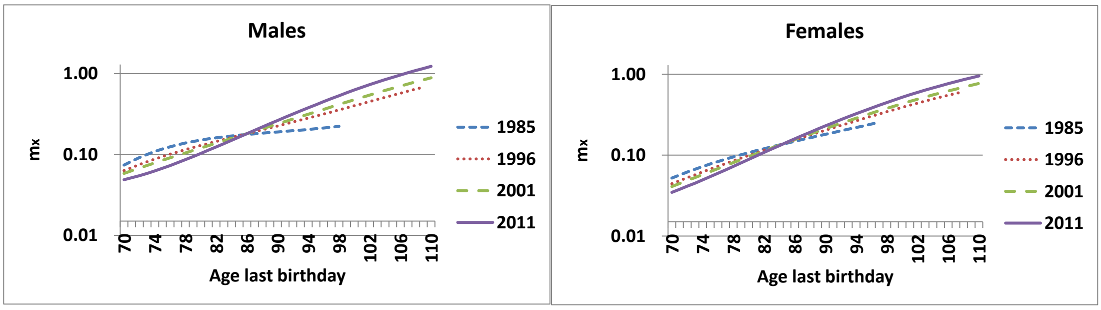
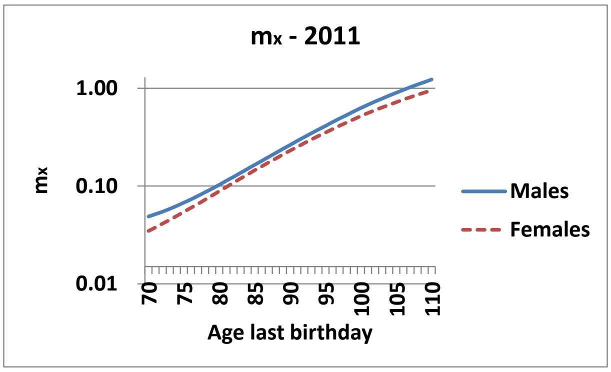


Figure 4.32: m_x**, Males and Females, 2011, Indians, log scale**



Compared to the mortality rates in the HMD shown in Figure 4.33, rates for the Indian population group in 2011 appear realistic, remaining at the high end of the range until the oldest ages when the rates tend towards the average.

A comparison of estimated life expectancy is shown in Figure 4.34. The NEG-GAM estimates begin at the same level as the estimates of Sadie (1988) in the 1980s, but increase faster than the rest of that series. The estimates in the SALT 80 and SALT85 appear too high.

Figure 4.33: NEG-GAM mortality rates, Indians in 2011, Males and Females, compared to estimates of m_x from the Human Mortality Database, log scale

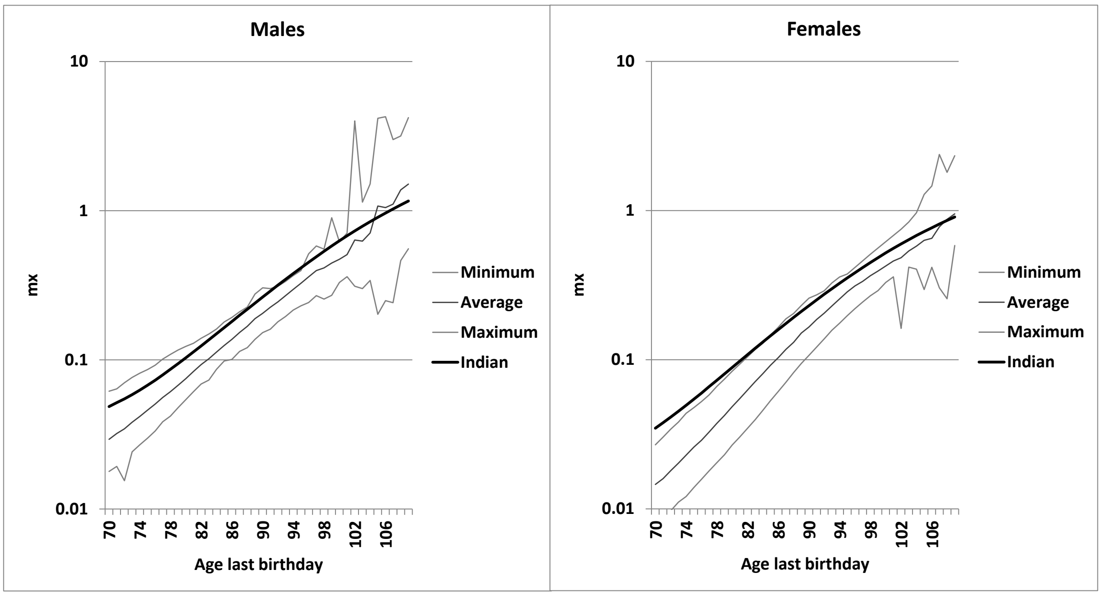
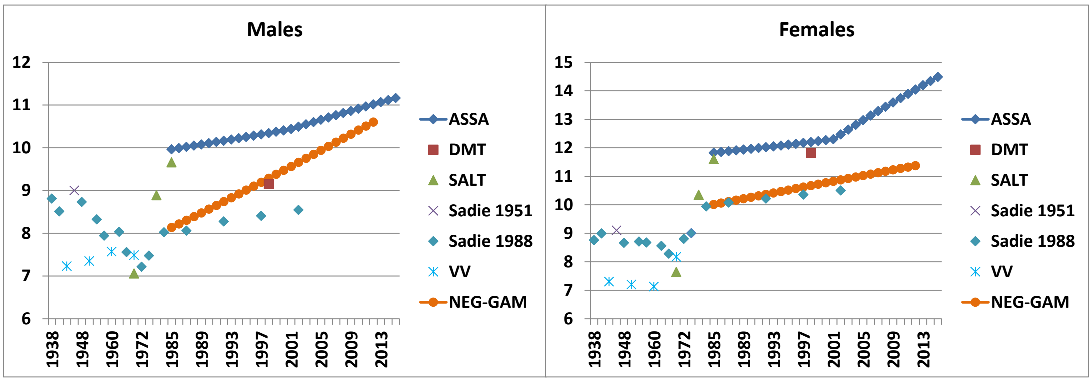


Figure 4.34: Estimates of e_{70}, Indians



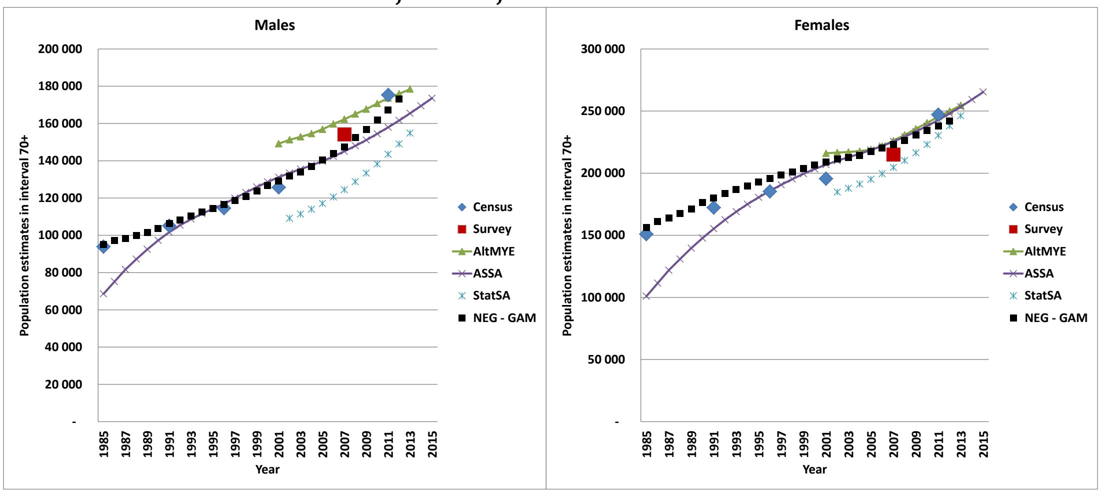
ASSA - Actuarial Society of South Africa, DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

4.5.5 Whites

The NEG-GAM population estimates are shown in Figure 4.35 to track all of the censuses since 1985 closely, except in 2011, when the NEG-GAM estimates are lower than the census estimates by 4.6 per cent for males and 3.7 per cent for females. The

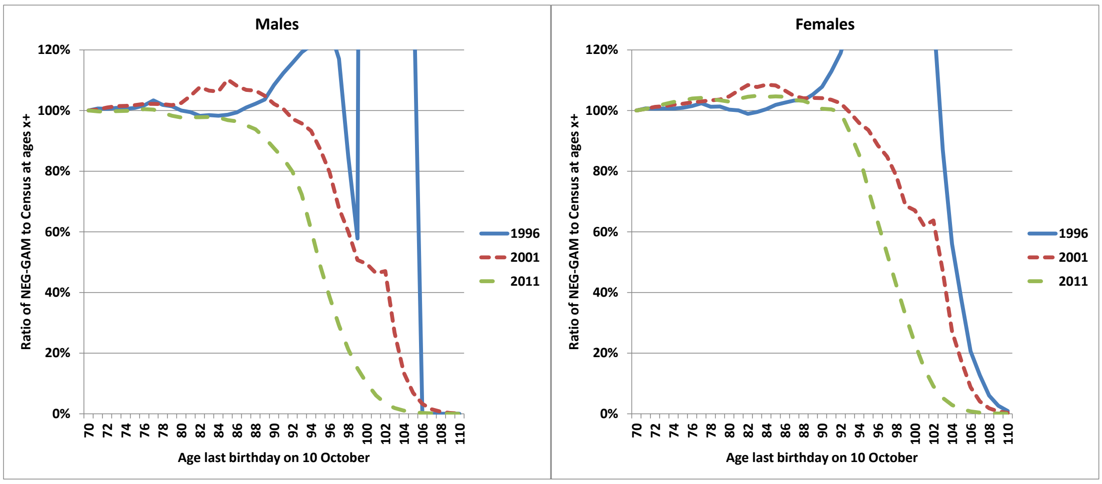
AltMYE estimates, which were back-projected from the 2011 census counts, are significantly higher than the census 2001 estimates for males, and somewhat less so for females, implying that the census 2011 count may be too high for the White population group (possibly due to problems with the Post Enumeration Survey adjustments to the 2011 census data), and that the NEG-GAM estimates are perhaps more reasonable. The comparison of the age structure of the NEG-GAM estimates to the census estimates in Figure 4.36 shows that the 1996 census appears to be free of age exaggeration and that age exaggeration appears in the census data in 2001 and 2011 from the ages of 92 and 78 for males and 93 and 92 for females.

Figure 4.35: Population estimates and unconstrained NEG-GAM projections, 70+, Males and Females, Whites, 1996-2015.



AltMYE – Alternative mid-year estimates, ASSA – Actuarial Society of South Africa (2009); Stats SA – Statistics South Africa (2015)

Figure 4.36: Ratio of NEG-GAM estimates to Census results, Males and Females, Whites



Mortality rates in the 70-79 age range have fallen since 1985 since at the rate of 1.5 per cent for males and 0.9 per cent for females, as shown in Figure 4.37. Rates appear to have risen at the older ages, with the effect larger for males than for females. Bearing in mind that the volume of deaths at the oldest ages (95+) has increased rapidly in recent years for both sexes as the White population ages, the rise in rates at the older ages is probably due to more statistically credible estimates of mortality rates being produced in the more recent years.

Rates for males are higher than for females over the entire age range with the female mortality advantage diminishing with age, as shown in Figure 4.38.

Figure 4.37: m_x, Males and Females, 1985-2011, Whites, log scale

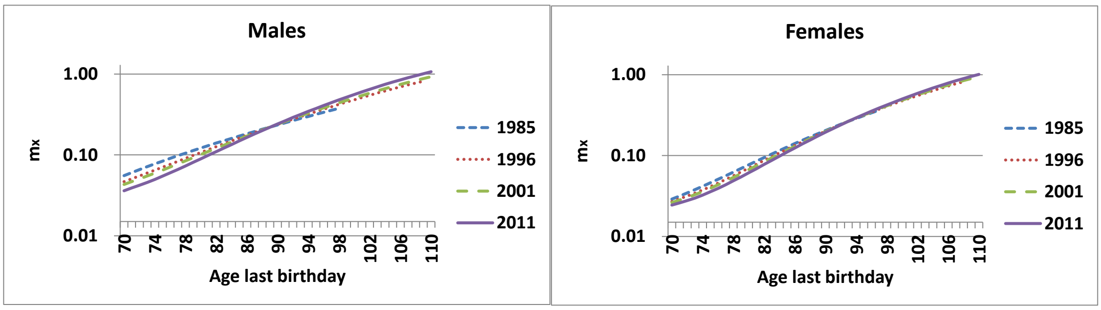


Figure 4.38: m_x, Males and Females, 2011, Whites, log scale

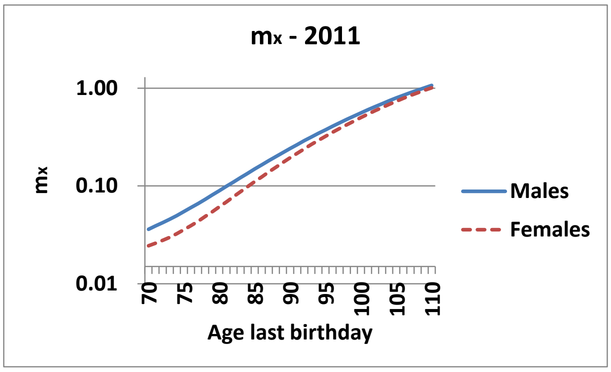
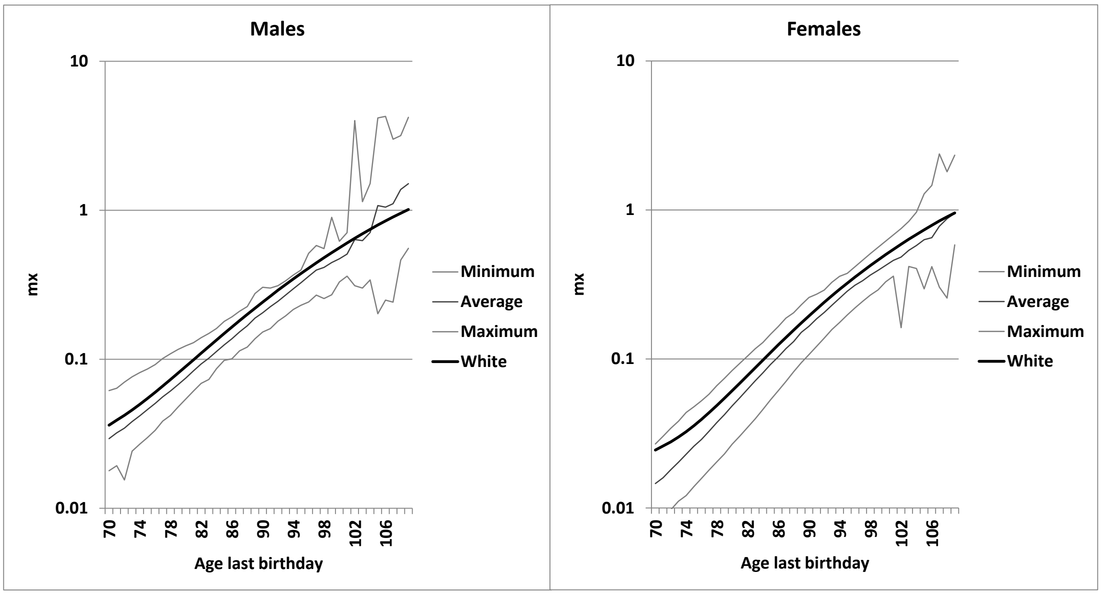


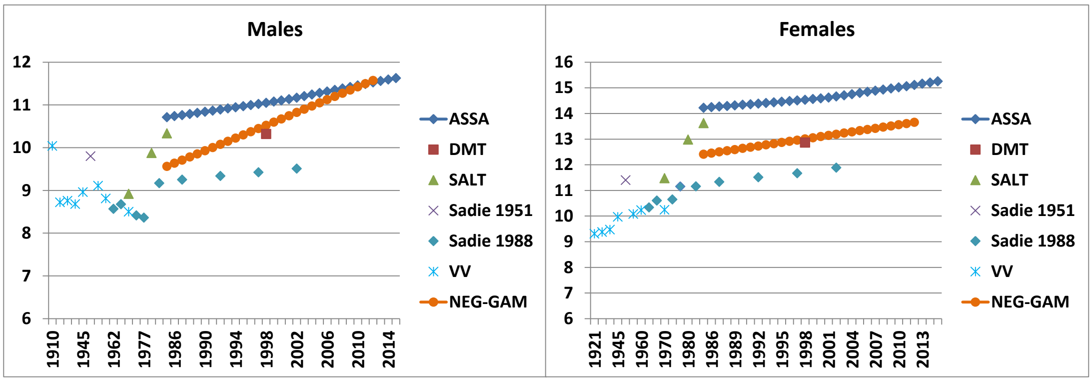
Figure 4.39 shows that mortality rates for the White population group are higher than the average in the HMD and remain at the upper end of the range before tending towards the average at the highest ages for females. The rates for males at the highest ages are somewhat lower than the average mortality rate in the HMD, but not implausibly so (m_{110} is estimated at 1.067, which is higher than the assumption of Coale and Kisker (1990) of 1).

Figure 4.39: NEG-GAM mortality rates, Whites in 2011, Males and Females, compared to estimates of m_x from the Human Mortality Database, log scale



Estimates of life expectancy are shown in Figure 4.40. The NEG-GAM estimates are consistent with Dorrington, Moultrie and Timæus (2004) for males and females and lie between the estimates of Sadie (1988) and the ASSA AIDS model. Compared to the NEG-GAM estimates, both the SALT80 and SALT85 overestimated life expectancy for whites.

Figure 4.40: Estimates of e_{70}, Whites



ASSA - Actuarial Society of South Africa (2009), DMT-Dorrington, Moultrie and Timæus (2004), DBW - Dorrington, Bradshaw and Wegner (1999), SALT – South African Life Tables, VV - Van Tonder and Van Eeden (1975)

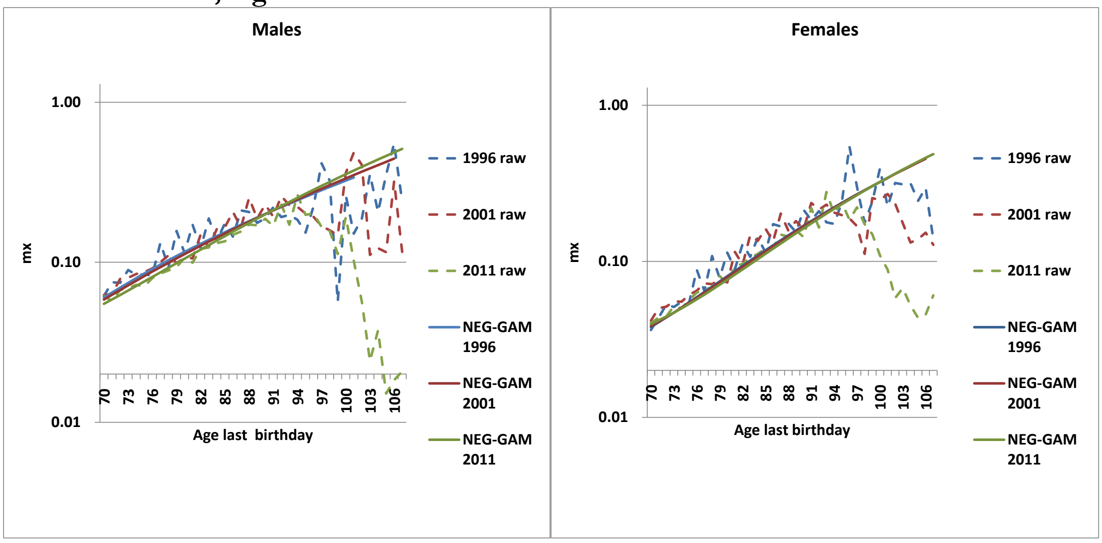
4.5.6 Summary

The mortality rates derived using the NEG-GAM at a national level are compared in Figure 4.41 to the rates implied by the census exposures. Mortality rates were calculated using the deaths[49] occurring 6 months preceding and following the census

[49] Rates were estimated after the death data were corrected for incomplete reporting of deaths.

(approximated assuming that deaths occur uniformly over the year) divided by the census population, and these rates are referred to as the "raw" rates in the comparison below. Similar to the findings of Das Gupta (1991) on the 1985 US census, the raw rates flatten from around age 90 for both males and females. The raw rates in 2011 appear to be worst affected by age exaggeration (consistent with the comparison of the NEG-GAM and census age distributions), forming a "bell shape" as the rates drop from around age 95 for both males and females. The NEG-GAM rates roughly follow the census estimates until the age of 90, after which the NEG-GAM estimates keep rising. The age heaping which visibly affects the raw rates has been removed from the NEG-GAM rates, showing that the NEG-GAM model has produced a more realistic set of rates than those produced using the census data (and also than the original NEG methods, which also flatten, as shown above).

Figure 4.41: National mortality rates, Males and Females, 1996, 2001 and 2011, raw rates derived from deaths and census data compared to NEG-GAM mortality rates, log scale



The mortality rates in 2011 derived for each population group are shown in Figure 4.42 and are presented in the tables in Appendix 0. At the younger ages, mortality rates for males are highest for Africans, followed by Coloureds, Indians and Whites, which is in line with Dorrington, Moultrie and Timæus (2004). For females, mortality rates are highest for Africans, followed by Indians, Coloureds and Whites. The rates for Indian females are slightly higher (on average, 2.1 per cent at ages 70-79) than those for Coloured females, whereas the rates for females in Dorrington, Moultrie and Timæus (2004) were higher for Coloureds than for Indians. The difference is explained by the rapid mortality improvement modelled in the NEG-GAM rates for Coloureds, of 2.5 per cent per annum at ages between 70 and 79; in 1999, the NEG-GAM rates for

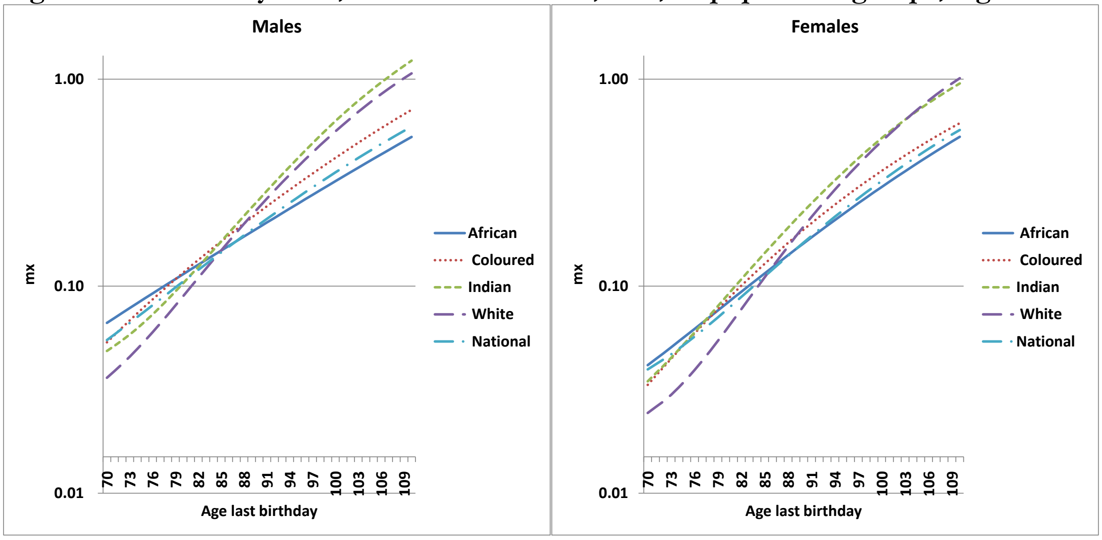
Coloured females were higher than those for Indian females by an average of 6.4 per cent. National mortality rates closely follow the rates for the African population group.

The rates for Africans cross over the rates for Whites at age 85 for males and 87 for females and for Coloureds at age 89 for males and 90 for females, which is suspicious since a population group that exhibits high mortality at the younger ages would be expected to suffer from a high level of mortality at the older ages as well, as discussed in Preston and Elo (2006), Coale and Kisker (1986, 1990) and Preston, Elo, Rosenwaike *et al.* (1996). These crossovers provide further evidence that the death data[50] for the African and Coloured population groups suffer from age exaggeration and therefore these mortality rates at the advanced ages should be treated sceptically; since the national rates are weighted towards the rates for Africans, they should similarly be treated with suspicion.

The mortality rates for Indian males remain above those for White males over the entire age range, which is reasonable considering that mortality is higher for Indians than Whites at the younger ages, but cross over the rates for White females at the oldest ages. Given the small amount of death data at the very oldest ages in both population groups, this crossover does not appear to be statistically credible. From the comparison with the mortality rates in the HMD, the rates for Indians and Whites of both sexes seem to be reasonable over the entire age range. The level of mortality at the highest ages modelled by the NEG-GAM model also seems demographically reasonable. For Whites, m_{110} is 1.07 for males and 1.01 for females and for Indians, m_{110} is 1.07 for males and 0.95 for females. These values are all higher than the corresponding assumptions for m_{110} in Coale and Kisker (1990) of 1 for males and 0.8 for females, indicating that the death data for these population groups do not seem to suffer from obvious age exaggeration.

[50] As noted in Footnote 47, the NEG-GAM model is fit only to the death data, and is therefore not biased by age exaggeration in the population data.

Figure 4.42: Mortality rates, Males and Females, 2011, all population groups, log scale



4.6 The question of exponentially increasing mortality

The question of whether mortality rates in general increase at an exponential rate was discussed in detail in Section 2.6.1. The spline function that was used in the NEG-GAM model is flexible enough to capture various shapes of the mortality curve implied by the death data. If a Gompertz curve approximates the NEG-GAM mortality rates well, then it can be inferred that mortality does indeed increase exponentially with age. To test this, Gompertz curves were fit to the mortality rates derived in this chapter, over the age range 70-110. The ratios of the fitted Gompertz mortality rates to the NEG-GAM rates are shown in Figure 4.43 and Figure 4.44.

Figure 4.43: Ratio of fitted Gompertz mortality rates to NEG-GAM mortality rates, 2011, National and African, Males and Females

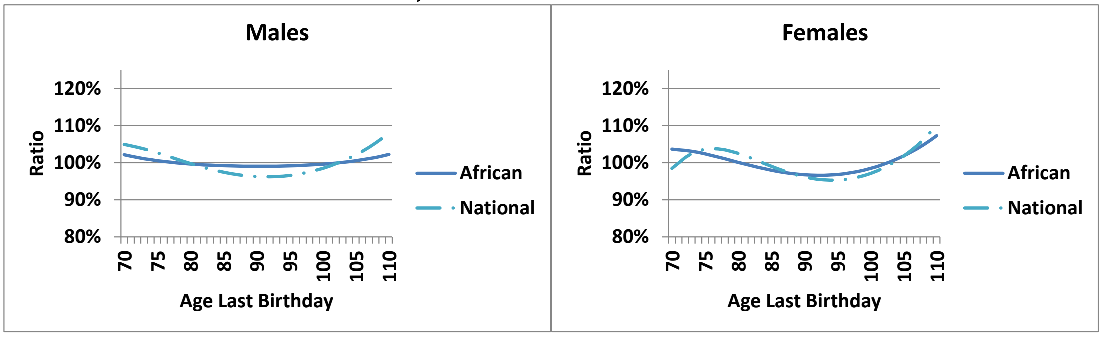
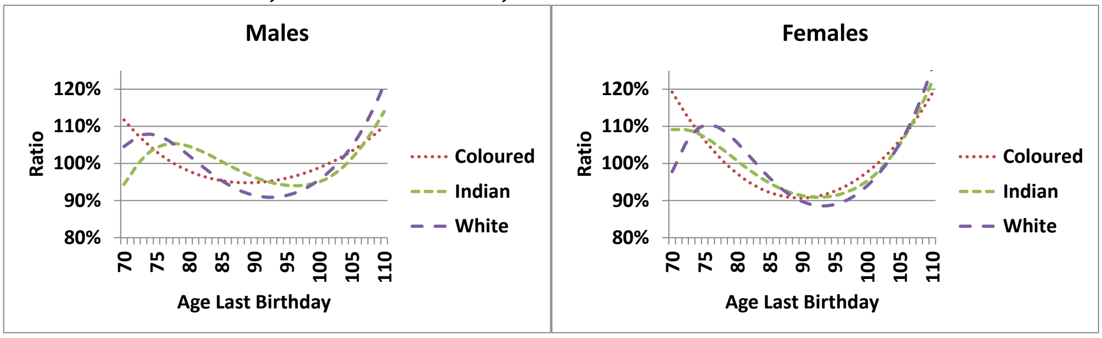


Figure 4.44: Ratio of fitted Gompertz mortality rates to NEG-GAM mortality rates, 2011, Coloured, Indian and White, Males and Females



The fit of the Gompertz curve appears to be best for those groups with the highest mortality at the younger ages, fitting well for African males, reasonably well for African females and nationally for both sexes, and poorly for Coloureds, Indians and Whites of both sexes. The Gompertz curves generally predict mortality rates that are too high for the younger and older ages and too low for the middle ages. Since no defects in the rates for Indians and Whites have been identified above, the poor fit of the Gompertz curve to these rates provides evidence to argue that the mortality rates of the oldest old do not increase exponentially.

Since it is suspected that the rates for the African and Coloured population groups are biased due to age exaggeration in the death data (i.e. the shape of these mortality curves are suspected to be inaccurate), no general inference can be drawn from the NEG-GAM rates for these population groups about whether mortality rates in developing countries increase exponentially. However, since the Gompertz curve fits the rates for Africans well, what can be inferred is that the problem with the NEG-GAM rates for Africans is not that mortality decelerates at the older ages, but rather that the rate of increase of mortality with age is too low over the entire age range. That this is the problem with the estimates is confirmed by examining the first difference of the log mortality rates (the rate of increase in mortality or k_x) which is shown in Figure 4.45 where it can be seen that k_x is flat for African males and is slightly curved and decreases slowly over the age range for African females.

In contrast, k_x is bell-shaped for the Indian and White population groups as shown in Figure 4.46. k_x has been shown by, amongst others, Horiuchi and Wilmoth (1998) and Wilmoth (1995), to be distinctively bell-shaped for females and somewhat less so for males, and Coale and Kisker (1990) assume that k_x is bell-shaped for both sexes. Figure 4.46 confirms that the NEG-GAM mortality rates for both the Indian and

White population groups exhibit this feature, lending credence to the argument that these rates are reasonable.

The values of k_x for the Coloured population group are higher than those for Indians and Whites at the younger ages but decline almost linearly over the age range, without displaying the characteristic bell-shape, implying that the problem with the NEG-GAM rates for Coloureds is that the rate of increase in mortality decelerates over the age range, without first undergoing the acceleration shown by the rates for Indians and Whites.

Figure 4.45: k_x, National and Africa, Males and Females, 2011

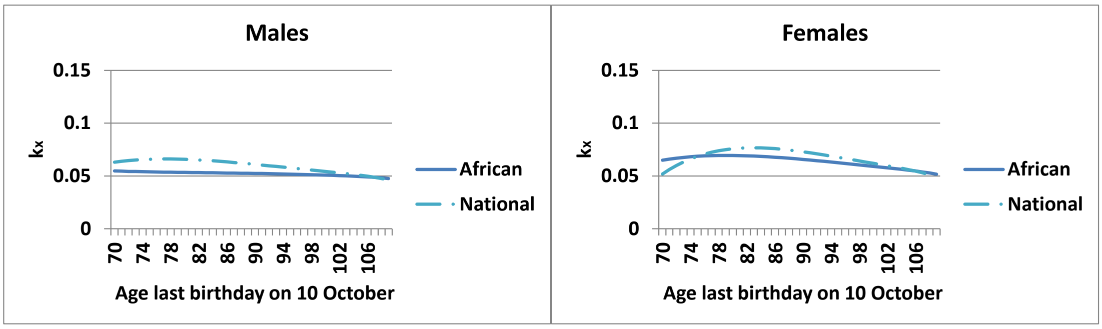
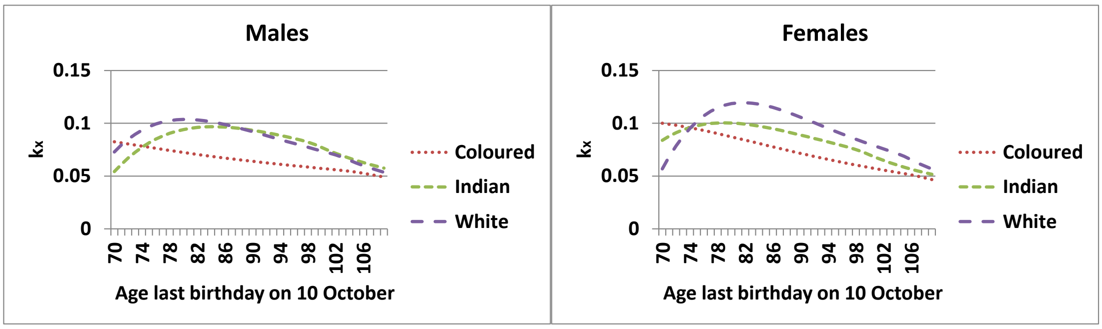


Figure 4.46: k_x, Coloured, Indian and White,, Males and Females, 2011



4.7 Conclusion

This chapter has discussed the reasons that the original NEG methods cannot be applied to death data that suffer from age and year of birth heaping and has suggested a new NEG model to solve these problems by modelling heaping explicitly. Mortality rates calculated using the exposures produced with this new model were presented and checked for internal consistency as well as consistency with other sources.

The rates presented in this chapter for the Indian and White population groups appear to be reasonable over the entire age range, based on comparison with the Human Mortality Database and the assumptions in Coale and Kisker (1990), and by inspection of the rate of increase in mortality with age.

The cross-over of mortality rates for Africans and Coloureds with the rates for Whites indicates that the death data for Africans and Coloureds probably suffer from age exaggeration at the oldest ages. However, the fact that these crossovers occur at ages 85+ and the relative consistency of the estimated life expectancy at age 70 with other sources indicate that the rates at the younger ages are probably robust. Accordingly, the estimates of annual mortality improvement at the ages 70-79 in the period 1985-2011 (and in the period 1996-2011 for Coloureds) shown in Table 4.1 are likewise probably robust and represent the best estimate of the trend in mortality over the period of this study.

Table 4.1: Annual mortality improvement rates, ages 70-79, 1985-2011 (Coloureds 1996-2011)

Group	Sex	Rate of improvement per annum
African	Male	-0.04%
African	Female	-0.20%
Coloured	Male	2.70%
Coloured	Female	2.50%
Indian	Male	1.80%
Indian	Female	1.20%
White	Male	1.50%
White	Female	0.90%
National	Male	0.70%
National	Female	0.10%

5 DISCUSSION

5.1 Introduction

This chapter presents conclusions drawn from the research results. The chapter also discusses the insights gained into the methods applied in this study, and highlights possible areas for future research before ending with high-level conclusions.

5.2 Results

This research set out to investigate the level and trend of old age mortality rates in South Africa for the period 1985-2011. The approach taken to accomplish this goal was to use the death data to re-estimate the population. These population estimates served as exposures in the calculation of mortality and thus mortality rates that were less biased due to misreporting of age in the population data were produced. Two main classes of techniques were applied in this study to accomplish this objective. The Death Distribution Methods (DDMs) were used to correct the data for incomplete reporting of deaths and of population group and the Near Extinct Generations (NEG) methods were used to project the future deaths for each cohort to derive estimates of the population.

The DDMs were applied first to correct the death data, and the NEG methods were then used to reconstruct the population. One could perhaps argue that there is no need to apply the DDMs first since estimates of the completeness of reporting of deaths could be derived by comparing the NEG population estimates based on the ratio of the reported deaths to the census population. However, this approach is problematic for several reasons. The unconstrained NEG methods have been shown to underestimate the population (Andreev 2004; Thatcher, Kannisto and Andreev 2002) and, therefore, estimates of completeness produced in this way may have been too low. Furthermore, since completeness of reporting has changed over time (as shown in section 3.3.8), the estimates of the population produced using the NEG methods applied to the reported death data would have been biased, perhaps severely, as discussed in section 2.4.2.2. Lastly, it would have been difficult to assess the level of bias produced by applying the NEG methods to the uncorrected death data.

Correcting the death data before applying the NEG methods allowed the hypothesis that the corrected death data was at the same (relative) level of completeness to be tested when fitting the NEG-GAM model and corrections to be made when this was shown not to be the case. Since the death data were corrected to be 100 per cent complete compared to the census 2011 counts, it was furthermore possible to check whether the population estimates produced using the NEG-GAM model were close to the census 2011 counts. Therefore, by correcting the death data first, a benchmark was provided to test the workings of the NEG-GAM model and ensure that the results appeared to be relatively free from bias. A further reason to apply the DDMs first is that the estimates of the census under- or over-count produced by the DDMs (shown in Table 3.10) allow the deaths to be corrected relative to the most recent census, whereas estimates of completeness produced using the NEG methods alone would have been relative to the coverage of the earlier censuses which was less than 100 per cent relative to the census in 2011.

Taking these arguments into account, the approach taken of correcting the death data followed by applying the NEG methods to the corrected data appears to be sound.

The results of the DDMs were consistent with previous studies, as shown in Table 3.16. It was argued in section 3.3.5 that it is important to apply the DDMs for each population group separately and then to estimate the national completeness of reporting as a weighted average of the population group estimates (as opposed to applying the DDMs to data for the population as a whole). Because of this, the final estimates of completeness used in this research were different from the consensus in the literature.

Table 3.7 shows that national completeness of reporting for males improved by 13 per cent from the period 1996-2001 to 2001-2011. Completeness of reporting for females also improved over this period, but by only 5 per cent. Also, as noted in Section 3.3.5, the difference between the estimated completeness for males and females is approximately 9 per cent, which is in line with the difference between the estimated completeness for African males and females in the period 1996-2001. The rapid increase in completeness of reporting for males (relative to the increase for females) and the difference in estimated completeness between males and females should be investigated further in future research to validate the estimates produced in this research using the DDMs. In particular, the appropriateness of the backward extrapolation of the deaths reported in 1999-2001 that underlies these estimates of completeness could be investigated further.

The DDMs were applied to population and death data that were shown by the NEG-GAM to suffer from age and year of birth heaping and age exaggeration, which might have led to inaccuracies in the estimated completeness of reporting. However, as the data were aggregated into five year age groups with a maximum open interval of 85+, below which most of these errors were found, and the estimates of completeness were, in most cases, calculated in the age range 25-64 and the estimates of life expectancy used in the SEG method were based on mortality curves graduated in the age range 45-80, the worst effects of the misreporting on the DDMs were probably avoided.

This study identified that age and year of birth heaping are part of the reason why the original NEG methods produced flattened mortality rates when applied to the South African death data. By recognising that a regression model can be fit to produce results that are almost the same as the Das Gupta method (and also that the Survivor Ratio method is equivalent to the Das Gupta method applied to smoothed data, as shown in Appendix A.4), a new NEG technique was proposed to solve these two particular problems. Thus, more reasonable estimates of mortality rates were made in this research than would have been the case if the original NEG methods were applied. A critical demonstration that the new technique is valid is that the census populations aged 70+ were reproduced accurately in 2011 and the NEG-GAM population estimates since 1985 generally followed a reasonable trajectory over time. Therefore, it is fair to conclude that the NEG-GAM technique is less biased when applied to the South African death data than the original NEG techniques.

However, the new technique cannot deal with the problem of age exaggeration in the death data and there is evidence of age exaggeration in the mortality rates that were produced. The mortality rates are therefore most probably biased downward at the older ages and, in this respect, certain objectives of the study could not be entirely achieved since the actual level of mortality rates in this period cannot be estimated from the death data for the African and Coloured population groups, and hence nationally. Also, while the research set out to estimate mortality rates for the years 1985-2011, it was found that including the death data for the Coloured population group in the years 1985-1991 within the NEG-GAM model produced estimates of the population in 2011 that were too low. The NEG-GAM model for Coloureds was, therefore, fit using the data for the years 1999-2011, and mortality rates were only produced for the years 1996-2011, as discussed in section 4.5.3. Thus, another objective of the study could not be entirely accomplished.

Nonetheless, the mortality rates in the age range 70-79 appear to be robust for all population groups, the estimates of life expectancy produced at age 70 are consistent with the estimates in the literature and mortality improvement rates calculated at these ages appear to be sensible. Furthermore, the mortality rates (and the rate of increase of mortality with age) for the Indian and White population groups appear reasonable over the entire age range. The research has shown that mortality appears to be improving at the ages 70-79 for the Coloured, Indian and White population groups and deteriorating slowly for the African population group. Therefore, the objectives of the research can be said to have been partly achieved.

A significant result of the study is that the level of the SALT80 and SALT85 mortality rates appears to be incorrect at the older ages, producing estimates of life expectancy that are too high for the years 1980 and 1985 relative to most of the other sources examined.

5.3 Techniques

5.3.1 Death Distribution Methods

In line with Dorrington, Moultrie and Timæus (2004), this research found that estimating the completeness of reporting using the DDMs for the population as a whole may produce results that are biased if the groups comprising the population have different demographic profiles, as discussed in section 3.3.5. It is therefore recommended that, whenever possible, the DDMs be applied for different population groups separately and national completeness then derived from these estimates.

The estimates of completeness in this study were based on the adult ages. There was little evidence of falling completeness of reporting at the older ages, which had been hypothesised to occur in earlier studies (Machemedze 2009; Machemedze and Dorrington 2011). However, a more accurate approach may have been to fit curves to the age-specific estimates of completeness derived using the SEG method, instead of assuming that completeness estimated for the adult ages (25-64) applies equally to the older ages. Also, it was assumed that the estimates of completeness of reporting are the same for males and females. Although comparing the NEG-GAM estimates to the census population in 2011 shows that there does not seem to be a systematic bias introduced into the estimates by these simplifications, this could be investigated further.

Several assumptions were needed to produce annual estimates of completeness. These assumptions included assuming that the estimates follow a smooth logistic curve and that the level of completeness estimated using the DDMs occurs at the midpoint of the intercensal interval. Other adjustments for late reported deaths, falling completeness of reporting and increased reporting of population group were also made. The NEG-GAM technique allowed some of these assumptions to be tested and modified. Based on the coefficients of NEG-GAM model shown in Appendix C.1.2, it seems as if completeness of reporting for those aged 70+ was lower than initially assumed in 1988-1994 and in the years since 2009, as discussed in section 4.4.5.

This study applied the General Growth Balance method of Hill (1987) and the Synthetic Extinct Generations method of Bennett and Horiuchi (1981) to derive estimates of completeness. The estimates were checked against estimates produced using the shortcut method of Preston and Lahiri (1991). The results were similar, but there were indications of a possible upward bias in the SEG estimates for the African, Indian and Coloured population groups compared to the shortcut estimates. The estimates of completeness for these population groups later proved to require adjustments to be made via the NEG-GAM model (for those aged 70+). The shortcut method, therefore, is a potentially useful diagnostic for researchers who are applying the DDMs and more research into the method seems to be warranted.

5.3.2 Near Extinct Generations Methods

This study provided some insights into the pitfalls that may be encountered when applying the NEG methods to developing country data. If the death data are not free of age and year of birth heaping before using the NEG methods, then population and mortality estimates produced by the NEG methods will likely be biased. The NEG-GAM technique offers a new way to deal with the problem of heaping in death data.

Much of the literature discussing the NEG methods has focussed on how best to allow for possible trends in mortality rates in the NEG methods (Andreev 1999, 2004; Terblanche and Wilson 2015; Thatcher, Kannisto and Andreev 2002). The NEG-GAM model offers a new option for the study of old age mortality in developed countries by fitting the trend in the death data in a statistically sound manner.

5.4 Areas of future research

A critical area for future research is how death data may be corrected for age exaggeration. The current literature, discussed in section 2.5.2, provides a method based on the general population theory (i.e. not restricted to stable populations) of Preston and Coale (1982) but this technique, unfortunately, requires that one knows the mortality rates, which is not the case at the older ages in developing countries.

Future research should be undertaken to understand the reasons for age exaggeration in the South African population and death data. Matching studies represent one form that this research could take. In particular, the reasons why the 2011 census data appear to suffer from the worse age exaggeration than the 1996 and 2001 censuses should be considered. Similarly, the reasons for incomplete reporting of deaths and of population group, as well as the reason for the improvement in completeness of reporting in recent years, should be investigated, since the process of reporting of deaths could perhaps be changed to facilitate complete death registration.

It can be seen in Table 4.1 that mortality improvement rates for females are lower than for males in all of the population groups. It is important to investigate why there appears to be a gender differential in mortality improvement rates in South Africa. Similarly, understanding why mortality rates for the African population group seem to be deteriorating at the older ages is an important area of research, which could be performed by studying the cause of death information.

In this research, curves were fit through the estimates of completeness derived using the DDMs to produce annual estimates of completeness.

For the years 1985-1987, completeness of reporting was assumed to be constant at 54 per cent, and the reasonability of this assumption was examined in Footnote 35 by comparison with the estimates in Dorrington (1998). An alternative to assuming constant completeness in these years is to include the estimates of completeness in Dorrington (1998) in the fit of the curve of annual completeness, and this could be tested in future research.

A useful area of future research would be to investigate if it is possible find a method to calculate annual estimates of completeness of reporting directly from the population and death data, which would obviate the need to assume a particular type of curve.

In order to complete the near extinct cohorts, an allowance for late reported deaths was made by assuming that the ratio of late-reported deaths to deaths that have already been reported remains constant in the future at the same level as was observed in 2013. Future research could test the sensitivity of the NEG-GAM results to this assumption.

As was acknowledged in Section 4.4.3 there is a risk that the NEG-GAM model may not model the death data adequately, for example, if not enough control variables are added to capture the effect of errors in the data. Future research should thoroughly investigate the NEG-GAM model and control variable approach to confirm that they produce reasonable results. Specifically, the model could be validated using the “fake-data” procedure described in Gelman and Hill (2006), by simulating death data from a known mortality table and checking if the NEG-GAM model recovers the mortality rates (both when the death data are free of age heaping and when contaminated with age heaping), as well as by applying the model to countries with high quality demographic data to ensure that the results are reasonable and in line with other sources. If the validity of the approach is confirmed, then the counterintuitive results of Section 4, that African and Coloured mortality was lighter than mortality in the other population groups at advanced ages, can be confidently be attributed to age exaggeration in the death data.

Embedding the NEG methods within a regression framework opens up several new possibilities for the study of old-age mortality that were not possible using the original NEG methods. The first is the opportunity to apply more advanced regression techniques to the death data. For example, instead of fitting separate models for each of the population groups and each sex, as was done in this research, a single hierarchical model could be fit to the death data for all of the population groups and both sexes at once, with the resulting estimates of mortality being made more robust via pooling of information (Gelman and Hill 2006). In a similar manner, a hierarchical model could be fit to death data from several developed countries at once to provide robust estimates of mortality and mortality improvement. Another possibility provided by the regression framework is the application of bootstrap techniques to assess the uncertainty of the estimated population, mortality and mortality improvement rates. In particular, studying the uncertainty of old-age mortality improvement rates could shed light on the appropriate calibration of the capital requirements for longevity risk for life insurers in South Africa.

This research applied the NEG methods from age 70 since population projections performed by Machemedze (2009)showed that age exaggeration began after age 75. Since evidence of age exaggeration was found for Coloureds and Indians at ages younger than 75, future research could attempt to apply the NEG methods at ages younger than 70 to determine whether age exaggeration may in fact begin at ages younger than 70 (Kannisto (1990) provides a precedent for implementing the method of extinct generations at ages 55+).

The research compared the mortality rates estimated using the NEG-GAM model to mortality rates from the HMD covering the period 2009-2011. A limitation of the research is that a more comprehensive comparison of the NEG-GAM results to mortality rates in countries at times when mortality was similar to that of the population groups in South Africa was not performed and this issue could be resolved in future research.

The Survivor Ratio method was shown to be equivalent to the Das Gupta method applied to smoothed death data in Appendix A.4. Future research could attempt to incorporate a similar smoothing step into the Das Gupta Advanced technique of Andreev (2004), which may enhance its accuracy to produce optimal results even in small populations.

In this research, control variables were added to the NEG-GAM model based on visual inspection of plots of the residuals from the model. Future research could address a statistical procedure to determine when control variables should be applied and thus reduce the extent of the somewhat subjective judgement needed to apply the NEG-GAM model to death data distorted by heaping or incompleteness of reporting.

REFERENCES

Actuarial Society of South Africa. 2009. *Aids and Demographic Model 2008*. Cape Town: ASSA. http://www.actuarialsociety.org.za/Societyactivities/CommitteeActivities/DemographyEpidemiologyCommittee/Models.aspx.

Andreev, K.F. 1999. "Demographic surfaces: Estimation, Assessment and Presentation, with Application to Danish Mortality." Unpublished thesis, Odense: University of Southern Denmark.

Andreev, K.F. 2004. "A method for estimating size of population aged 90 and over with application to the 2000 US census data", *Demographic Research* **11**:235-262.

Bah, S. 2005. "Technical Appraisal of Official South African Life Tables," in Zuberi, Tukufu, Amson Sibanda and Eric Udjo (eds). *The Demography of South Africa*. pp. 310.

Beard, R.E. 1961. "A theory of mortality based on actuarial, biological, and medical considerations," Paper presented at Proceedings of International Population Conference, New York. Vol. 1:611-625.

Beard, R.E. 1971. "Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes," in Brass, William (ed). *Biological Aspects of. Demography*. London: Taylor and Francis, pp. 57-68.

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

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

Bhat, P. 1990. "Estimating transition probabilities of age misstatement", *Demography* **27**(1):149-163.

Bhat, P. 1992. "Changing demography of elderly in India", *Current Science* **63**(8):440-448.

Bhat, P. 1995. "Age misreporting and its impact on adult mortality estimates in South Asia", *Demography India* **24**(1):59-80.

Bhat, P. 2002. "General growth balance method: A reformulation for populations open to migration", *Population Studies* **56**(1):23-34.

Bourbeau, R. and B. Desjardins. 2002. "Dealing with problems in data quality for the measurement of mortality at advanced ages in Canada", *North American Actuarial Journal* **6**(3):1-13.

Bourbeau, R. and A. Lebel. 2000. "Mortality statistics for the oldest-old: An evaluation of Canadian data", *Demographic Research* **2**(2):36.

Bradshaw, D., M. Schneider, R. Laubscher and B. Nojilana. 2002. "Cause of death profile, South Africa 1996", *Cape Town, South Africa: South African Medical Research Council*

Brass, W. 1975. *Methods for estimating fertility and mortality from limited and defective data*. Laboratories for Population Studies: Chapel Hill.

Caldwell, J. and A. Igun. 1971. "An experiment with census-type age enumeration in Nigeria", *Population Studies* **25**(2):287-302.

Camarda, C.G., P.H. Eilers and J. Gampe. 2008. "Modelling general patterns of digit preference", *Statistical Modelling* **8**(4):385-401.

Central Statistical Service. 1992a. *Population census 1991. Summarised results before adjustment for undercount*. Pretoria: Sentrale Statistiekdiens.

Central Statistical Service. 1992b. *Population Census 1991: Adjustment for the Undercount*. Pretoria: Sentrale Statistiekdiens.

Chambers, J.M. and T.J. Hastie. 1991. *Statistical models in S*. Chapman and Hall/CRC

Cleveland, W., E. Grosse and W. Shyu. 1992. "Local Regression Models," in Chambers, John M and Trevor J Hastie (eds). *Statistical Models in S*. Chapman and Hall/CRC

Coale, A.J. and G. Caselli. 1990. "Estimation of the number of persons at advanced ages from the number of deaths at each age in the given year and adjacent years", *Genus*:1-23.

Coale, A.J. and E.E. Kisker. 1986. "Mortality crossovers: Reality or bad data?", *Population Studies* **40**(3):389-401.

Coale, A.J. and E.E. Kisker. 1990. "Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages", *Asian and Pacific Population Forum* **4**(1):1-31.

Coale, A.J. and S. Li. 1991. "The effect of age misreporting in China on the calculation of mortality rates at very high ages", *Demography* **28**(2):293-301.

Cohen, B., J. Menken, V. Hosegood and I.M. Timaeus. 2006. "HIV/AIDS and Older People in South Africa," in Cohen, Barney and Jane Menken (eds). *Aging in Sub-Saharan Africa: Recommendations for Furthering Research*. Washington (DC): National Academies Press, pp. 250-275.

Condran, G.A., C. Himes and S.H. Preston. 1991. "Old age mortality patterns in low-mortality countries: an evaluation of population and death data at advanced ages 1950 to the present", *Population Bulletin of the United Nations* **30**:23-60.

Continuous Mortality Investigation. 2016. *CMI Mortality Projections Model consultation*. Working Paper 90. UK: The Institute and Faculty of Actuaries.

Cooper-Williams, J., L.-M. Albertyn and P. Lewis. 2012. "Mortality improvements in South Africa," Paper presented at 2012 Actuarial Convention, Actuarial Society of South Africa Cape Town.

Courbage, Y. and P. Fargues. 1979. "A method for deriving mortality estimates from incomplete vital statistics", *Population Studies* **33**(1):165-180.

CSI. 2012. *Annuitant Mortality 2001-2004*. Continuous Statistical Investigation Committee, Actuarial Society of South Africa.

Das Gupta, P. 1975. "A general method of correction for age misreporting in census populations", *Demography* **12**(2):303-312.

Das Gupta, P. 1991. "Reconstruction of the Age Distribution of the Extreme Aged in the 1980 Census by the Method of Extinct Generations," Paper presented at American Statistical Association Proceedings of the Social Statistics Section. 154-159.

Dechter, A.R. and S.H. Preston. 1990. "Age misreporting and its effects on adult mortality estimates in Latin America", *Population Bulletin of the United Nations* (31-32):1-16.

Demeny, P.G. and F.C. Shorter. 1968. *Estimating Turkish Mortality, Fertility, and Age Structure: Application of Some New Techniques*. publisher not identified.

Dépoid, F. 1973. "La mortalité des grands vieillards", *Population (french edition)*:755-792.

Dorrington, R.E. 1998. "Estimates of the level and shape of mortality rates in South Africa around 1985 and 1990 derived by applying indirect demographic techniques to reported deaths." Unpublished thesis, University of Cape Town.

Dorrington, R.E. 2013a. *Alternative South African mid-year estimates, 2013*. CARe Monograph No. 13. Cape Town: Centre for Actuarial Research, University of Cape Town. http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono13.pdf.

Dorrington, R.E. 2013b. *The Generalized Growth Balance Method*. Paris: International Union for the Scientific Study of Population.

http://demographicestimation.iussp.org/content/generalized-growth-balance-method. Accessed: 2016/01/31 2016.
Dorrington, R.E. 2013c. *Synthetic extinct generations methods*. Paris: International Union for the Scientific Study of Population. http://demographicestimation.iussp.org/content/synthetic-extinct-generations-methods. Accessed: 2016/01/31 2016.
Dorrington, R.E., D. Bourne, D. Bradshaw, R. Laubscher *et al.* 2001. "The impact of HIV/AIDS on adult mortality in South Africa",
Dorrington, R.E., D. Bradshaw, R. Laubscher and N. Nannan. 2015. *Rapid mortality surveillance report 2014*. Cape Town, South Africa: South African Medical Research Council.
Dorrington, R.E., 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", *Cape Town: South African Medical Research Council*
Dorrington, R.E., T.A. Moultrie and I. Timæus. 2004. *Estimation of mortality using the South African Census 2001 data*. Centre for Actuarial Research, University of Cape Town.
Dorrington, R.E., I.M. Timæus and T.A. Moultrie. 2008. "Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors, revisited", *Working Paper*
Dorrington, R.E. and S. Tootla. 2007. "South African annuitant standard mortality tables 1996–2000 (SAIML98 and SAIFL98)", *South African Actuarial Journal* **7**:161-184.
Elo, I.T. and S.H. Preston. 1994. "Estimating African-American mortality from inaccurate data", *Demography* **31**(3):427-458.
Feeney, G. 1979. "A technique for correcting age distributions for heaping on multiples of five," Paper presented at Asian and Pacific Census Forum. Vol. 5:12-14.
Feeney, G. 1990. "Untilting age distributions: A transformation for graphical analysis", *Asia and Pacific Population Forum* **4**(3):13-20.
Feeney, G. 1997. *Census Survival Ratio Consistency Check*. Unpublished Working Paper. http://www.gfeeney.com/research-notes/1997.census.survival.ratio.check/1997-census-survival-ratio-check.pdf
FSB. 2014. *Life SCR Longevity Risk*. Position Paper 64. Pretoria: Financial Services Board.
FSB. 2016. "Comprehensive Parallel Run Feedback ", Paper presented at 2016 SAM Workshop. Pretoria. Financial Services Board.
Gavrilov, L.A. and N.S. Gavrilova. 2011. "Mortality measurement at advanced ages: a study of the Social Security Administration Death Master File", *North American Actuarial Journal* **15**(3):432-447.
Gelman, A. and J. Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
Gomes, M.M.F. and C.M. Turra. 2009. "The number of centenarians in Brazil: indirect estimates based on death certificates", *Demographic Research* **20**(20):495-502.
Gompertz, B. 1825. "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies", *Philosophical transactions of the Royal Society of London*:513-583.
Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized Additive Models*. Chapman and Hall/CRC.

Heitjan, D.F. and D.B. Rubin. 1990. "Inference from Coarse Data Via Multiple Imputation with Application to Age Heaping", *Journal of the American Statistical Association,* **85**(410):304-314.
Hill, K. 1981. "An evaluation of indirect methods for estimating mortality," Paper presented at Seminar on Methodology and Data Collection in Mortality Studies Dakar, Senegal. International Union for the Scientific Study of Population.
Hill, K. 1987. "Estimating census and death registration completeness", *Asian and Pacific Population Forum* **1**(3):8-13.
Hill, K. 2001. *Methods for measuring adult mortality in developing countries: a comparative review*. Hopkins Population Center.
Hill, K. 2003. *Adult mortality in the developing world: what we know and how we know it.* New York: Population Division, Department of Economic and Social Affairs, United Nations.
Hill, K. and Y. Choi. 2004. "Performance of GGB and SEG given various simulated data errors", Paper presented at Workshop on Adult Mortality in the Developing World: Methods and Measures. Marconi Conference Center, Marin County, California, July 8-11, 2004.
Hill, K., Y. Choi and I. Timæus. 2005. "Unconventional approaches to mortality estimation", *Demographic Research* **13**(12):281-300.
Hill, K. and B. Queiroz. 2010. "Adjusting the general growth balance method for migration", *Revista Brasileira de Estudos de População* **27**(1):7-20.
Hill, K., D. You and Y. Choi. 2009. "Death distribution methods for estimating adult mortality", *Demographic Research* **21**:235-254.
Hill, M.E., S.H. Preston, I.T. Elo and I. Rosenwaike. 1997. "Age-linked institutions and age reporting among older African Americans", *Social Forces* **75**(3):1007-1030.
Hill, M.E., S.H. Preston and I. Rosenwaike. 2000. "Age reporting among white Americans aged 85+: Results of a record linkage study", *Demography* **37**(2):175-186.
Horiuchi, S. and A.J. Coale. 1982. "A simple equation for estimating the expectation of life at old ages", *Population Studies* **36**(2):317-326.
Horiuchi, S. and S.H. Preston. 1988. "Age-specific growth rates: The legacy of past population dynamics", *Demography* **25**(3):429-441.
Horiuchi, S. and J.R. Wilmoth. 1998. "Deceleration in the age pattern of mortality at older ages", *Demography* **35**(4):391-412.
Human Mortality Database (2008) University of California and Max Planck Institute for Demographic Research, Berkeley (USA) and Rostock (Germany).
Humphrey, G. 1970. "Mortality at the oldest ages", *Journal of the Institute of Actuaries* **96**(1):105-119.
Jdanov, D.A., D. Jasilionis, E. Soroko, R. Rau *et al.* 2008. *Beyond the Kannisto-Thatcher database on old age mortality: An assessment of data quality at advanced ages*. Rostock, Germany: Max Planck Institute for Demographic Research.
Jowett, A.J. and Y.-Q. Li. 1992. "Age-heaping: contrasting patterns from China", *GeoJournal* **28**(4):427-442.
Kannisto, V. 1988. "On the survival of centenarians and the span of life", *Population Studies* **42**(3):389-406.
Kannisto, V. 1990. *Mortality of the elderly in late 19th and early 20th century Finland*. Helsinki: Central Statsitical Office of Finalnd.
Kannisto, V., J. Lauritsen, A.R. Thatcher and J.W. Vaupel. 1994. "Reductions in mortality at advanced ages: several decades of evidence from 27 countries", *Population and Development Review* **20**(4):793-810.

Kirkwood, T.B. 2015. "Deciphering death: a commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'", *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**(1666):20140379.

Koch, R. 2016. *The Quantum Yearbook*. Port Elizabeth: Van Zyl Rudd.

Lee, R.D. 1982. *Correcting census age distributions: extensions and applications of the Demeny-Shorter technique*. Berkeley: University of California.

Lee, R.D. and D. Lam. 1983. "Age distribution adjustments for English censuses, 1821 to 1931", *Population Studies* **37**(3):445-464.

Machemedze, T. 2009. "Old age mortality in South Africa." Unpublished thesis, Cape Town: University of Cape Town.

Machemedze, T. and R. Dorrington. 2011. "Levels of mortality of the South African aged population using the method of extinct generations", *African Population Studies* **25**(1 (Supplement)):63-76.

Makeham, W.M. 1860. "On the law of mortality and the construction of annuity tables", *Journal of the Institute of Actuaries* **8**(6):301-310.

Makeham, W.M. 1867. "On the law of mortality", *Journal of the Institute of Actuaries* **13**(6):325-358.

Medvedev, Z.A. 1974. "Caucasus and Altay longevity: A biological or social problem?", *The Gerontologist* **14**(5 Part 1):381-387.

Meslé, F. and J. Vallin. 2002. "Improving the Accuracy of Life Tables for the Oldest Old", *Population* **57**(4):601-629.

Moultrie, T. and R. Dorrington. 2015. "Where do they all come from? Where do they all belong? Searching for Eleanor Rigby in the 2011 Census data on migration.", Paper presented at Seventh African Population Conference. Johannesburg.

Moultrie, T., R. Dorrington, A. Hill, K. Hill *et al.* 2013. *Tools for demographic estimation*. http://demographicestimation.iussp.org/. Accessed: 2016/01/31

Moultrie, T.A. and R.E. Dorrington. 2012. "Used for ill; used for good: a century of collecting data on race in South Africa", *Ethnic and Racial Studies* **35**(8):1447-1465.

Murray, C.J., J.K. Rajaratnam, J. Marcus, T. Laakso *et al.* 2010. "What can we conclude from death registration? Improved methods for evaluating completeness", *PLoS Med* **7**(4):e1000262.

Myers, R.J. 1954. "Accuracy of age reporting in the 1950 United States census", *Journal of the American Statistical Association* **49**(268):826-831.

Myers, R.J. 1976. "An instance of reverse heaping of ages", *Demography* **13**(4):577-580.

Ntozi, J.P.M. 1978. "The Demeny-Shorter and three-census methods for correcting age data", *Demography* **15**(4):509-521.

Ortega, A. and V. Garcia. 1986. *Estudio experimental sobre la mortalidad y algunas características socioeconómicas de las personas de la tercera edad: informe de la investigación efectuada en los cantones de Puriscal y Coronado, 3-20 junio 1985*. San José, Costa Rica: Naciones Unidas Centro Latinoamericano De Demografía (CELADE).

Palloni, A., G. Pinto and H. Beltrán-Sánchez. 2016. "Estimation of Life Tables in the Latin American Data Base (LAMBdA): Adjustments for Relative Completeness and Age Misreporting", Paper presented at United Nations Expert Group Meeting New York, 3 November 2016.

Perks, W. 1932. "On some experiments in the graduation of mortality statistics", *Journal of the Institute of Actuaries (1886-1994)* **63**(1):12-57.

Preston, S.H. 1984. *Use of direct and indirect techniques for estimating the completeness of death registration systems*. New York: Population Division, Department of Economic and Social Affairs, United Nations.

Preston, S.H. and N.G. Bennett. 1983. "A census-based method for estimating adult mortality", *Population Studies* **37**(1):91-104.
Preston, S.H. and A.J. Coale. 1982. "Age structure, growth, attrition, and accession: A new synthesis", *Population Index*:217-259.
Preston, S.H., A.J. Coale, J. Trussell and M. Weinstein. 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index* **46**(2):179-202.
Preston, S.H. and I.T. Elo. 2006. "Black Mortality at Very Old Ages in Official Us Life Tables: A Skeptical Appraisal", *Population and Development Review* **32**(3):557-565.
Preston, S.H., I.T. Elo, I. Rosenwaike and M. Hill. 1996. "African-American mortality at older ages: Results of a matching study", *Demography* **33**(2):193-209.
Preston, S.H., I.T. Elo and Q. Stewart. 1999. "Effects of age misreporting on mortality estimates at older ages", *Population Studies* **53**(2):165-177.
Preston, S.H. and K. Hill. 1980. "Estimating the completeness of death registration", *Population Studies* **34**(2):349-366.
Preston, S.H. and S. Lahiri. 1991. "A short-cut method for estimating death registration completeness in destabilized populations*", *Mathematical Population Studies* **3**(1):39-51.
Preston, S.H. and A. Stokes. 2012. "Sources of population aging in more and less developed countries", *Population and Development Review* **38**(2):221-236.
Pullum, T.W. 1991. "Statistical methods to adjust for date and age misreporting to improve estimates of vital rates in Pakistan", *Statistics in Medicine* **10**(2):191-200.
R Development Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/
Renshaw, A.E. and R.J. Verrall. 1998. "A stochastic model underlying the chain-ladder technique", *British Actuarial Journal* **4**(4):903-923.
Rosenwaike, I. 1968. "On measuring the extreme aged in the population", *Journal of the American Statistical Association* **63**(321):29-40.
Rosenwaike, I. 1979. "A new evaluation of United States census data on the extreme aged", *Demography* **16**(2):279-288.
Rosenwaike, I. 1981. "A Note on New Estimates of the Mortality of the Extreme Aged", *Demography* **18**(2):257-266.
Rosenwaike, I. and M.E. Hill. 1996. "The accuracy of age reporting among elderly african americans evidence of a birth registration effect", *Research on Aging* **18**(3):310-324.
Rosenwaike, I. and B. Logue. 1983. "Accuracy of death certificate ages for the extreme aged", *Demography* **20**(4):569-585.
Rubin, D.B. 1976. "Inference and missing data", *Biometrika* **63**(3):581-592.
Sadie, J.L. 1951. "Differential mortality in South Africa", *South African Journal of Economics* **19**(4):361-369.
Sadie, J.L. 1988. *A reconstruction and projection of demographic movements in the RSA and TBVC countries*. Research Report No. 148. Pretoria: Bureau of Market Reserach, University of South Africa.
Saxena, P. and B. Gogte. 1985. "On Feeneys method for correcting age distributions for heaping on multiples of five", *Asian and Pacific Census Forum* **11**(3):5-9.
Siegel, J.S. and J.S. Passel. 1976. "New estimates of the number of centenarians in the United States", *Journal of the American Statistical Association* **71**(355):559-566.
Stats SA. 2006. *Mortality and Causes of Death in South Africa, 2003 and 2004: Findings from Death Notification*. Statistical Release P0309.3. Pretoria: Statistics South Africa.

Stats SA. 2008. *Mortality and causes of death in South Africa, 2006: Findings from death notification*. Statistical Release P0309.3. Pretoria: Statistics South Africa.
Stats SA. 2009. *Mortality and causes of death in South Africa, 2007: Findings from death notification*. Statistical Release P0309.3. Pretoria: Statistics South Africa.
Stats SA. 2012. *Social profile of vulnerable groups in South Africa, 2002–2011*. Report No. 03-19-00 (2002–2012) Pretoria: Statistics South Africa.
Stats SA. 2014. *Mortality and Causes of Death in South Africa, 2011: Findings from Death Notification*. Statistical Release P0309.3. Pretoria: Statistics South Africa.
Stats SA. 2015. *Mid-year population estimates*. P0302. Pretoria: Statistics South Africa.
Taylor, G. 2012. *Loss reserving: an actuarial perspective*. Springer Science & Business Media.
Terblanche, W. and T. Wilson. 2014. "Understanding the Growth of Australia's Very Elderly Population, 1976 to 2012", *Journal of Population Ageing* **7**(4):301-322.
Terblanche, W. and T. Wilson. 2015. "An Evaluation of Nearly-Extinct Cohort Methods for Estimating the Very Elderly Populations of Australia and New Zealand", *PloS one* **10**(4):e0123692.
Thatcher, A.R. 1987. "Mortality at the highest ages", *Journal of the Institute of Actuaries* **114**(2):327-338.
Thatcher, A.R. 1992. "Trends in numbers and mortality at high ages in England and Wales", *Population Studies* **46**(3):411-426.
Thatcher, A.R. 1999. "The long-term pattern of adult mortality and the highest attained age", *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **162**(1):5-43.
Thatcher, A.R., V. Kannisto and K. Andreev. 2002. "The survivor ratio method for estimating numbers at high ages", *Demographic Research* **6**(1):2-15.
Thatcher, A.R., V. Kannisto and J.W. Vaupel. 1998. *The force of mortality at ages 80 to 120*. Odense: Odense University.
Timæus, I.M. 2004. "Impact of HIV on mortality in Southern Africa: Evidence from demographic surveillance", Paper presented at Seminar of the IUSSP Committee "Emerging Health Threats - HIV, Resurgent Infections and Population Change in Africa". Ougadougou, 12-14 February.
Timaeus, I.M. 1991. "Measurement of adult mortality in less developed countries: a comparative review", *Population Index* **57**(4):552-568.
Towers Perrin. 2009. *Longevity Risk Investigation*. Madrid, Spain: Towers Perrin.
United Nations. 1982. *Model life tables for developing countries*. New York: Population Division, Department of Economic and Social Affairs.
United Nations. 2013. *World Population Prospects: The 2012 Revision*. New York: Population Division, Department of Economic and Social Affairs.
United Nations. 2015. *World Population Prospects: The 2015 Revision*. New York: Population Division, Department of Economic and Social Affairs.
United Nations. 2017. *World Population Prospects: The 2017 Revision*. New York: Population Division, Department of Economic and Social Affairs.
US Census Bureau. 2015. *International Database*. Washington DC: International Programs Center.
Van Tonder, J.L. and I. Van Eeden. 1975. *Abridged Life Tables for All the Population Groups in the Republic of South Africa (1921-70)*. Institute for Sociological, Demographic and Criminological Research, Human Sciences Research Council.
Verrall, R. 1996. "Claims reserving and generalised additive models", *Insurance: mathematics and economics* **19**(1):31-43.
Vincent, P. 1951. "La mortalité des vieillards", *Population (french edition)* **6**(2):181-204.
Whelan, S. 2009a. "Mortality in Ireland at advanced ages, 1950-2006: Part 1: Crude rates", *Annals of Actuarial Science* **4**(1):33-66.

Whelan, S. 2009b. "Mortality in Ireland at advanced ages, 1950-2006: Part 2: Graduated rates", *Annals of Actuarial Science* **4**(1):67-104.

Wilmoth, J.R. 1995. "Are mortality rates falling at extremely high ages? An investigation based on a model proposed by Coale and Kisker", *Population Studies* **49**(2):281-295.

Wilmoth, J.R., K. Andreev, D. Jdanov, D.A. Glei *et al.* 2007. *Methods protocol for the human mortality database*. Rostock: Max Planck Institute for Demographic Research. http://mortality.org

Wüthrich, M. and M. Merz. 2008. *Stochastic claims reserving methods in insurance.* John Wiley & Sons.

Yi, Z. and J.W. Vaupel. 2006. "Oldest-old mortality in China," in Robine, J M , E M Crimmins, S Horiuchi and Y Zeng (eds). *Human Longevity, Individual Life Duration, and the Growth of the Oldest-Old Population*. Springer, pp. 87-110.

Zehnwirth, B. 1994. "Probabilistic development factor models with applications to loss reserve variability, prediction intervals and risk based capital", *Casualty Actuarial Society Forum* **2**:447-606.

APPENDIX A - RESULTS FROM CHAPTER 2

A.1 Mortality rates implied by the Das Gupta method

$$\hat{N}(x,t) = D(x-1,t-1) * \sum_{i=1}^{\omega-x+1} \prod_{j=1}^{i} CR(x+j-1,t+j-1,m)$$

$$= D(x-1,t-1) * CR(x,t,m) * (1 + \sum_{i=2}^{\omega-x+1} \prod_{j=2}^{i} CR(x+j-1,t+j-1,m))$$

$$= \hat{D}(x,t) * (1 + \sum_{i=2}^{\omega-x+1} \prod_{j=2}^{i} CR(x+j-1,t+j-1,m))$$

$$\rightarrow \frac{\hat{D}(x,t)}{\hat{N}(x,t)} = \hat{q}_{x,t} = \frac{1}{\left(1 + \sum_{i=2}^{\omega-x+1} \prod_{j=2}^{i} CR(x+j-1,t+j-1,m)\right)}$$

A.2 Survivor Ratio expressed as a weighted average

$$SR(x,t,k,m) = \frac{\sum_{j=1}^{m} N(x,t-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$$

$$= \frac{\dfrac{\sum_{j=1}^{m} N(x,t-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)} * \sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$$

$$= \frac{\sum_{j=1}^{m}\left(\dfrac{N(x,t-j)}{\sum_{i=1}^{k} D(x-i,t-i-j)} * \sum_{i=1}^{k} D(x-i,t-i-j)\right)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$$

$$= \frac{\sum_{j=1}^{m}\left(SR(x,t-j,k,1) * \sum_{i=1}^{k} D(x-i,t-i-j)\right)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$$

A.3 Equivalence of definitions of the Survivor Ratio given by Andreev (1999) and Thatcher, Kannisto and Andreev (2002)

$$\frac{\tilde{s}(x-k,t-k,k)}{1-\tilde{s}(x-k,t-k,k)} = \frac{\dfrac{\sum_{i=1}^{m} N(x,t-i)}{\sum_{i=1}^{m} N(x-k,t-k-i)}}{1-\dfrac{\sum_{i=1}^{m} N(x,t-i)}{\sum_{i=1}^{m} N(x-k,t-k-i)}}$$

$$= \frac{\sum_{i=1}^{m} N(x,t-i) / \sum_{i=1}^{m} N(x-k,t-k-i)}{\left(\sum_{i=1}^{m} N(x-k,t-k-i) - N(x,t-i)\right) / \sum_{i=1}^{m} N(x-k,t-k-i)}$$

$$= \frac{\sum_{i=1}^{m} N(x,t-i)}{\sum_{i=1}^{m}\sum_{j=1}^{k} D(x-j,t-j-i)}$$

$$= SR(x,t,k,m)$$

A.4 Expression of the Survivor Ratio method as an adaptation of the Das Gupta method

A.4.1 Rearrangement of the product of cohort ratios

$$\sum_{i=1}^{\omega-x+1} \prod_{j=1}^{i} CR(x+j-1,t+j-1,m)$$

$$= CR(x,t,m) + CR(x,t,m) * CR(x+1,t+1,m) + \ldots$$
$$+ CR(x,t,m) * CR(x+1,t+1,m) * \ldots * CR(\omega,t+\omega-x,m)$$

$$= \frac{\sum_{j=1}^{m} D(x,t-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)} + \frac{\sum_{j=1}^{m} D(x,t-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)} * \frac{\sum_{j=1}^{m} D(x+1,t+1-j)}{\sum_{j=1}^{m} D(x,t-j)} + \ldots$$

$$+ \frac{\sum_{j=1}^{m} D(x,t-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)} * \frac{\sum_{j=1}^{m} D(x+1,t+1-j)}{\sum_{j=1}^{m} D(x,t-j)} * \ldots * \frac{\sum_{j=1}^{m} D(\omega,t+\omega-x-j)}{\sum_{j=1}^{m} D(\omega-1,t+\omega-x-1-j)}$$

$$= \frac{\sum_{j=1}^{m} D(x,t-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)} + \frac{\sum_{j=1}^{m} D(x+1,t+1-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)} + \ldots + \frac{\sum_{j=1}^{m} D(\omega,t+\omega-x-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)}$$

$$= \sum_{i=0}^{\omega-x} \frac{\sum_{j=1}^{m} D(x+i,t+i-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)}$$

$$\rightarrow \sum_{i=1}^{\omega-x+1} \prod_{j=1}^{i} CR(x+j-1,t+j-1,m) = \sum_{i=0}^{\omega-x} \frac{\sum_{j=1}^{m} D(x+i,t+i-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)}$$

A.4.2 Redefinition of the Survivor Ratio method

$$\hat{N}(x,t) = \sum_{i=1}^{k} D(x-i,t-i) * SR(x,t,k,m)$$

where $SR(x,t,k,m) = \frac{\sum_{j=1}^{m} N(x,t-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)}$

Thus $\hat{N}(x,t) = \sum_{i=1}^{k} D(x-i,t-i) * \frac{\sum_{j=1}^{m} N(x,t-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)})$

$$= \sum_{i=1}^{k} D(x-i,t-i) * \frac{\sum_{j=1}^{m}\sum_{i=0}^{\omega-x} D(x+i,t+i-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)})$$

$$= \sum_{i=1}^{k} D(x-i,t-i) * \sum_{i=0}^{\omega-x} \frac{\sum_{j=1}^{m} D(x+i,t+i-j)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)})$$

$$= \sum_{i=1}^{k} D(x-i,t-i) * \frac{\sum_{j=1}^{m} D(x-1,t-j-1)}{\sum_{j=1}^{m}\sum_{i=1}^{k} D(x-i,t-i-j)} \sum_{i=0}^{\omega-x} \frac{\sum_{j=1}^{m} D(x+i,t+i-j)}{\sum_{j=1}^{m} D(x-1,t-j-1)})$$

As shown previously,

$$\sum_{i=1}^{\omega-x+1}\prod_{j=1}^{i}CR(x+j-1,t+j-1,m)=\sum_{i=0}^{\omega-x}\frac{\sum_{j=1}^{m}D(x+i,t+i-j)}{\sum_{j=1}^{m}D(x-1,t-j-1)}$$

$$\hat{N}(x,t)=\sum_{i=1}^{k}D(x-i,t-i)*SR(x,t,k,m)=$$

$$\sum_{i=1}^{k}D(x-i,t-i)*\frac{\sum_{j=1}^{m}D(x-1,t-j-1)}{\sum_{j=1}^{m}\sum_{i=1}^{k}D(x-i,t-i-j)}\sum_{i=1}^{\omega-x+1}\prod_{j=1}^{i}CR(x+j-1,t+j-1,m)$$

The Das Gupta method is produced when $k=1$, i.e.:

$$\hat{N}(x,t)=D(x-1,t-1)*SR(x,t,1,m)$$

$$=D(x-1,t-1)*\sum_{i=1}^{\omega-x+1}\prod_{j=1}^{i}CR(x+j-1,t+j-1,m)$$

When $k>1$, instead of projecting future deaths using $D(x-1,t-1)$, $D(x-1,t-1)$ is replaced by a smoothed estimate based on the experience of the m previous cohorts. The ratio of deaths at age $x-1$ to deaths at the ages from $x-1$ to $x-k$ is calculated in the m previous cohorts, and then the deaths at the ages from $x-1$ to $x-k$ in the current cohort are multiplied by this ratio. Since the ratio is based on more than one year of experience, it can be expected that the estimate of $D(x-1,t-1)$ will fluctuate less than the observed numbers in single years.

A.5 Effect of c on the estimates of mortality

$$\hat{N}(x,t)=D(x-1,t-1)*\sum_{i=1}^{\omega-x+1}\prod_{j=1}^{i}CR(x+j-1,t+j-1,m)*c$$

$$\hat{N}(x,t)=\hat{D}(x,t)*c*\left(1+\sum_{i=2}^{\omega-x+1}\prod_{j=2}^{i}CR(x+j-1,t+j-1,m)\right)$$

$$\rightarrow\frac{\hat{D}(x,t)}{\hat{N}(x,t)}=\hat{q}_{x,t}=\frac{1}{\left(1+\sum_{i=2}^{\omega-x+1}\prod_{j=2}^{i}CR(x+j-1,t+j-1,m)\right)}*\frac{1}{c}$$

When mortality rates are falling, c will be greater than unity, and therefore mortality rates will be reduced in proportion to c. Since the mortality rates implied by the Das Gupta and Survivor Ratio methods are the same, as shown above, this point is equally applicable to both methods.

APPENDIX B - RESULTS FROM CHAPTER 3

B.1 Population data, 1985 and 1991

B.1.1 1985

Table B.1: Population of South Africa, 1985, Males and Females

	Stats SA - Coloureds, Indians and Whites		Dorrington (1999) - Africans		Total	
Age group	Males	Females	Males	Females	Males	Females
0-4	415 036	403 503	1 427 164	1 439 138	1 842 200	1 842 641
5-9	398 943	386 980	1 408 114	1 414 645	1 807 057	1 801 625
10-14	456 305	443 589	1 365 284	1 395 373	1 821 589	1 838 962
15-19	423 086	418 473	1 082 317	1 197 588	1 505 403	1 616 061
20-24	378 156	396 162	914 491	1 093 225	1 292 647	1 489 387
25-29	340 656	356 382	769 642	902 135	1 110 298	1 258 517
30-34	303 060	317 256	598 215	694 281	901 275	1 011 537
35-39	276 559	284 512	495 217	569 630	771 776	854 142
40-44	233 604	237 163	407 116	480 480	640 720	717 643
45-49	197 537	200 422	356 308	401 811	553 845	602 233
50-54	162 540	170 430	265 867	317 024	428 407	487 454
55-59	133 063	144 335	204 372	250 158	337 435	394 493
60-64	110 212	128 975	171 063	267 607	281 275	396 582
65-69	83 428	101 417	154 243	217 354	237 671	318 771
70-74	62 494	85 832	91 585	131 369	154 079	217 201
75-79	36 879	58 145	56 501	79 691	93 380	137 836
80-84	16 694	32 026	27 768	46 726	44 462	78 752
85+	8 278	20 673	23 971	42 034	32 249	62 707
Total	4 036 530	4 186 275	9 819 238	10 940 269	13 855 768	15 126 544

B.1.2 1991

Table B.2: Population of South Africa, 1991, Males and Females

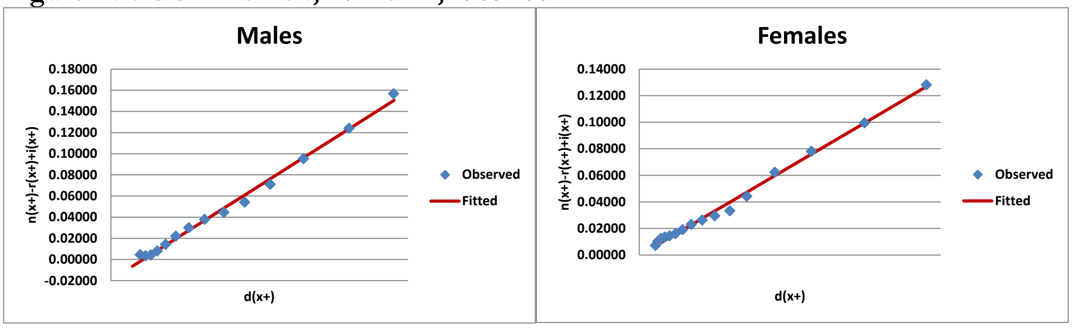
	CSS(1992a) - Total population of RSA		CSS(1992b) - Estimated population of TBVC countries		Total	
	Male	Female	Male	Female	Male	Female
0-4	1 346 636	1 350 024	567 200	561 400	1 913 836	1 911 424
5-9	1 468 035	1 455 956	568 900	562 500	2 036 935	2 018 456
10-14	1 405 193	1 419 704	508 500	502 100	1 913 693	1 921 804
15-19	1 339 384	1 411 625	396 200	394 300	1 735 584	1 805 925
20-24	1 238 065	1 342 672	265 200	273 100	1 503 265	1 615 772
25-29	1 165 972	1 231 684	186 300	230 400	1 352 272	1 462 084
30-34	1 110 298	1 101 897	140 900	202 000	1 251 198	1 303 897
35-39	871 068	870 161	123 800	167 800	994 868	1 037 961
40-44	745 247	743 730	92 700	129 600	837 947	873 330
45-49	560 837	554 875	78 700	110 400	639 537	665 275
50-54	470 727	488 360	62 700	93 600	533 427	581 960
55-59	325 502	365 626	49 700	78 400	375 202	444 026
60-64	276 856	384 490	45 100	67 000	321 956	451 490
65-69	197 568	261 852	44 000	64 100	241 568	325 952
70-74	161 783	220 475	29 100	51 800	190 883	272 275
75-79	78 780	117 437	20 200	37 600	98 980	155 037
80-84	46 101	80 024	9 200	20 200	55 301	100 224
85+	25 961	53 785	5 000	11 000	30 961	64 785
Total	12 834 013	13 454 377	3 193 400	3 557 300	16 027 413	17 011 677

B.2 Results of the GGB method

B.2.1 1985-1991

The GGB method produced an estimate of completeness of 53% for males and 61% for females in the period 1985-1991.

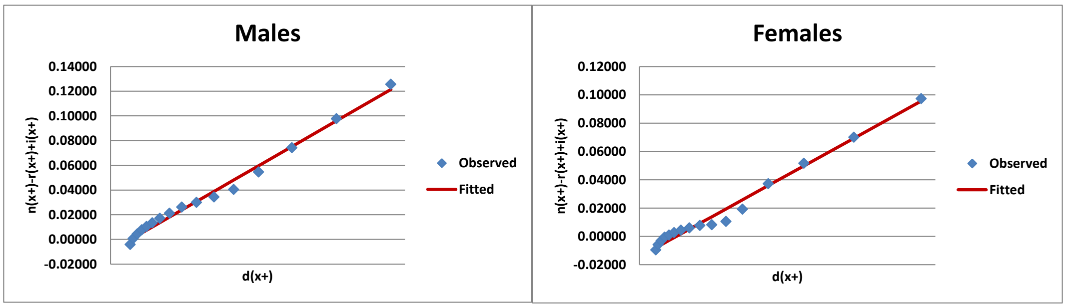
Figure B.1: GGB method, National, 1985-1991



B.2.2 1991-1996

Estimates of 78% and 75% for males and females respectively were produced by the GGB method in the period 1991-1996, with the somewhat poor fits shown in Figure B.2.

Figure B.2: GGB method, National, 1991-1996

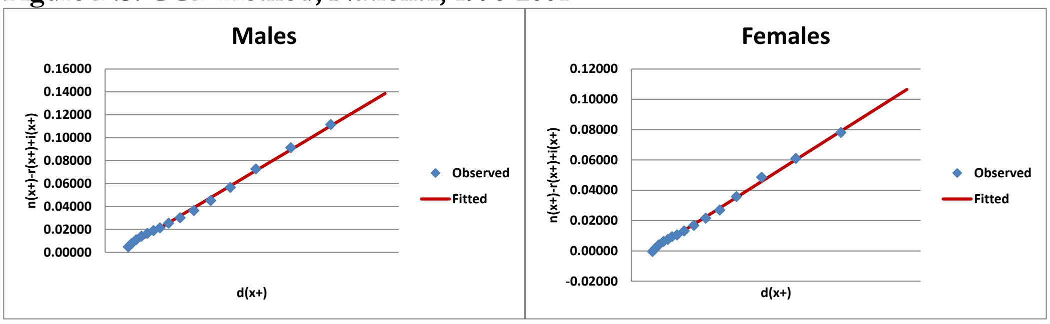


B.2.3 1996-2001

B.2.3.1 National

The GGB method produced estimates of completeness of 81% and 85% for males and females respectively in the period 1996-2001. A large residual was produced in the 85+ age group for males, and this point was excluded from the estimate. These results are shown in Figure B.3.

Figure B.3: GGB method, National, 1996-2001



B.2.3.2 Population Group

The GGB method produced estimates of completeness as shown in Table B.3 and Figure B.4 and Figure B.5.

Table B.3: Estimates of completeness, GGB method, 1996-2001

	Males	Females
African	55%	59%
Coloured	63%	65%
Indian	72%	82%
White	72%	76%

Figure B.4: Application of the GGB method, 1996-2001, Males

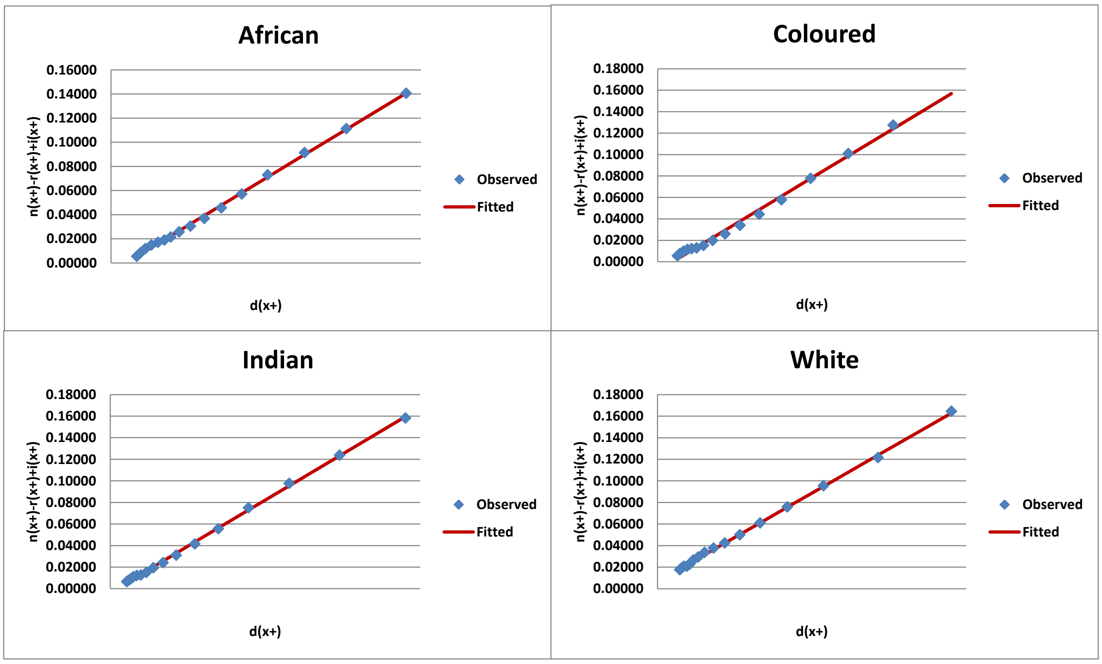
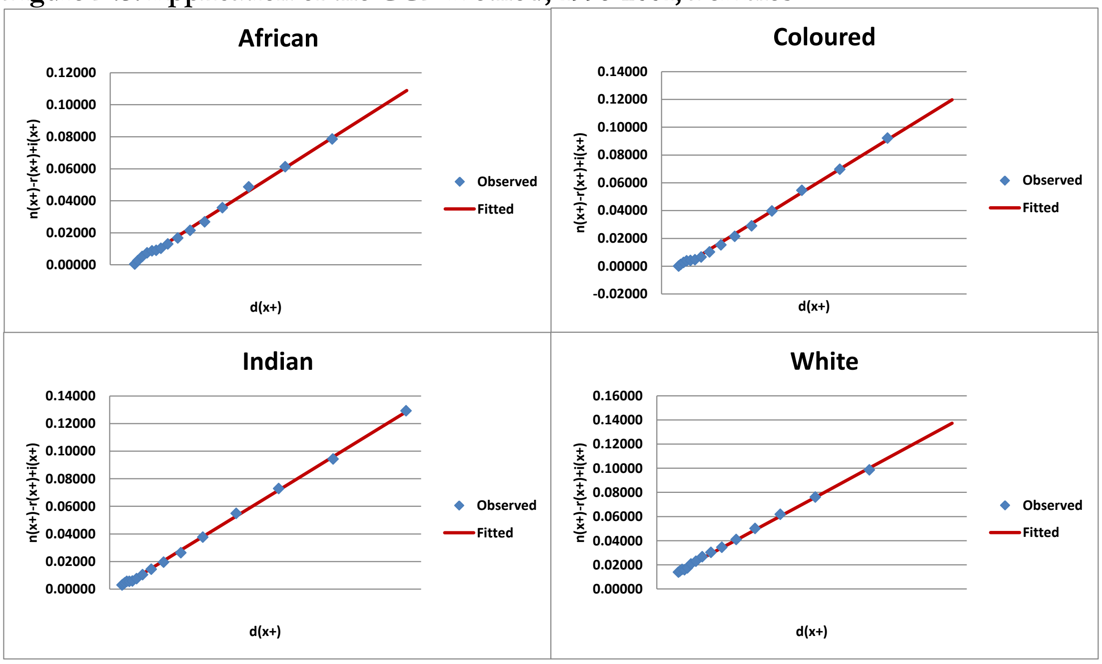


Figure B.5: Application of the GGB method, 1996-2001, Females



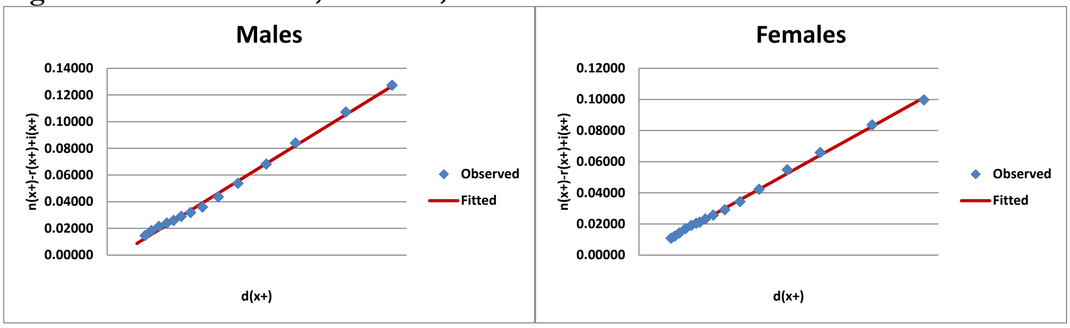
The method produced large residuals for males from the Coloured population group aged 80-84, and for females from the African, Coloured and White population groups, and in these cases, the age group 80-84 was excluded from the GGB calculations.

B.2.4 2001-2011

B.2.4.1 National

The GGB diagnostics showed large residuals until age 15 for both sexes, and at ages 80-84 for females. The regression plots, after excluding these points, are shown in Figure B.6.

Figure B.6: GGB method, National, 2001-2011



The GGB method applied to data for males and females produced estimates of completeness of 89% and 94% in the period 2001-2011.

B.2.4.2 Population group

The GGB method produced estimates of completeness as shown in Table B.4, Figure B.7 and Figure B.8.

Table B.4: Estimates of completeness, GGB method, 2001-2011

	Males	Females
African	62%	65%
Coloured	62%	62%
Indian	89%	97%
White	73%	77%

Figure B.7: Application of the GGB method, 2001-2011, Males

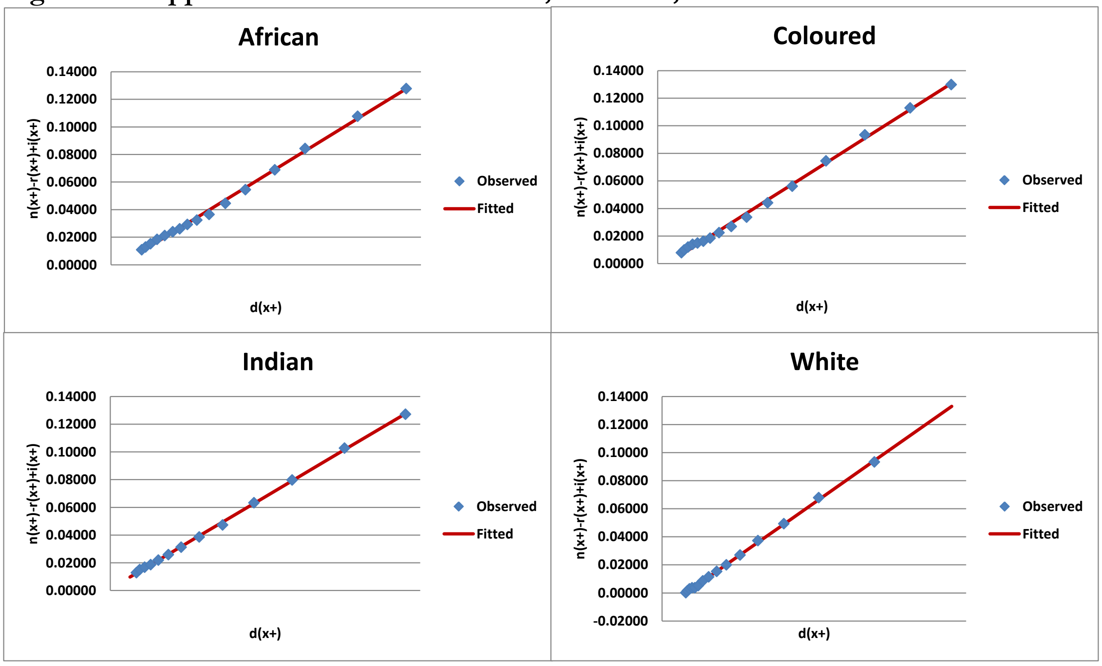
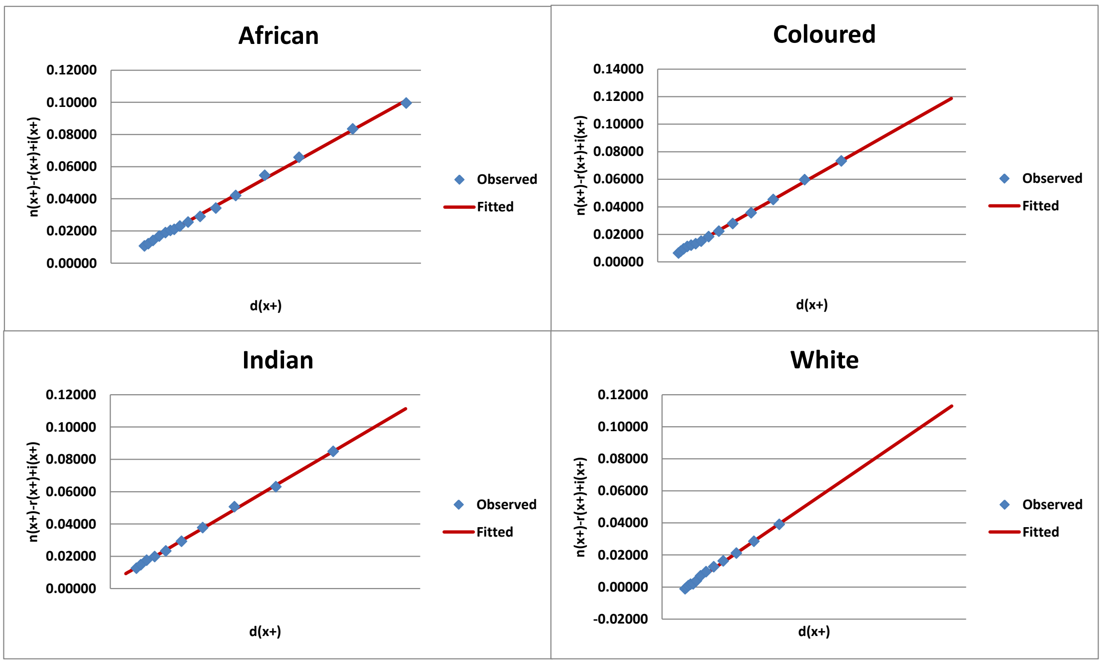


Figure B.8: Application of the GGB method, 2001-2011, Females



B.2.5 Annual estimates of completeness

The blank entries in Table B.5 represent the years in which population group was not recorded in the death data.

Table B.5: Annual estimates of completeness, 1984-2013

Year	National	African	White	Indian	Coloured
1984	54%	39%	100%	100%	100%
1985	54%	39%	100%	100%	100%
1986	54%	39%	100%	100%	100%
1987	54%	39%	100%	100%	100%
1988	54%	40%	100%	100%	100%
1989	54%	39%	100%	100%	100%
1990	58%	38%	100%	100%	100%
1991	62%				
1992	65%				
1993	68%				
1994	71%				
1995	73%				
1996	76%				
1997	77%				
1998	79%				
1999	81%	56%	66%	76%	67%
2000	82%	60%	76%	86%	67%
2001	83%	60%	77%	90%	67%
2002	84%	60%	78%	91%	67%
2003	85%	61%	78%	91%	67%
2004	85%	61%	78%	91%	67%
2005	86%	62%	78%	91%	67%
2006	86%	62%	78%	91%	67%
2007	87%	62%	78%	91%	67%
2008	87%	62%	78%	91%	67%
2009	87%	62%	77%	91%	66%
2010	87%	64%	77%	91%	66%
2011	87%	70%	82%	85%	74%
2012[51]	86%	71%	81%	83%	72%
2013[38]	83%	71%	78%	80%	70%

[51] 2013 completeness, and to a lesser extent, 2012 completeness, is likely to increase significantly with the inclusion of late registrations for these years in the 2014 and 2015 reporting on deaths.

APPENDIX C – RESULTS FROM CHAPTER 4

C.1.1 Fitted models

Table C.1: Fitted NEG-GAM models, National and Population Group models, Males and Females

Term	National	National	African	African	Coloured	Coloured	Indian	Indian	White	White
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Intercept	6.67	6.42	5.44	5.04	8.27	7.46	4.15	2.99	5.90	5.31
Spline - t|qx 1	7.82	34.26	32.14	27.85	-34.34	-123.57	34.72	-5.23	0.49	26.76
Spline - t|qx 2	-102.04	-26.11	-97.74	-75.44	-297.50	-187.06	-252.59	-143.92	-171.89	-92.64
Spline - t|qx 3	17.97	-10.58	204.84	176.62	-165.35	-84.68	55.38	26.84	13.05	-1.98
Spline - t|qx*Year of Death 1	-0.00	-0.02	-0.02	-0.01	0.02	0.06	-0.02	0.00	-0.00	-0.01
Spline - t|qx*Year of Death 2	0.05	0.01	0.05	0.04	0.15	0.09	0.13	0.07	0.09	0.05
Spline - t|qx*Year of Death 3	-0.01	0.00	-0.10	-0.09	0.08	0.04	-0.03	-0.02	-0.01	-0.00
Spline - Year of Birth 1	1.65	0.78	1.76	1.37	-3.04	-1.57	-0.41	0.14	1.12	1.12
Spline - Year of Birth 2	1.16	1.76	2.37	2.92	-2.78	-2.14	-0.13	1.45	0.06	0.59
Spline - Year of Birth 3	1.55	1.79	2.38	2.90	-2.39	-1.60	0.80	1.77	0.58	0.86

C.1.2 Control variables

The tables in this section present the effect of the control variables over and above the baseline smooth functions. Using age as an example, these values were derived as $Effect = e^{\alpha_x} - 1$, where α represents the control variable for age x.

Table C.2: Control variables, Age

Term	National	National	African	African	Coloured	Coloured	Indian	Indian	White	White
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Age 70	8%									
Age 77								-6%		
Age 79	4%	4%								
Age 80	4%									
Age 89		4%	11%	8%						
Age 90				7%		4%				
Age 93				-5%	-15%					
Age 96	-10%	-7%	-15%			-14%				
Age 97		-5%				-23%	96%			
Age 98	5%	0%	4%	6%		-24%				

Age 99	70%	21%	85%	33%	140%	12%	230%	89%	-5%	
Age 100	122%	29%	135%	51%	209%	44%	425%	107%		
Age 101	71%		64%	19%	143%	29%				
Age 102					126%					
Age 104					286%				-16%	
Age 105	80%									
Age 106	98%	46%			441%					
Age 107	155%	78%			1075%	145%				
Age 108	233%	117%				286%				

Table C.3: Control variables, Year of Birth

Year of Birth	National	National	African	African	Coloured	Coloured	Indian	Indian	White	White
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1886	463%	201%	231%	214%			-4%			
1887	349%	143%	216%	179%				-41%		
1888	211%	101%	133%	134%			68%			
1889	167%	82%	126%	116%				98%		
1890	106%	58%	84%	88%				59%		
1892							-56%			
1893							-40%			
1896							30%	-19%		
1897						-43%		-31%		-11%
1898						-48%	-29%			-16%
1899		-26%		-11%		-52%		37%	-13%	-18%
1900	73%	63%	139%	143%	-33%	-18%	33%	44%	-23%	-26%
1901	42%	26%	122%	110%			-6%	10%	-38%	-40%
1902	26%	20%	73%	69%	24%		3%	8%	-23%	-24%
1903									-12%	-13%
1904		35%		29%	-15%					
1905					-38%	-12%				
1906					-22%					
1907							-17%			
1909					2%					
1910	38%	52%	58%	89%	10%	7%	5%	6%	-14%	-16%
1911	-17%	-10%	-19%	-9%					9%	7%
1912		18%		24%						
1913	-24%	-19%	-32%	-29%						6%
1914	45%	46%	77%	72%	17%	10%			7%	11%
1915	-17%	-13%	-17%	-14%	8%	17%			-13%	-11%
1916	-13%	-6%	-16%	-9%	-10%	-9%	26%	25%		3%
1917	-25%	-18%	-31%	-23%	-13%	-17%	-30%	-30%	-11%	-9%
1918	47%	64%	89%	111%	6%	4%			-6%	-3%
1919		17%	19%	39%					-6%	-6%
1920	28%	33%	57%	61%	14%	11%	-19%	-11%		
1921					-14%	-16%	15%	21%	4%	

1922		-11%		-17%						
1923	-16%	-19%	-21%	-28%						
1924					-7%					
1925					7%	10%				
1926					-10%	-9%	-10%	-9%		
1927					-5%					
1928				9%	3%					
1929		14%		17%	6%				4%	
1930	24%	43%	31%	58%	17%	7%			5%	4%
1931	-14%		-24%	-11%	-3%	-10%				
1932					7%	4%				
1933			-14%		9%	8%	-14%	-15%	-3%	-6%
1934	-11%		-18%							-6%
1935	-12%		-19%	-6%						
1936				6%						
1937				-9%		-5%	10%	4%		
1938			10%		3%			-7%	5%	
1939								8%		6%
1940	24%	26%	39%	34%						
1942	4%		26%		2%	0%	-8%	-5%		

Table C.4: Control variables, Year of Death

Year of Death	National	National	African	African	Coloured	Coloured	Indian	Indian	White	White
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1984		-3%		-5%			-9%		-2%	
1985							-8%			
1986	2%								-3%	-2%
1987	7%		5%							
1988	-3%	-6%	-6%	-10%					-9%	-8%
1989	-15%	-18%	-15%	-7%			-25%	-18%	-16%	-15%
1990	-16%	-20%								
1991	-11%	-17%								
1992	-16%	-22%								
1993	-11%	-16%								
1994	-6%	-12%								
1995		-7%								
1996	-2%	-8%								
1997		-3%								
1998		-6%								
1999	-2%	-3%	-3%	0%	-8%	-6%	1%	3%		3%
2000			0%	2%			3%	8%		2%
2002	4%	3%	5%	6%					3%	4%
2003		-4%								
2004		-4%	-5%	-5%						-2%
2006			1%				-8%	-2%		

2007			-1%				-12%	-10%	-2%	
2008							-13%	-13%		
2009			-1%	-3%			-13%	-14%		
2010	-1%	0%	-4%	-4%	-3%		-5%	-1%		
2011	-2%	-1%	-5%	-5%				5%		
2012	-3%	-1%	-7%	-6%	4%		-19%	-11%		-1%

Table C.5: Control variables, Age and Year of Death

Term	National	National	African	African	Coloured	Coloured	Indian	Indian	White	White
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Ages 99+ in 2008	210%	105%	239%	123%						
Age 98 in years prior to 1999	102%	108%	246%	265%			204%	210%	102%	69%
Ages 70, and 80 in years prior to 1999		10%								
Ages 70,79 and 80 in years prior to 1999			51%	66%						

C.1.3 Mortality rates, 1985-2011

The empty cells in the tables in this section in the earlier years (1985-1994) at the advanced ages represent mortality rates that could not be estimated using the NEG-GAM model because death data for those dying in these years at these ages were not available in the data.

Table C.6: m_x, National, Males, 1985-1998.

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.064	0.064	0.063	0.063	0.063	0.062	0.062	0.062	0.061	0.061	0.061	0.060	0.060	0.060
71	0.069	0.069	0.068	0.068	0.068	0.067	0.067	0.066	0.066	0.065	0.065	0.065	0.064	0.064
72	0.075	0.074	0.074	0.073	0.073	0.072	0.072	0.071	0.071	0.070	0.070	0.069	0.069	0.068
73	0.080	0.080	0.079	0.079	0.078	0.078	0.077	0.077	0.076	0.075	0.075	0.074	0.074	0.073
74	0.086	0.086	0.085	0.084	0.084	0.083	0.083	0.082	0.081	0.081	0.080	0.080	0.079	0.078
75	0.092	0.091	0.091	0.090	0.089	0.089	0.088	0.088	0.087	0.086	0.086	0.085	0.084	0.084
76	0.098	0.098	0.097	0.096	0.095	0.095	0.094	0.093	0.093	0.092	0.091	0.091	0.090	0.089
77	0.105	0.104	0.103	0.102	0.102	0.101	0.100	0.099	0.099	0.098	0.097	0.097	0.096	0.095
78	0.111	0.110	0.109	0.109	0.108	0.107	0.106	0.106	0.105	0.104	0.104	0.103	0.102	0.101
79	0.118	0.117	0.116	0.115	0.115	0.114	0.113	0.112	0.112	0.111	0.110	0.109	0.109	0.108
80	0.124	0.124	0.123	0.122	0.121	0.120	0.120	0.119	0.118	0.118	0.117	0.116	0.115	0.115
81	0.131	0.130	0.130	0.129	0.128	0.127	0.127	0.126	0.125	0.124	0.124	0.123	0.122	0.122
82	0.138	0.138	0.137	0.136	0.135	0.135	0.134	0.133	0.132	0.132	0.131	0.130	0.130	0.129
83	0.145	0.145	0.144	0.143	0.143	0.142	0.141	0.141	0.140	0.139	0.139	0.138	0.137	0.137
84	0.153	0.152	0.152	0.151	0.150	0.150	0.149	0.148	0.148	0.147	0.146	0.146	0.145	0.145
85	0.160	0.160	0.159	0.159	0.158	0.157	0.157	0.156	0.156	0.155	0.155	0.154	0.153	0.153
86	0.168	0.168	0.167	0.167	0.166	0.166	0.165	0.165	0.164	0.164	0.163	0.163	0.162	0.162
87	0.176	0.176	0.175	0.175	0.174	0.174	0.174	0.173	0.173	0.172	0.172	0.171	0.171	0.171
88	0.184	0.184	0.184	0.183	0.183	0.183	0.182	0.182	0.182	0.181	0.181	0.181	0.180	0.180
89	0.192	0.192	0.192	0.192	0.192	0.192	0.191	0.191	0.191	0.191	0.191	0.190	0.190	0.190
90	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.201	0.200	0.200	0.200	0.200	0.200
91	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.210	0.211	0.211	0.211	0.211	0.211
92	0.219	0.219	0.219	0.220	0.220	0.220	0.220	0.221	0.221	0.221	0.221	0.221	0.222	0.222
93	0.228	0.228	0.229	0.229	0.230	0.230	0.231	0.231	0.231	0.232	0.232	0.233	0.233	0.233
94	0.237	0.238	0.239	0.239	0.240	0.241	0.241	0.242	0.242	0.243	0.244	0.244	0.245	0.245
95	0.247	0.248	0.249	0.250	0.251	0.251	0.252	0.253	0.254	0.255	0.256	0.256	0.257	0.258
96	0.257	0.258	0.259	0.260	0.262	0.263	0.264	0.265	0.266	0.267	0.268	0.269	0.270	0.271
97	0.268	0.269	0.270	0.272	0.273	0.274	0.275	0.277	0.278	0.279	0.281	0.282	0.283	0.284
98	0.278	0.280	0.281	0.283	0.285	0.286	0.288	0.289	0.291	0.292	0.294	0.295	0.297	0.298
99		0.291	0.293	0.295	0.297	0.298	0.300	0.302	0.304	0.306	0.308	0.309	0.311	0.313
100			0.305	0.307	0.309	0.311	0.313	0.315	0.318	0.320	0.322	0.324	0.326	0.328
101				0.319	0.322	0.324	0.327	0.329	0.332	0.334	0.337	0.339	0.341	0.344
102					0.335	0.338	0.341	0.344	0.346	0.349	0.352	0.355	0.357	0.360
103						0.352	0.355	0.358	0.362	0.365	0.368	0.371	0.374	0.377
104							0.370	0.374	0.377	0.381	0.384	0.388	0.391	0.395
105								0.390	0.393	0.397	0.401	0.405	0.409	0.413
106									0.410	0.414	0.419	0.423	0.427	0.431
107										0.432	0.437	0.442	0.446	0.451
108											0.456	0.461	0.466	0.471
109												0.480	0.486	0.491
110													0.507	0.513

Table C.7: m_x, National, Males, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.059	0.059	0.058	0.058	0.058	0.057	0.057	0.057	0.056	0.056	0.056	0.055	0.055
71	0.063	0.063	0.063	0.062	0.062	0.061	0.061	0.061	0.060	0.060	0.059	0.059	0.059
72	0.068	0.067	0.067	0.067	0.066	0.066	0.065	0.065	0.064	0.064	0.063	0.063	0.062
73	0.073	0.072	0.072	0.071	0.071	0.070	0.070	0.069	0.069	0.068	0.068	0.067	0.067
74	0.078	0.077	0.077	0.076	0.076	0.075	0.074	0.074	0.073	0.073	0.072	0.072	0.071
75	0.083	0.083	0.082	0.081	0.081	0.080	0.079	0.079	0.078	0.078	0.077	0.076	0.076
76	0.089	0.088	0.087	0.087	0.086	0.085	0.085	0.084	0.083	0.083	0.082	0.082	0.081
77	0.095	0.094	0.093	0.092	0.092	0.091	0.090	0.090	0.089	0.088	0.088	0.087	0.086
78	0.101	0.100	0.099	0.099	0.098	0.097	0.096	0.096	0.095	0.094	0.094	0.093	0.092
79	0.107	0.106	0.106	0.105	0.104	0.104	0.103	0.102	0.101	0.101	0.100	0.099	0.099
80	0.114	0.113	0.112	0.112	0.111	0.110	0.110	0.109	0.108	0.107	0.107	0.106	0.105
81	0.121	0.120	0.119	0.119	0.118	0.117	0.117	0.116	0.115	0.115	0.114	0.113	0.112
82	0.128	0.128	0.127	0.126	0.125	0.125	0.124	0.123	0.123	0.122	0.121	0.121	0.120
83	0.136	0.135	0.135	0.134	0.133	0.133	0.132	0.131	0.131	0.130	0.129	0.129	0.128
84	0.144	0.143	0.143	0.142	0.141	0.141	0.140	0.140	0.139	0.138	0.138	0.137	0.137
85	0.152	0.152	0.151	0.151	0.150	0.150	0.149	0.148	0.148	0.147	0.147	0.146	0.146
86	0.161	0.161	0.160	0.160	0.159	0.159	0.158	0.158	0.157	0.157	0.156	0.156	0.155
87	0.170	0.170	0.169	0.169	0.169	0.168	0.168	0.167	0.167	0.167	0.166	0.166	0.165
88	0.180	0.179	0.179	0.179	0.179	0.178	0.178	0.178	0.177	0.177	0.177	0.176	0.176
89	0.190	0.190	0.189	0.189	0.189	0.189	0.188	0.188	0.188	0.188	0.188	0.187	0.187
90	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.199	0.199	0.199	0.199	0.199	0.199
91	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211	0.211
92	0.222	0.222	0.223	0.223	0.223	0.223	0.223	0.224	0.224	0.224	0.224	0.224	0.225
93	0.234	0.234	0.235	0.235	0.235	0.236	0.236	0.237	0.237	0.237	0.238	0.238	0.238
94	0.246	0.247	0.247	0.248	0.248	0.249	0.249	0.250	0.251	0.251	0.252	0.252	0.253
95	0.259	0.260	0.260	0.261	0.262	0.263	0.263	0.264	0.265	0.266	0.266	0.267	0.268
96	0.272	0.273	0.274	0.275	0.276	0.277	0.278	0.279	0.280	0.281	0.282	0.283	0.284
97	0.286	0.287	0.288	0.289	0.291	0.292	0.293	0.294	0.296	0.297	0.298	0.299	0.300
98	0.300	0.301	0.303	0.304	0.306	0.307	0.309	0.310	0.312	0.313	0.315	0.316	0.318
99	0.315	0.317	0.318	0.320	0.322	0.324	0.325	0.327	0.329	0.331	0.332	0.334	0.336
100	0.330	0.332	0.334	0.336	0.338	0.341	0.343	0.345	0.347	0.349	0.351	0.353	0.355
101	0.346	0.349	0.351	0.353	0.356	0.358	0.360	0.363	0.365	0.367	0.370	0.372	0.374
102	0.363	0.366	0.368	0.371	0.374	0.376	0.379	0.382	0.384	0.387	0.390	0.392	0.395
103	0.380	0.383	0.386	0.389	0.392	0.395	0.398	0.401	0.404	0.407	0.410	0.413	0.416
104	0.398	0.401	0.405	0.408	0.412	0.415	0.418	0.422	0.425	0.429	0.432	0.435	0.439
105	0.416	0.420	0.424	0.428	0.432	0.436	0.439	0.443	0.447	0.451	0.454	0.458	0.462
106	0.436	0.440	0.444	0.448	0.452	0.457	0.461	0.465	0.469	0.473	0.478	0.482	0.486
107	0.455	0.460	0.465	0.469	0.474	0.479	0.483	0.488	0.492	0.497	0.502	0.506	0.511
108	0.476	0.481	0.486	0.491	0.496	0.501	0.506	0.511	0.516	0.521	0.526	0.531	0.536
109	0.497	0.502	0.508	0.513	0.519	0.524	0.530	0.535	0.541	0.546	0.552	0.557	0.563
110	0.518	0.524	0.530	0.536	0.542	0.548	0.554	0.560	0.566	0.572	0.578	0.584	0.590

Table C.8: m_x, National, Females, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.038	0.038	0.038	0.038	0.038	0.038
71	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.041	0.041	0.041	0.041
72	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.044	0.044	0.044	0.044	0.044	0.044
73	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047
74	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.050
75	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.054	0.054
76	0.060	0.060	0.060	0.060	0.060	0.060	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
77	0.065	0.065	0.065	0.065	0.065	0.065	0.064	0.064	0.064	0.064	0.064	0.064	0.063	0.063
78	0.071	0.071	0.070	0.070	0.070	0.070	0.070	0.069	0.069	0.069	0.069	0.069	0.069	0.068
79	0.077	0.077	0.076	0.076	0.076	0.076	0.075	0.075	0.075	0.075	0.074	0.074	0.074	0.074
80	0.083	0.083	0.083	0.082	0.082	0.082	0.082	0.081	0.081	0.081	0.081	0.080	0.080	0.080
81	0.090	0.090	0.089	0.089	0.089	0.089	0.088	0.088	0.088	0.087	0.087	0.087	0.086	0.086
82	0.097	0.097	0.097	0.096	0.096	0.096	0.095	0.095	0.095	0.094	0.094	0.094	0.093	0.093
83	0.105	0.105	0.104	0.104	0.104	0.103	0.103	0.103	0.102	0.102	0.101	0.101	0.101	0.100
84	0.113	0.113	0.113	0.112	0.112	0.111	0.111	0.111	0.110	0.110	0.109	0.109	0.109	0.108
85	0.122	0.122	0.121	0.121	0.120	0.120	0.120	0.119	0.119	0.118	0.118	0.118	0.117	0.117
86	0.131	0.131	0.131	0.130	0.130	0.129	0.129	0.128	0.128	0.128	0.127	0.127	0.126	0.126
87	0.141	0.141	0.140	0.140	0.139	0.139	0.139	0.138	0.138	0.137	0.137	0.136	0.136	0.136
88	0.152	0.151	0.151	0.150	0.150	0.149	0.149	0.148	0.148	0.148	0.147	0.147	0.146	0.146
89	0.162	0.162	0.162	0.161	0.161	0.160	0.160	0.159	0.159	0.158	0.158	0.158	0.157	0.157
90	0.174	0.174	0.173	0.173	0.172	0.172	0.171	0.171	0.170	0.170	0.170	0.169	0.169	0.168
91	0.186	0.186	0.185	0.185	0.184	0.184	0.183	0.183	0.183	0.182	0.182	0.181	0.181	0.180
92	0.199	0.198	0.198	0.197	0.197	0.197	0.196	0.196	0.195	0.195	0.195	0.194	0.194	0.193
93	0.212	0.212	0.211	0.211	0.210	0.210	0.210	0.209	0.209	0.208	0.208	0.208	0.207	0.207
94	0.226	0.225	0.225	0.225	0.224	0.224	0.224	0.223	0.223	0.223	0.222	0.222	0.222	0.221
95	0.240	0.240	0.240	0.239	0.239	0.239	0.238	0.238	0.238	0.237	0.237	0.237	0.236	0.236
96	0.255	0.255	0.255	0.255	0.254	0.254	0.254	0.253	0.253	0.253	0.253	0.252	0.252	0.252
97	0.271	0.271	0.271	0.270	0.270	0.270	0.270	0.270	0.269	0.269	0.269	0.269	0.269	0.268
98	0.287	0.287	0.287	0.287	0.287	0.287	0.287	0.286	0.286	0.286	0.286	0.286	0.286	0.286
99		0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304
100			0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322
101				0.341	0.341	0.341	0.341	0.341	0.341	0.342	0.342	0.342	0.342	0.342
102					0.361	0.361	0.361	0.361	0.361	0.362	0.362	0.362	0.362	0.362
103						0.381	0.381	0.382	0.382	0.382	0.383	0.383	0.383	0.383
104							0.403	0.403	0.403	0.404	0.404	0.405	0.405	0.405
105								0.425	0.425	0.426	0.427	0.427	0.428	0.428
106									0.448	0.449	0.450	0.450	0.451	0.452
107										0.473	0.474	0.474	0.475	0.476
108											0.498	0.499	0.500	0.501
109												0.524	0.525	0.526
110													0.551	0.552

Table C.9: m_x, National, Females, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.038	0.038	0.038	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.040	0.040
71	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.042	0.042	0.042	0.042	0.042
72	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
73	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047
74	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
75	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.053	0.053	0.053
76	0.059	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.057	0.057	0.057	0.057
77	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.061	0.061
78	0.068	0.068	0.068	0.068	0.067	0.067	0.067	0.067	0.067	0.066	0.066	0.066	0.066
79	0.074	0.073	0.073	0.073	0.073	0.072	0.072	0.072	0.072	0.071	0.071	0.071	0.071
80	0.079	0.079	0.079	0.079	0.078	0.078	0.078	0.078	0.077	0.077	0.077	0.077	0.076
81	0.086	0.086	0.085	0.085	0.085	0.084	0.084	0.084	0.083	0.083	0.083	0.083	0.082
82	0.093	0.092	0.092	0.092	0.091	0.091	0.091	0.090	0.090	0.090	0.089	0.089	0.089
83	0.100	0.100	0.099	0.099	0.099	0.098	0.098	0.098	0.097	0.097	0.096	0.096	0.096
84	0.108	0.108	0.107	0.107	0.106	0.106	0.106	0.105	0.105	0.105	0.104	0.104	0.103
85	0.116	0.116	0.116	0.115	0.115	0.114	0.114	0.114	0.113	0.113	0.112	0.112	0.112
86	0.125	0.125	0.125	0.124	0.124	0.123	0.123	0.123	0.122	0.122	0.121	0.121	0.120
87	0.135	0.135	0.134	0.134	0.133	0.133	0.133	0.132	0.132	0.131	0.131	0.130	0.130
88	0.145	0.145	0.144	0.144	0.144	0.143	0.143	0.142	0.142	0.141	0.141	0.141	0.140
89	0.156	0.156	0.155	0.155	0.155	0.154	0.154	0.153	0.153	0.152	0.152	0.151	0.151
90	0.168	0.167	0.167	0.167	0.166	0.166	0.165	0.165	0.164	0.164	0.164	0.163	0.163
91	0.180	0.180	0.179	0.179	0.178	0.178	0.177	0.177	0.177	0.176	0.176	0.175	0.175
92	0.193	0.193	0.192	0.192	0.191	0.191	0.190	0.190	0.190	0.189	0.189	0.188	0.188
93	0.207	0.206	0.206	0.205	0.205	0.205	0.204	0.204	0.203	0.203	0.203	0.202	0.202
94	0.221	0.220	0.220	0.220	0.219	0.219	0.219	0.218	0.218	0.218	0.217	0.217	0.217
95	0.236	0.236	0.235	0.235	0.235	0.234	0.234	0.234	0.233	0.233	0.233	0.232	0.232
96	0.252	0.251	0.251	0.251	0.251	0.250	0.250	0.250	0.250	0.249	0.249	0.249	0.248
97	0.268	0.268	0.268	0.268	0.267	0.267	0.267	0.267	0.267	0.266	0.266	0.266	0.266
98	0.285	0.285	0.285	0.285	0.285	0.285	0.285	0.284	0.284	0.284	0.284	0.284	0.284
99	0.304	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303	0.303
100	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322	0.322
101	0.342	0.342	0.342	0.342	0.342	0.343	0.343	0.343	0.343	0.343	0.343	0.343	0.343
102	0.363	0.363	0.363	0.363	0.363	0.363	0.364	0.364	0.364	0.364	0.364	0.365	0.365
103	0.384	0.384	0.384	0.385	0.385	0.385	0.386	0.386	0.386	0.387	0.387	0.387	0.387
104	0.406	0.406	0.407	0.407	0.408	0.408	0.408	0.409	0.409	0.410	0.410	0.410	0.411
105	0.429	0.429	0.430	0.430	0.431	0.431	0.432	0.433	0.433	0.434	0.434	0.435	0.435
106	0.452	0.453	0.454	0.454	0.455	0.456	0.457	0.457	0.458	0.459	0.459	0.460	0.461
107	0.477	0.478	0.478	0.479	0.480	0.481	0.482	0.483	0.483	0.484	0.485	0.486	0.487
108	0.502	0.503	0.504	0.505	0.506	0.507	0.508	0.509	0.510	0.511	0.512	0.513	0.513
109	0.527	0.528	0.530	0.531	0.532	0.533	0.534	0.535	0.536	0.537	0.539	0.540	0.541
110	0.553	0.554	0.556	0.557	0.558	0.560	0.561	0.562	0.563	0.565	0.566	0.567	0.568

Table C.10: m_x, African, Males, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.062	0.062	0.063	0.063	0.063	0.063	0.063	0.063	0.064	0.064	0.064	0.064	0.064	0.064
71	0.067	0.067	0.067	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.069	0.069	0.069
72	0.072	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073
73	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078
74	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
75	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088
76	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093
77	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098
78	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103
79	0.107	0.107	0.107	0.107	0.107	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108
80	0.111	0.112	0.112	0.112	0.112	0.112	0.112	0.113	0.113	0.113	0.113	0.113	0.113	0.113
81	0.116	0.116	0.116	0.116	0.117	0.117	0.117	0.117	0.118	0.118	0.118	0.118	0.118	0.119
82	0.120	0.120	0.120	0.121	0.121	0.121	0.122	0.122	0.122	0.123	0.123	0.123	0.124	0.124
83	0.123	0.124	0.124	0.125	0.125	0.126	0.126	0.127	0.127	0.128	0.128	0.129	0.129	0.129
84	0.127	0.128	0.128	0.129	0.130	0.130	0.131	0.132	0.132	0.133	0.133	0.134	0.134	0.135
85	0.131	0.131	0.132	0.133	0.134	0.135	0.136	0.136	0.137	0.138	0.139	0.139	0.140	0.141
86	0.134	0.135	0.136	0.137	0.138	0.139	0.140	0.141	0.142	0.143	0.144	0.145	0.146	0.146
87	0.137	0.139	0.140	0.141	0.142	0.143	0.145	0.146	0.147	0.148	0.149	0.150	0.151	0.152
88	0.140	0.142	0.143	0.145	0.146	0.148	0.149	0.151	0.152	0.153	0.155	0.156	0.157	0.158
89	0.144	0.145	0.147	0.149	0.150	0.152	0.154	0.155	0.157	0.159	0.160	0.162	0.163	0.165
90	0.147	0.149	0.151	0.153	0.155	0.156	0.158	0.160	0.162	0.164	0.166	0.168	0.169	0.171
91	0.149	0.152	0.154	0.156	0.159	0.161	0.163	0.165	0.167	0.170	0.172	0.174	0.176	0.178
92	0.152	0.155	0.158	0.160	0.163	0.165	0.168	0.170	0.173	0.175	0.178	0.180	0.182	0.185
93	0.155	0.158	0.161	0.164	0.167	0.170	0.173	0.175	0.178	0.181	0.184	0.186	0.189	0.192
94	0.158	0.161	0.165	0.168	0.171	0.174	0.178	0.181	0.184	0.187	0.190	0.193	0.196	0.199
95	0.161	0.165	0.168	0.172	0.175	0.179	0.183	0.186	0.190	0.193	0.197	0.200	0.203	0.207
96	0.164	0.168	0.172	0.176	0.180	0.184	0.188	0.192	0.196	0.199	0.203	0.207	0.211	0.215
97	0.167	0.171	0.176	0.180	0.184	0.189	0.193	0.197	0.202	0.206	0.210	0.215	0.219	0.223
98	0.169	0.174	0.179	0.184	0.189	0.194	0.199	0.203	0.208	0.213	0.218	0.222	0.227	0.232
99		0.178	0.183	0.189	0.194	0.199	0.204	0.210	0.215	0.220	0.225	0.230	0.235	0.241
100			0.187	0.193	0.199	0.205	0.210	0.216	0.222	0.227	0.233	0.239	0.244	0.250
101				0.198	0.204	0.210	0.216	0.223	0.229	0.235	0.241	0.247	0.253	0.260
102					0.209	0.216	0.223	0.230	0.236	0.243	0.250	0.256	0.263	0.270
103						0.222	0.229	0.237	0.244	0.251	0.259	0.266	0.273	0.280
104							0.236	0.244	0.252	0.260	0.268	0.276	0.284	0.291
105								0.252	0.261	0.269	0.278	0.286	0.294	0.303
106									0.270	0.279	0.288	0.297	0.306	0.315
107										0.289	0.298	0.308	0.318	0.328
108											0.310	0.320	0.331	0.341
109												0.333	0.344	0.355
110													0.358	0.370

Table C.11: m_x, African, Males, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.066	0.066	0.066	0.066	0.066	0.066
71	0.069	0.069	0.069	0.069	0.069	0.069	0.070	0.070	0.070	0.070	0.070	0.070	0.070
72	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074
73	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078
74	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
75	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.088	0.087	0.087
76	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.093	0.092	0.092	0.092	0.092
77	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.097	0.097	0.097
78	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103
79	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108	0.108
80	0.113	0.113	0.114	0.114	0.114	0.114	0.114	0.114	0.114	0.114	0.114	0.114	0.114
81	0.119	0.119	0.119	0.119	0.119	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120
82	0.124	0.124	0.125	0.125	0.125	0.125	0.126	0.126	0.126	0.126	0.127	0.127	0.127
83	0.130	0.130	0.131	0.131	0.131	0.132	0.132	0.132	0.133	0.133	0.133	0.134	0.134
84	0.135	0.136	0.136	0.137	0.137	0.138	0.138	0.139	0.139	0.140	0.140	0.141	0.141
85	0.141	0.142	0.143	0.143	0.144	0.145	0.145	0.146	0.146	0.147	0.148	0.148	0.149
86	0.147	0.148	0.149	0.150	0.151	0.151	0.152	0.153	0.154	0.154	0.155	0.156	0.157
87	0.153	0.154	0.155	0.156	0.157	0.158	0.159	0.160	0.161	0.162	0.163	0.164	0.165
88	0.160	0.161	0.162	0.163	0.165	0.166	0.167	0.168	0.169	0.170	0.172	0.173	0.174
89	0.166	0.168	0.169	0.171	0.172	0.173	0.175	0.176	0.178	0.179	0.180	0.182	0.183
90	0.173	0.175	0.176	0.178	0.180	0.181	0.183	0.185	0.186	0.188	0.190	0.191	0.193
91	0.180	0.182	0.184	0.186	0.188	0.190	0.192	0.194	0.196	0.197	0.199	0.201	0.203
92	0.187	0.189	0.192	0.194	0.196	0.198	0.201	0.203	0.205	0.207	0.210	0.212	0.214
93	0.194	0.197	0.200	0.202	0.205	0.208	0.210	0.213	0.215	0.218	0.220	0.223	0.225
94	0.202	0.205	0.208	0.211	0.214	0.217	0.220	0.223	0.226	0.229	0.231	0.234	0.237
95	0.210	0.214	0.217	0.220	0.224	0.227	0.230	0.233	0.237	0.240	0.243	0.246	0.250
96	0.219	0.222	0.226	0.230	0.233	0.237	0.241	0.245	0.248	0.252	0.255	0.259	0.263
97	0.227	0.231	0.236	0.240	0.244	0.248	0.252	0.256	0.260	0.264	0.268	0.272	0.276
98	0.236	0.241	0.245	0.250	0.255	0.259	0.264	0.268	0.273	0.277	0.282	0.286	0.291
99	0.246	0.251	0.256	0.261	0.266	0.271	0.276	0.281	0.286	0.291	0.296	0.301	0.306
100	0.255	0.261	0.267	0.272	0.278	0.283	0.289	0.294	0.300	0.305	0.311	0.316	0.322
101	0.266	0.272	0.278	0.284	0.290	0.296	0.302	0.308	0.314	0.320	0.326	0.332	0.338
102	0.276	0.283	0.290	0.296	0.303	0.310	0.316	0.323	0.329	0.336	0.343	0.349	0.356
103	0.288	0.295	0.302	0.309	0.316	0.324	0.331	0.338	0.345	0.352	0.360	0.367	0.374
104	0.299	0.307	0.315	0.323	0.331	0.338	0.346	0.354	0.362	0.370	0.377	0.385	0.393
105	0.311	0.320	0.328	0.337	0.345	0.354	0.362	0.371	0.379	0.388	0.396	0.404	0.413
106	0.324	0.333	0.342	0.351	0.361	0.370	0.379	0.388	0.397	0.406	0.415	0.425	0.434
107	0.337	0.347	0.357	0.367	0.377	0.387	0.396	0.406	0.416	0.426	0.436	0.446	0.455
108	0.351	0.362	0.372	0.383	0.393	0.404	0.415	0.425	0.436	0.446	0.457	0.467	0.478
109	0.366	0.377	0.389	0.400	0.411	0.422	0.434	0.445	0.456	0.467	0.479	0.490	0.501
110	0.382	0.394	0.405	0.417	0.429	0.441	0.453	0.465	0.477	0.490	0.502	0.514	0.526

Table C.12: m_x, African, Females, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.038	0.038	0.038	0.038	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.040	0.040	0.040
71	0.041	0.041	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.043	0.043	0.043	0.043
72	0.045	0.045	0.045	0.045	0.045	0.045	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046
73	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.050	0.050	0.050	0.050
74	0.052	0.052	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
75	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057
76	0.060	0.060	0.060	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061
77	0.064	0.064	0.064	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.065	0.066	0.066
78	0.068	0.068	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.070	0.070	0.070	0.070	0.070
79	0.072	0.073	0.073	0.073	0.073	0.073	0.074	0.074	0.074	0.074	0.074	0.074	0.075	0.075
80	0.077	0.077	0.077	0.077	0.078	0.078	0.078	0.078	0.079	0.079	0.079	0.079	0.079	0.080
81	0.081	0.081	0.082	0.082	0.082	0.082	0.083	0.083	0.083	0.084	0.084	0.084	0.084	0.085
82	0.085	0.086	0.086	0.086	0.087	0.087	0.088	0.088	0.088	0.089	0.089	0.089	0.090	0.090
83	0.090	0.090	0.091	0.091	0.092	0.092	0.093	0.093	0.093	0.094	0.094	0.095	0.095	0.096
84	0.094	0.095	0.095	0.096	0.096	0.097	0.098	0.098	0.099	0.099	0.100	0.100	0.101	0.102
85	0.098	0.099	0.100	0.101	0.101	0.102	0.103	0.104	0.104	0.105	0.106	0.106	0.107	0.108
86	0.103	0.104	0.105	0.106	0.107	0.108	0.108	0.109	0.110	0.111	0.112	0.112	0.113	0.114
87	0.108	0.109	0.110	0.111	0.112	0.113	0.114	0.115	0.116	0.117	0.118	0.119	0.120	0.121
88	0.113	0.114	0.115	0.116	0.117	0.119	0.120	0.121	0.122	0.123	0.124	0.126	0.127	0.128
89	0.117	0.119	0.120	0.122	0.123	0.124	0.126	0.127	0.129	0.130	0.131	0.133	0.134	0.135
90	0.122	0.124	0.126	0.127	0.129	0.131	0.132	0.134	0.135	0.137	0.138	0.140	0.141	0.143
91	0.128	0.129	0.131	0.133	0.135	0.137	0.139	0.140	0.142	0.144	0.146	0.147	0.149	0.151
92	0.133	0.135	0.137	0.139	0.141	0.143	0.145	0.147	0.149	0.151	0.153	0.155	0.157	0.159
93	0.138	0.141	0.143	0.145	0.148	0.150	0.152	0.155	0.157	0.159	0.161	0.164	0.166	0.168
94	0.144	0.146	0.149	0.152	0.154	0.157	0.159	0.162	0.165	0.167	0.170	0.172	0.175	0.177
95	0.150	0.152	0.155	0.158	0.161	0.164	0.167	0.170	0.173	0.176	0.178	0.181	0.184	0.187
96	0.155	0.159	0.162	0.165	0.168	0.172	0.175	0.178	0.181	0.184	0.188	0.191	0.194	0.197
97	0.162	0.165	0.169	0.172	0.176	0.179	0.183	0.187	0.190	0.194	0.197	0.201	0.204	0.207
98	0.168	0.172	0.176	0.180	0.184	0.188	0.191	0.195	0.199	0.203	0.207	0.211	0.215	0.218
99		0.179	0.183	0.187	0.192	0.196	0.200	0.205	0.209	0.213	0.217	0.222	0.226	0.230
100			0.191	0.195	0.200	0.205	0.209	0.214	0.219	0.223	0.228	0.233	0.237	0.242
101				0.204	0.209	0.214	0.219	0.224	0.229	0.234	0.239	0.245	0.250	0.255
102					0.218	0.224	0.229	0.235	0.240	0.246	0.251	0.257	0.262	0.268
103						0.233	0.239	0.245	0.252	0.258	0.264	0.270	0.276	0.282
104							0.250	0.257	0.263	0.270	0.276	0.283	0.289	0.296
105								0.269	0.276	0.283	0.290	0.297	0.304	0.311
106									0.289	0.296	0.304	0.311	0.319	0.327
107										0.310	0.318	0.327	0.335	0.343
108											0.334	0.342	0.351	0.360
109												0.359	0.368	0.378
110													0.386	0.396

Table C.13: m_x, African, Females, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.040	0.040	0.040	0.040	0.040	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.042
71	0.043	0.043	0.043	0.043	0.043	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
72	0.046	0.046	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047
73	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.051	0.051	0.051
74	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054
75	0.057	0.057	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058
76	0.061	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
77	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.067
78	0.070	0.070	0.070	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071
79	0.075	0.075	0.075	0.075	0.075	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
80	0.080	0.080	0.080	0.080	0.081	0.081	0.081	0.081	0.081	0.081	0.082	0.082	0.082
81	0.085	0.085	0.085	0.086	0.086	0.086	0.086	0.087	0.087	0.087	0.087	0.088	0.088
82	0.090	0.091	0.091	0.091	0.092	0.092	0.092	0.093	0.093	0.093	0.093	0.094	0.094
83	0.096	0.096	0.097	0.097	0.098	0.098	0.098	0.099	0.099	0.100	0.100	0.100	0.101
84	0.102	0.103	0.103	0.104	0.104	0.105	0.105	0.106	0.106	0.106	0.107	0.107	0.108
85	0.108	0.109	0.110	0.110	0.111	0.111	0.112	0.113	0.113	0.114	0.114	0.115	0.115
86	0.115	0.116	0.116	0.117	0.118	0.119	0.119	0.120	0.121	0.121	0.122	0.123	0.124
87	0.122	0.123	0.123	0.124	0.125	0.126	0.127	0.128	0.129	0.130	0.130	0.131	0.132
88	0.129	0.130	0.131	0.132	0.133	0.134	0.135	0.136	0.137	0.138	0.139	0.140	0.141
89	0.136	0.138	0.139	0.140	0.141	0.143	0.144	0.145	0.146	0.147	0.149	0.150	0.151
90	0.144	0.146	0.147	0.149	0.150	0.151	0.153	0.154	0.156	0.157	0.159	0.160	0.161
91	0.152	0.154	0.156	0.158	0.159	0.161	0.162	0.164	0.166	0.167	0.169	0.171	0.172
92	0.161	0.163	0.165	0.167	0.169	0.171	0.173	0.174	0.176	0.178	0.180	0.182	0.184
93	0.170	0.172	0.175	0.177	0.179	0.181	0.183	0.185	0.187	0.190	0.192	0.194	0.196
94	0.180	0.182	0.185	0.187	0.190	0.192	0.194	0.197	0.199	0.202	0.204	0.206	0.209
95	0.190	0.192	0.195	0.198	0.201	0.203	0.206	0.209	0.212	0.214	0.217	0.220	0.222
96	0.200	0.203	0.206	0.209	0.212	0.215	0.218	0.222	0.225	0.228	0.231	0.234	0.237
97	0.211	0.214	0.218	0.221	0.225	0.228	0.231	0.235	0.238	0.242	0.245	0.248	0.252
98	0.222	0.226	0.230	0.234	0.237	0.241	0.245	0.249	0.253	0.256	0.260	0.264	0.268
99	0.234	0.238	0.243	0.247	0.251	0.255	0.259	0.263	0.268	0.272	0.276	0.280	0.284
100	0.247	0.251	0.256	0.260	0.265	0.270	0.274	0.279	0.283	0.288	0.293	0.297	0.302
101	0.260	0.265	0.270	0.275	0.280	0.285	0.290	0.295	0.300	0.305	0.310	0.315	0.320
102	0.273	0.279	0.284	0.290	0.295	0.301	0.306	0.312	0.317	0.323	0.328	0.334	0.339
103	0.288	0.294	0.300	0.306	0.312	0.318	0.324	0.329	0.335	0.341	0.347	0.353	0.359
104	0.302	0.309	0.315	0.322	0.328	0.335	0.341	0.348	0.354	0.361	0.367	0.374	0.380
105	0.318	0.325	0.332	0.339	0.346	0.353	0.360	0.367	0.374	0.381	0.388	0.395	0.402
106	0.334	0.342	0.349	0.357	0.365	0.372	0.380	0.387	0.395	0.403	0.410	0.418	0.425
107	0.351	0.359	0.367	0.376	0.384	0.392	0.400	0.408	0.417	0.425	0.433	0.441	0.449
108	0.369	0.377	0.386	0.395	0.404	0.413	0.421	0.430	0.439	0.448	0.457	0.465	0.474
109	0.387	0.396	0.406	0.415	0.425	0.434	0.443	0.453	0.462	0.472	0.481	0.491	0.500
110	0.406	0.416	0.426	0.436	0.446	0.456	0.466	0.476	0.486	0.496	0.507	0.517	0.527

Table C.14: m_x, Coloureds, Males, 1996-2008

Age	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
70	0.086	0.084	0.082	0.079	0.077	0.075	0.073	0.071	0.068	0.066	0.064	0.062	0.060
71	0.093	0.091	0.089	0.086	0.084	0.081	0.079	0.076	0.074	0.072	0.069	0.067	0.065
72	0.101	0.099	0.096	0.093	0.091	0.088	0.085	0.083	0.080	0.078	0.075	0.073	0.070
73	0.109	0.106	0.103	0.101	0.098	0.095	0.092	0.089	0.087	0.084	0.081	0.079	0.076
74	0.117	0.114	0.111	0.108	0.105	0.102	0.099	0.096	0.093	0.090	0.088	0.085	0.082
75	0.126	0.122	0.119	0.116	0.113	0.110	0.106	0.103	0.100	0.097	0.094	0.091	0.089
76	0.134	0.131	0.127	0.124	0.120	0.117	0.114	0.111	0.108	0.104	0.101	0.098	0.095
77	0.142	0.139	0.135	0.132	0.128	0.125	0.122	0.118	0.115	0.112	0.109	0.106	0.102
78	0.151	0.147	0.144	0.140	0.137	0.133	0.130	0.126	0.123	0.120	0.116	0.113	0.110
79	0.160	0.156	0.152	0.149	0.145	0.142	0.138	0.135	0.131	0.128	0.124	0.121	0.118
80	0.169	0.165	0.161	0.157	0.154	0.150	0.147	0.143	0.140	0.136	0.133	0.129	0.126
81	0.178	0.174	0.170	0.166	0.163	0.159	0.156	0.152	0.148	0.145	0.142	0.138	0.135
82	0.187	0.183	0.179	0.176	0.172	0.168	0.165	0.161	0.158	0.154	0.151	0.147	0.144
83	0.196	0.192	0.189	0.185	0.181	0.178	0.174	0.171	0.167	0.164	0.160	0.157	0.154
84	0.205	0.202	0.198	0.195	0.191	0.187	0.184	0.181	0.177	0.174	0.170	0.167	0.164
85	0.215	0.212	0.208	0.204	0.201	0.198	0.194	0.191	0.187	0.184	0.181	0.177	0.174
86	0.225	0.222	0.218	0.215	0.211	0.208	0.205	0.201	0.198	0.195	0.192	0.189	0.185
87	0.235	0.232	0.228	0.225	0.222	0.219	0.216	0.212	0.209	0.206	0.203	0.200	0.197
88	0.245	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221	0.218	0.215	0.212	0.209
89	0.256	0.253	0.250	0.247	0.244	0.242	0.239	0.236	0.233	0.230	0.228	0.225	0.222
90	0.267	0.264	0.261	0.259	0.256	0.254	0.251	0.248	0.246	0.243	0.241	0.238	0.236
91	0.278	0.276	0.273	0.271	0.268	0.266	0.264	0.261	0.259	0.257	0.255	0.252	0.250
92	0.290	0.287	0.285	0.283	0.281	0.279	0.277	0.275	0.273	0.271	0.269	0.267	0.265
93	0.301	0.300	0.298	0.296	0.294	0.293	0.291	0.289	0.287	0.286	0.284	0.282	0.281
94	0.314	0.312	0.311	0.309	0.308	0.307	0.305	0.304	0.302	0.301	0.300	0.298	0.297
95	0.326	0.325	0.324	0.323	0.322	0.321	0.320	0.319	0.318	0.317	0.316	0.315	0.314
96	0.339	0.339	0.338	0.338	0.337	0.336	0.336	0.335	0.335	0.334	0.333	0.333	0.332
97	0.353	0.353	0.353	0.353	0.352	0.352	0.352	0.352	0.352	0.351	0.351	0.351	0.351
98	0.367	0.367	0.368	0.368	0.368	0.369	0.369	0.369	0.370	0.370	0.370	0.370	0.371
99	0.382	0.383	0.383	0.384	0.385	0.386	0.387	0.387	0.388	0.389	0.390	0.391	0.391
100	0.397	0.398	0.400	0.401	0.402	0.404	0.405	0.406	0.408	0.409	0.410	0.412	0.413
101	0.413	0.415	0.417	0.419	0.420	0.422	0.424	0.426	0.428	0.430	0.432	0.434	0.436
102	0.429	0.432	0.434	0.437	0.439	0.442	0.444	0.447	0.449	0.452	0.454	0.457	0.459
103	0.446	0.449	0.452	0.456	0.459	0.462	0.465	0.468	0.471	0.475	0.478	0.481	0.484
104	0.464	0.468	0.472	0.475	0.479	0.483	0.487	0.491	0.495	0.498	0.502	0.506	0.510
105	0.482	0.487	0.491	0.496	0.500	0.505	0.510	0.514	0.519	0.523	0.528	0.532	0.537
106	0.501	0.507	0.512	0.517	0.523	0.528	0.533	0.538	0.544	0.549	0.554	0.559	0.564
107	0.521	0.527	0.533	0.539	0.545	0.551	0.557	0.564	0.570	0.576	0.582	0.588	0.594
108	0.542	0.549	0.555	0.562	0.569	0.576	0.583	0.590	0.596	0.603	0.610	0.617	0.623
109	0.563	0.571	0.578	0.586	0.594	0.601	0.609	0.616	0.624	0.632	0.639	0.647	0.654
110		0.594	0.602	0.611	0.619	0.627	0.636	0.644	0.653	0.661	0.669	0.678	0.686

Table C.15: m_x, Coloureds, Males, 2009-2011

Age	2009	2010	2011
70	0.058	0.056	0.054
71	0.063	0.060	0.058
72	0.068	0.065	0.063
73	0.073	0.071	0.068
74	0.079	0.077	0.074
75	0.086	0.083	0.080
76	0.092	0.089	0.086
77	0.099	0.096	0.093
78	0.107	0.104	0.100
79	0.115	0.111	0.108
80	0.123	0.120	0.116
81	0.131	0.128	0.125
82	0.141	0.137	0.134
83	0.150	0.147	0.144
84	0.160	0.157	0.154
85	0.171	0.168	0.165
86	0.182	0.179	0.176
87	0.194	0.191	0.188
88	0.207	0.204	0.201
89	0.220	0.217	0.214
90	0.233	0.231	0.228
91	0.248	0.246	0.243
92	0.263	0.261	0.259
93	0.279	0.277	0.275
94	0.296	0.294	0.293
95	0.313	0.312	0.311
96	0.331	0.331	0.330
97	0.351	0.350	0.350
98	0.371	0.371	0.371
99	0.392	0.393	0.394
100	0.414	0.415	0.417
101	0.437	0.439	0.441
102	0.462	0.464	0.466
103	0.487	0.490	0.493
104	0.513	0.517	0.521
105	0.541	0.545	0.550
106	0.570	0.575	0.580
107	0.599	0.605	0.611
108	0.630	0.637	0.644
109	0.662	0.669	0.677
110	0.694	0.703	0.711

Table C.16: m_x, Coloureds, Females, 1996-2008

Age	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
70	0.063	0.060	0.058	0.056	0.053	0.051	0.049	0.047	0.045	0.043	0.042	0.040	0.038
71	0.066	0.063	0.061	0.059	0.057	0.055	0.053	0.051	0.049	0.047	0.045	0.043	0.042
72	0.069	0.067	0.065	0.062	0.060	0.058	0.056	0.054	0.052	0.051	0.049	0.047	0.045
73	0.073	0.071	0.068	0.066	0.064	0.062	0.060	0.058	0.056	0.055	0.053	0.051	0.049
74	0.077	0.075	0.072	0.070	0.068	0.066	0.065	0.063	0.061	0.059	0.057	0.056	0.054
75	0.081	0.079	0.077	0.075	0.073	0.071	0.069	0.067	0.065	0.064	0.062	0.060	0.059
76	0.085	0.083	0.081	0.080	0.078	0.076	0.074	0.072	0.071	0.069	0.067	0.065	0.064
77	0.090	0.088	0.087	0.085	0.083	0.081	0.079	0.078	0.076	0.074	0.073	0.071	0.069
78	0.096	0.094	0.092	0.090	0.088	0.087	0.085	0.083	0.082	0.080	0.079	0.077	0.075
79	0.101	0.099	0.098	0.096	0.094	0.093	0.091	0.090	0.088	0.086	0.085	0.083	0.082
80	0.107	0.105	0.104	0.102	0.101	0.099	0.098	0.096	0.095	0.093	0.092	0.090	0.089
81	0.114	0.112	0.110	0.109	0.108	0.106	0.105	0.103	0.102	0.100	0.099	0.098	0.096
82	0.120	0.119	0.118	0.116	0.115	0.113	0.112	0.111	0.109	0.108	0.107	0.106	0.104
83	0.128	0.126	0.125	0.124	0.122	0.121	0.120	0.119	0.118	0.116	0.115	0.114	0.113
84	0.135	0.134	0.133	0.132	0.131	0.129	0.128	0.127	0.126	0.125	0.124	0.123	0.122
85	0.143	0.142	0.141	0.140	0.139	0.138	0.137	0.136	0.135	0.134	0.133	0.132	0.131
86	0.152	0.151	0.150	0.149	0.148	0.148	0.147	0.146	0.145	0.144	0.143	0.142	0.142
87	0.161	0.160	0.160	0.159	0.158	0.157	0.157	0.156	0.155	0.155	0.154	0.153	0.152
88	0.171	0.170	0.170	0.169	0.168	0.168	0.167	0.167	0.166	0.166	0.165	0.164	0.164
89	0.181	0.181	0.180	0.180	0.179	0.179	0.178	0.178	0.178	0.177	0.177	0.176	0.176
90	0.192	0.191	0.191	0.191	0.191	0.190	0.190	0.190	0.190	0.189	0.189	0.189	0.189
91	0.203	0.203	0.203	0.203	0.203	0.203	0.203	0.202	0.202	0.202	0.202	0.202	0.202
92	0.215	0.215	0.215	0.215	0.215	0.215	0.216	0.216	0.216	0.216	0.216	0.216	0.216
93	0.227	0.228	0.228	0.228	0.229	0.229	0.229	0.229	0.230	0.230	0.230	0.231	0.231
94	0.240	0.241	0.241	0.242	0.242	0.243	0.243	0.244	0.244	0.245	0.245	0.246	0.246
95	0.254	0.255	0.256	0.256	0.257	0.258	0.258	0.259	0.260	0.261	0.261	0.262	0.263
96	0.268	0.269	0.270	0.271	0.272	0.273	0.274	0.275	0.276	0.277	0.278	0.279	0.280
97	0.283	0.284	0.286	0.287	0.288	0.289	0.290	0.292	0.293	0.294	0.295	0.296	0.297
98	0.299	0.300	0.302	0.303	0.305	0.306	0.307	0.309	0.310	0.312	0.313	0.314	0.316
99	0.315	0.317	0.319	0.320	0.322	0.324	0.325	0.327	0.329	0.330	0.332	0.333	0.335
100	0.332	0.334	0.336	0.338	0.340	0.342	0.344	0.346	0.348	0.349	0.351	0.353	0.355
101	0.350	0.352	0.354	0.356	0.359	0.361	0.363	0.365	0.367	0.369	0.372	0.374	0.376
102	0.368	0.371	0.373	0.375	0.378	0.380	0.383	0.385	0.388	0.390	0.393	0.395	0.398
103	0.387	0.390	0.393	0.395	0.398	0.401	0.404	0.406	0.409	0.412	0.415	0.417	0.420
104	0.407	0.410	0.413	0.416	0.419	0.422	0.425	0.428	0.431	0.434	0.437	0.440	0.443
105	0.427	0.431	0.434	0.437	0.441	0.444	0.447	0.451	0.454	0.457	0.461	0.464	0.467
106	0.448	0.452	0.456	0.459	0.463	0.467	0.470	0.474	0.478	0.481	0.485	0.489	0.492
107	0.470	0.474	0.478	0.482	0.486	0.490	0.494	0.498	0.502	0.506	0.510	0.514	0.518
108	0.493	0.497	0.501	0.505	0.510	0.514	0.518	0.523	0.527	0.531	0.536	0.540	0.544
109	0.515	0.520	0.525	0.529	0.534	0.539	0.543	0.548	0.553	0.557	0.562	0.566	0.571
110		0.544	0.549	0.554	0.559	0.564	0.569	0.574	0.579	0.584	0.588	0.593	0.598

Table C.17: m_x, Coloureds, Females, 2009-2011

Age	2009	2010	2011
70	0.037	0.035	0.033
71	0.040	0.038	0.037
72	0.044	0.042	0.041
73	0.048	0.046	0.045
74	0.052	0.051	0.049
75	0.057	0.055	0.054
76	0.062	0.061	0.059
77	0.068	0.066	0.065
78	0.074	0.072	0.071
79	0.080	0.079	0.078
80	0.087	0.086	0.085
81	0.095	0.094	0.092
82	0.103	0.102	0.100
83	0.112	0.110	0.109
84	0.121	0.120	0.119
85	0.130	0.129	0.128
86	0.141	0.140	0.139
87	0.152	0.151	0.150
88	0.163	0.163	0.162
89	0.175	0.175	0.175
90	0.188	0.188	0.188
91	0.202	0.202	0.202
92	0.216	0.216	0.216
93	0.231	0.231	0.232
94	0.247	0.247	0.248
95	0.263	0.264	0.265
96	0.281	0.281	0.282
97	0.298	0.300	0.301
98	0.317	0.319	0.320
99	0.337	0.338	0.340
100	0.357	0.359	0.361
101	0.378	0.380	0.383
102	0.400	0.403	0.405
103	0.423	0.426	0.428
104	0.446	0.449	0.452
105	0.471	0.474	0.477
106	0.496	0.500	0.503
107	0.522	0.526	0.530
108	0.549	0.553	0.557
109	0.576	0.580	0.585
110	0.603	0.608	0.613

Table C.18: m_x, Indians, Males, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.074	0.073	0.072	0.071	0.070	0.069	0.068	0.067	0.066	0.065	0.064	0.063	0.062	0.061
71	0.082	0.081	0.080	0.079	0.077	0.076	0.075	0.074	0.073	0.071	0.070	0.069	0.068	0.067
72	0.091	0.090	0.088	0.087	0.085	0.084	0.082	0.081	0.079	0.078	0.077	0.075	0.074	0.072
73	0.100	0.099	0.097	0.095	0.093	0.092	0.090	0.088	0.086	0.085	0.083	0.081	0.080	0.078
74	0.109	0.107	0.105	0.103	0.101	0.099	0.097	0.095	0.093	0.091	0.090	0.088	0.086	0.084
75	0.117	0.115	0.113	0.111	0.108	0.106	0.104	0.102	0.100	0.098	0.096	0.094	0.092	0.090
76	0.125	0.123	0.120	0.118	0.116	0.114	0.111	0.109	0.107	0.105	0.103	0.101	0.099	0.097
77	0.132	0.130	0.127	0.125	0.123	0.121	0.118	0.116	0.114	0.112	0.110	0.108	0.106	0.104
78	0.139	0.137	0.134	0.132	0.130	0.128	0.125	0.123	0.121	0.119	0.117	0.115	0.113	0.111
79	0.145	0.143	0.141	0.139	0.136	0.134	0.132	0.130	0.128	0.126	0.124	0.122	0.120	0.118
80	0.151	0.149	0.147	0.145	0.143	0.141	0.139	0.137	0.135	0.133	0.131	0.129	0.128	0.126
81	0.156	0.154	0.153	0.151	0.149	0.147	0.146	0.144	0.142	0.141	0.139	0.137	0.135	0.134
82	0.161	0.160	0.158	0.157	0.155	0.154	0.152	0.151	0.150	0.148	0.147	0.145	0.144	0.142
83	0.165	0.164	0.163	0.162	0.161	0.160	0.159	0.158	0.157	0.156	0.155	0.154	0.152	0.151
84	0.170	0.169	0.169	0.168	0.167	0.167	0.166	0.165	0.165	0.164	0.163	0.162	0.161	0.161
85	0.173	0.173	0.173	0.173	0.173	0.173	0.173	0.173	0.172	0.172	0.172	0.171	0.171	0.171
86	0.177	0.178	0.178	0.179	0.179	0.180	0.180	0.180	0.180	0.181	0.181	0.181	0.181	0.181
87	0.181	0.182	0.183	0.184	0.185	0.186	0.187	0.188	0.189	0.190	0.190	0.191	0.192	0.193
88	0.184	0.186	0.188	0.190	0.191	0.193	0.195	0.196	0.198	0.199	0.201	0.202	0.203	0.205
89	0.187	0.190	0.192	0.195	0.198	0.200	0.202	0.205	0.207	0.209	0.211	0.213	0.215	0.217
90	0.190	0.194	0.197	0.201	0.204	0.207	0.211	0.214	0.217	0.220	0.223	0.225	0.228	0.231
91	0.194	0.198	0.202	0.207	0.211	0.215	0.219	0.223	0.227	0.231	0.235	0.238	0.242	0.246
92	0.197	0.202	0.208	0.213	0.218	0.223	0.228	0.233	0.238	0.243	0.247	0.252	0.257	0.261
93	0.200	0.207	0.213	0.219	0.226	0.232	0.238	0.244	0.249	0.255	0.261	0.267	0.272	0.278
94	0.204	0.212	0.219	0.226	0.234	0.241	0.248	0.255	0.262	0.269	0.276	0.282	0.289	0.296
95	0.208	0.217	0.225	0.234	0.242	0.250	0.259	0.267	0.275	0.283	0.291	0.299	0.307	0.315
96	0.213	0.222	0.232	0.242	0.251	0.261	0.270	0.280	0.289	0.298	0.308	0.317	0.326	0.335
97	0.218	0.228	0.239	0.250	0.261	0.272	0.283	0.293	0.304	0.315	0.325	0.336	0.346	0.357
98	0.223	0.235	0.247	0.259	0.272	0.284	0.296	0.308	0.320	0.332	0.344	0.356	0.368	0.380
99		0.242	0.256	0.269	0.283	0.297	0.310	0.324	0.337	0.351	0.364	0.378	0.391	0.405
100			0.266	0.280	0.295	0.310	0.325	0.340	0.356	0.371	0.386	0.401	0.416	0.431
101				0.293	0.309	0.326	0.342	0.359	0.375	0.392	0.409	0.425	0.442	0.459
102					0.324	0.342	0.360	0.378	0.397	0.415	0.433	0.452	0.470	0.489
103						0.360	0.380	0.399	0.419	0.439	0.459	0.480	0.500	0.520
104							0.401	0.422	0.444	0.465	0.487	0.509	0.531	0.553
105								0.446	0.470	0.493	0.517	0.540	0.564	0.588
106									0.497	0.523	0.548	0.573	0.599	0.625
107										0.554	0.581	0.608	0.636	0.663
108											0.616	0.645	0.674	0.704
109												0.683	0.715	0.746
110													0.757	0.790

Table C.19: m_x, Indians, Males, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.060	0.059	0.058	0.057	0.056	0.055	0.054	0.054	0.053	0.052	0.051	0.050	0.049
71	0.065	0.064	0.063	0.062	0.061	0.060	0.058	0.057	0.056	0.055	0.054	0.053	0.052
72	0.071	0.069	0.068	0.067	0.065	0.064	0.063	0.061	0.060	0.059	0.057	0.056	0.055
73	0.076	0.075	0.073	0.072	0.070	0.069	0.067	0.066	0.064	0.063	0.061	0.060	0.058
74	0.082	0.081	0.079	0.077	0.075	0.074	0.072	0.070	0.069	0.067	0.066	0.064	0.063
75	0.089	0.087	0.085	0.083	0.081	0.079	0.078	0.076	0.074	0.072	0.071	0.069	0.068
76	0.095	0.093	0.091	0.089	0.087	0.085	0.084	0.082	0.080	0.078	0.076	0.075	0.073
77	0.102	0.100	0.098	0.096	0.094	0.092	0.090	0.088	0.086	0.085	0.083	0.081	0.079
78	0.109	0.107	0.105	0.103	0.101	0.099	0.097	0.095	0.093	0.092	0.090	0.088	0.086
79	0.116	0.114	0.112	0.110	0.109	0.107	0.105	0.103	0.101	0.099	0.098	0.096	0.094
80	0.124	0.122	0.120	0.119	0.117	0.115	0.113	0.112	0.110	0.108	0.106	0.105	0.103
81	0.132	0.131	0.129	0.127	0.126	0.124	0.122	0.121	0.119	0.118	0.116	0.115	0.113
82	0.141	0.139	0.138	0.137	0.135	0.134	0.132	0.131	0.130	0.128	0.127	0.125	0.124
83	0.150	0.149	0.148	0.147	0.145	0.144	0.143	0.142	0.141	0.140	0.139	0.137	0.136
84	0.160	0.159	0.158	0.157	0.157	0.156	0.155	0.154	0.153	0.152	0.151	0.151	0.150
85	0.170	0.170	0.169	0.169	0.169	0.168	0.168	0.167	0.167	0.166	0.166	0.165	0.165
86	0.181	0.181	0.182	0.182	0.182	0.182	0.181	0.181	0.181	0.181	0.181	0.181	0.181
87	0.193	0.194	0.194	0.195	0.195	0.196	0.196	0.197	0.197	0.198	0.198	0.199	0.199
88	0.206	0.207	0.208	0.209	0.211	0.212	0.213	0.214	0.215	0.216	0.217	0.218	0.219
89	0.219	0.221	0.223	0.225	0.227	0.229	0.230	0.232	0.234	0.235	0.237	0.238	0.240
90	0.234	0.236	0.239	0.242	0.244	0.247	0.249	0.252	0.254	0.256	0.259	0.261	0.263
91	0.249	0.253	0.256	0.260	0.263	0.266	0.270	0.273	0.276	0.279	0.283	0.286	0.289
92	0.266	0.270	0.275	0.279	0.283	0.287	0.292	0.296	0.300	0.304	0.308	0.312	0.316
93	0.283	0.289	0.294	0.300	0.305	0.310	0.316	0.321	0.326	0.331	0.336	0.341	0.346
94	0.302	0.309	0.315	0.322	0.328	0.335	0.341	0.347	0.354	0.360	0.366	0.372	0.379
95	0.322	0.330	0.338	0.346	0.353	0.361	0.369	0.376	0.384	0.391	0.399	0.406	0.413
96	0.344	0.353	0.362	0.371	0.380	0.389	0.398	0.407	0.416	0.424	0.433	0.442	0.451
97	0.367	0.378	0.388	0.398	0.409	0.419	0.429	0.440	0.450	0.460	0.471	0.481	0.491
98	0.392	0.404	0.416	0.428	0.439	0.451	0.463	0.475	0.487	0.499	0.510	0.522	0.534
99	0.418	0.432	0.445	0.459	0.472	0.486	0.499	0.512	0.526	0.539	0.553	0.566	0.579
100	0.446	0.461	0.476	0.492	0.507	0.522	0.537	0.552	0.567	0.582	0.598	0.613	0.628
101	0.476	0.493	0.510	0.526	0.543	0.560	0.577	0.594	0.611	0.628	0.645	0.662	0.679
102	0.507	0.526	0.544	0.563	0.582	0.600	0.619	0.638	0.656	0.675	0.694	0.713	0.732
103	0.540	0.561	0.581	0.601	0.622	0.642	0.663	0.683	0.704	0.725	0.745	0.766	0.787
104	0.575	0.597	0.620	0.642	0.664	0.686	0.709	0.731	0.754	0.776	0.799	0.821	0.844
105	0.612	0.636	0.660	0.684	0.708	0.732	0.757	0.781	0.805	0.830	0.854	0.878	0.903
106	0.650	0.676	0.702	0.728	0.754	0.780	0.806	0.832	0.859	0.885	0.911	0.938	0.964
107	0.691	0.718	0.746	0.774	0.802	0.830	0.858	0.886	0.914	0.943	0.971	0.999	1.027
108	0.733	0.763	0.792	0.822	0.852	0.882	0.912	0.942	0.972	1.002	1.032	1.063	1.093
109	0.777	0.809	0.840	0.872	0.904	0.936	0.968	1.000	1.032	1.064	1.096	1.128	1.160
110	0.823	0.857	0.890	0.924	0.958	0.991	1.025	1.059	1.093	1.127	1.162	1.196	1.230

Table C.20: m_x, Indians, Females, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.052	0.051	0.051	0.050	0.049	0.049	0.048	0.047	0.046	0.046	0.045	0.044	0.044	0.043
71	0.057	0.056	0.055	0.054	0.054	0.053	0.052	0.051	0.050	0.050	0.049	0.048	0.047	0.047
72	0.062	0.061	0.060	0.059	0.058	0.057	0.057	0.056	0.055	0.054	0.053	0.052	0.052	0.051
73	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.061	0.060	0.059	0.058	0.057	0.056	0.055
74	0.072	0.071	0.070	0.069	0.068	0.068	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.060
75	0.078	0.077	0.076	0.075	0.074	0.073	0.072	0.071	0.070	0.069	0.068	0.067	0.066	0.065
76	0.083	0.082	0.081	0.080	0.080	0.079	0.078	0.077	0.076	0.075	0.074	0.073	0.072	0.071
77	0.089	0.088	0.087	0.086	0.085	0.084	0.083	0.083	0.082	0.081	0.080	0.079	0.078	0.077
78	0.095	0.094	0.093	0.092	0.091	0.091	0.090	0.089	0.088	0.087	0.086	0.085	0.084	0.083
79	0.101	0.100	0.099	0.099	0.098	0.097	0.096	0.095	0.094	0.094	0.093	0.092	0.091	0.090
80	0.107	0.107	0.106	0.105	0.104	0.104	0.103	0.102	0.101	0.101	0.100	0.099	0.099	0.098
81	0.114	0.113	0.112	0.112	0.111	0.111	0.110	0.109	0.109	0.108	0.108	0.107	0.106	0.106
82	0.120	0.120	0.119	0.119	0.118	0.118	0.118	0.117	0.117	0.116	0.116	0.115	0.115	0.114
83	0.127	0.127	0.126	0.126	0.126	0.126	0.125	0.125	0.125	0.125	0.124	0.124	0.124	0.124
84	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.133	0.133	0.133
85	0.141	0.141	0.142	0.142	0.142	0.142	0.143	0.143	0.143	0.143	0.143	0.143	0.144	0.144
86	0.148	0.149	0.150	0.150	0.151	0.151	0.152	0.152	0.153	0.153	0.154	0.154	0.154	0.155
87	0.156	0.157	0.158	0.159	0.160	0.161	0.161	0.162	0.163	0.164	0.165	0.165	0.166	0.167
88	0.164	0.165	0.167	0.168	0.169	0.170	0.172	0.173	0.174	0.175	0.176	0.177	0.178	0.179
89	0.172	0.174	0.176	0.178	0.179	0.181	0.182	0.184	0.185	0.187	0.189	0.190	0.191	0.193
90	0.181	0.183	0.185	0.188	0.190	0.192	0.194	0.196	0.198	0.200	0.202	0.204	0.205	0.207
91	0.190	0.193	0.196	0.198	0.201	0.203	0.206	0.208	0.211	0.213	0.215	0.218	0.220	0.223
92	0.200	0.203	0.206	0.209	0.212	0.215	0.218	0.221	0.224	0.227	0.230	0.233	0.236	0.239
93	0.210	0.213	0.217	0.221	0.224	0.228	0.232	0.235	0.239	0.242	0.246	0.249	0.253	0.256
94	0.220	0.224	0.229	0.233	0.237	0.241	0.246	0.250	0.254	0.258	0.262	0.266	0.270	0.274
95	0.231	0.236	0.241	0.246	0.251	0.256	0.260	0.265	0.270	0.275	0.279	0.284	0.289	0.294
96	0.243	0.248	0.254	0.259	0.265	0.270	0.276	0.281	0.287	0.292	0.298	0.303	0.309	0.314
97	0.255	0.261	0.267	0.274	0.280	0.286	0.292	0.299	0.305	0.311	0.317	0.323	0.330	0.336
98	0.162	0.274	0.281	0.289	0.296	0.303	0.310	0.317	0.324	0.331	0.338	0.345	0.352	0.358
99		0.175	0.296	0.304	0.312	0.320	0.328	0.336	0.343	0.351	0.359	0.367	0.375	0.382
100			0.189	0.321	0.329	0.338	0.347	0.356	0.364	0.373	0.382	0.390	0.399	0.408
101				0.204	0.348	0.357	0.367	0.376	0.386	0.395	0.405	0.415	0.424	0.434
102					0.222	0.377	0.388	0.398	0.409	0.419	0.429	0.440	0.450	0.461
103						0.241	0.409	0.421	0.432	0.443	0.455	0.466	0.478	0.489
104							0.261	0.444	0.457	0.469	0.481	0.494	0.506	0.518
105								0.283	0.482	0.495	0.509	0.522	0.535	0.549
106									0.307	0.523	0.537	0.551	0.565	0.580
107										0.333	0.566	0.582	0.597	0.612
108											0.361	0.613	0.629	0.645
109												0.390	0.663	0.680
110													0.422	0.715

Table C.21: m_x , Indians, Females, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.042	0.042	0.041	0.040	0.040	0.039	0.038	0.038	0.037	0.037	0.036	0.035	0.035
71	0.046	0.045	0.045	0.044	0.043	0.042	0.042	0.041	0.040	0.040	0.039	0.038	0.038
72	0.050	0.049	0.049	0.048	0.047	0.046	0.046	0.045	0.044	0.043	0.043	0.042	0.041
73	0.055	0.054	0.053	0.052	0.051	0.050	0.050	0.049	0.048	0.047	0.047	0.046	0.045
74	0.059	0.058	0.058	0.057	0.056	0.055	0.054	0.053	0.053	0.052	0.051	0.050	0.049
75	0.064	0.064	0.063	0.062	0.061	0.060	0.059	0.058	0.058	0.057	0.056	0.055	0.054
76	0.070	0.069	0.068	0.067	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.061	0.060
77	0.076	0.075	0.074	0.073	0.073	0.072	0.071	0.070	0.069	0.068	0.068	0.067	0.066
78	0.083	0.082	0.081	0.080	0.079	0.078	0.078	0.077	0.076	0.075	0.074	0.073	0.073
79	0.090	0.089	0.088	0.087	0.086	0.086	0.085	0.084	0.083	0.083	0.082	0.081	0.080
80	0.097	0.096	0.096	0.095	0.094	0.094	0.093	0.092	0.091	0.091	0.090	0.089	0.089
81	0.105	0.105	0.104	0.103	0.103	0.102	0.102	0.101	0.100	0.100	0.099	0.098	0.098
82	0.114	0.113	0.113	0.112	0.112	0.111	0.111	0.110	0.110	0.109	0.109	0.108	0.108
83	0.123	0.123	0.123	0.122	0.122	0.122	0.121	0.121	0.121	0.120	0.120	0.119	0.119
84	0.133	0.133	0.133	0.133	0.133	0.133	0.132	0.132	0.132	0.132	0.132	0.132	0.131
85	0.144	0.144	0.144	0.144	0.144	0.144	0.144	0.144	0.145	0.145	0.145	0.145	0.145
86	0.155	0.156	0.156	0.156	0.157	0.157	0.157	0.158	0.158	0.158	0.159	0.159	0.159
87	0.167	0.168	0.169	0.169	0.170	0.171	0.171	0.172	0.173	0.173	0.174	0.174	0.175
88	0.180	0.181	0.183	0.184	0.184	0.185	0.186	0.187	0.188	0.189	0.190	0.191	0.192
89	0.194	0.196	0.197	0.199	0.200	0.201	0.203	0.204	0.205	0.207	0.208	0.209	0.210
90	0.209	0.211	0.213	0.215	0.216	0.218	0.220	0.222	0.223	0.225	0.227	0.229	0.230
91	0.225	0.227	0.229	0.232	0.234	0.236	0.238	0.241	0.243	0.245	0.247	0.249	0.252
92	0.242	0.244	0.247	0.250	0.253	0.256	0.258	0.261	0.264	0.266	0.269	0.272	0.274
93	0.259	0.263	0.266	0.270	0.273	0.276	0.279	0.283	0.286	0.289	0.293	0.296	0.299
94	0.278	0.282	0.286	0.290	0.294	0.298	0.302	0.306	0.310	0.314	0.318	0.321	0.325
95	0.298	0.303	0.308	0.312	0.317	0.321	0.326	0.331	0.335	0.340	0.344	0.349	0.353
96	0.319	0.325	0.330	0.335	0.341	0.346	0.351	0.357	0.362	0.367	0.372	0.378	0.383
97	0.342	0.348	0.354	0.360	0.366	0.372	0.378	0.384	0.390	0.396	0.402	0.408	0.414
98	0.365	0.372	0.379	0.386	0.393	0.400	0.407	0.414	0.420	0.427	0.434	0.441	0.448
99	0.390	0.398	0.406	0.413	0.421	0.429	0.437	0.444	0.452	0.460	0.467	0.475	0.483
100	0.416	0.425	0.433	0.442	0.451	0.459	0.468	0.477	0.485	0.494	0.502	0.511	0.520
101	0.443	0.453	0.462	0.472	0.481	0.491	0.501	0.510	0.520	0.529	0.539	0.548	0.558
102	0.471	0.482	0.492	0.503	0.513	0.524	0.534	0.545	0.555	0.566	0.576	0.587	0.597
103	0.500	0.512	0.523	0.535	0.546	0.558	0.569	0.581	0.592	0.603	0.615	0.626	0.638
104	0.531	0.543	0.555	0.568	0.580	0.593	0.605	0.617	0.630	0.642	0.655	0.667	0.680
105	0.562	0.575	0.589	0.602	0.615	0.629	0.642	0.655	0.669	0.682	0.696	0.709	0.723
106	0.594	0.608	0.623	0.637	0.651	0.666	0.680	0.695	0.709	0.723	0.738	0.752	0.767
107	0.627	0.643	0.658	0.673	0.689	0.704	0.719	0.735	0.750	0.766	0.781	0.796	0.812
108	0.662	0.678	0.694	0.711	0.727	0.743	0.760	0.776	0.793	0.809	0.825	0.842	0.858
109	0.697	0.714	0.732	0.749	0.766	0.784	0.801	0.819	0.836	0.853	0.871	0.888	0.906
110	0.734	0.752	0.770	0.788	0.807	0.825	0.844	0.862	0.880	0.899	0.917	0.936	0.954

Table C.22: m_x, Whites, Males, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.056	0.055	0.054	0.053	0.052	0.052	0.051	0.050	0.049	0.048	0.048	0.047	0.046	0.045
71	0.060	0.060	0.059	0.058	0.057	0.056	0.055	0.054	0.053	0.052	0.052	0.051	0.050	0.049
72	0.066	0.065	0.064	0.063	0.062	0.061	0.060	0.059	0.058	0.057	0.056	0.055	0.054	0.053
73	0.071	0.070	0.069	0.068	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.060	0.059	0.058
74	0.077	0.076	0.075	0.074	0.073	0.072	0.070	0.069	0.068	0.067	0.066	0.065	0.064	0.063
75	0.084	0.083	0.081	0.080	0.079	0.078	0.076	0.075	0.074	0.073	0.072	0.071	0.069	0.068
76	0.091	0.089	0.088	0.087	0.085	0.084	0.083	0.082	0.080	0.079	0.078	0.077	0.075	0.074
77	0.098	0.096	0.095	0.094	0.092	0.091	0.090	0.088	0.087	0.086	0.085	0.083	0.082	0.081
78	0.105	0.104	0.103	0.101	0.100	0.098	0.097	0.096	0.095	0.093	0.092	0.091	0.089	0.088
79	0.113	0.112	0.111	0.109	0.108	0.106	0.105	0.104	0.102	0.101	0.100	0.099	0.097	0.096
80	0.122	0.120	0.119	0.118	0.116	0.115	0.114	0.112	0.111	0.110	0.108	0.107	0.106	0.105
81	0.131	0.129	0.128	0.127	0.125	0.124	0.123	0.122	0.120	0.119	0.118	0.117	0.115	0.114
82	0.140	0.139	0.137	0.136	0.135	0.134	0.133	0.131	0.130	0.129	0.128	0.127	0.125	0.124
83	0.150	0.149	0.147	0.146	0.145	0.144	0.143	0.142	0.141	0.140	0.138	0.137	0.136	0.135
84	0.160	0.159	0.158	0.157	0.156	0.155	0.154	0.153	0.152	0.151	0.150	0.149	0.148	0.147
85	0.171	0.170	0.169	0.169	0.168	0.167	0.166	0.165	0.164	0.163	0.162	0.162	0.161	0.160
86	0.183	0.182	0.181	0.181	0.180	0.179	0.178	0.178	0.177	0.176	0.176	0.175	0.174	0.174
87	0.195	0.194	0.194	0.193	0.193	0.192	0.192	0.191	0.191	0.190	0.190	0.189	0.189	0.188
88	0.208	0.207	0.207	0.207	0.207	0.206	0.206	0.206	0.205	0.205	0.205	0.205	0.204	0.204
89	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.221
90	0.235	0.236	0.236	0.236	0.236	0.237	0.237	0.237	0.238	0.238	0.238	0.238	0.239	0.239
91	0.250	0.251	0.251	0.252	0.253	0.253	0.254	0.255	0.255	0.256	0.256	0.257	0.258	0.258
92	0.266	0.267	0.268	0.269	0.270	0.271	0.272	0.273	0.274	0.275	0.276	0.277	0.278	0.279
93	0.282	0.284	0.285	0.287	0.288	0.289	0.291	0.292	0.294	0.295	0.296	0.298	0.299	0.300
94	0.299	0.301	0.303	0.305	0.307	0.309	0.311	0.313	0.314	0.316	0.318	0.320	0.322	0.323
95	0.318	0.320	0.322	0.325	0.327	0.329	0.332	0.334	0.336	0.339	0.341	0.343	0.345	0.348
96	0.337	0.340	0.342	0.345	0.348	0.351	0.354	0.357	0.360	0.362	0.365	0.368	0.371	0.374
97	0.357	0.360	0.364	0.367	0.370	0.374	0.377	0.381	0.384	0.387	0.391	0.394	0.398	0.401
98	0.378	0.382	0.386	0.390	0.394	0.398	0.402	0.406	0.410	0.414	0.418	0.422	0.426	0.430
99		0.404	0.409	0.414	0.418	0.423	0.428	0.432	0.437	0.442	0.446	0.451	0.455	0.460
100			0.433	0.439	0.444	0.449	0.455	0.460	0.465	0.471	0.476	0.481	0.487	0.492
101				0.465	0.471	0.477	0.483	0.489	0.495	0.501	0.507	0.513	0.519	0.525
102					0.499	0.506	0.513	0.520	0.527	0.533	0.540	0.547	0.554	0.561
103						0.536	0.544	0.552	0.559	0.567	0.574	0.582	0.590	0.597
104							0.576	0.585	0.593	0.602	0.610	0.619	0.627	0.635
105								0.619	0.628	0.638	0.647	0.656	0.666	0.675
106									0.665	0.675	0.685	0.695	0.705	0.716
107										0.713	0.724	0.735	0.746	0.757
108											0.764	0.776	0.788	0.800
109												0.818	0.830	0.843
110													0.874	0.888

Table C.23: m_x, Whites, Males, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.045	0.044	0.043	0.042	0.042	0.041	0.040	0.040	0.039	0.038	0.037	0.037	0.036
71	0.048	0.047	0.047	0.046	0.045	0.044	0.043	0.043	0.042	0.041	0.040	0.040	0.039
72	0.052	0.051	0.050	0.050	0.049	0.048	0.047	0.046	0.045	0.044	0.044	0.043	0.042
73	0.057	0.056	0.055	0.054	0.053	0.052	0.051	0.050	0.049	0.048	0.047	0.047	0.046
74	0.062	0.061	0.060	0.059	0.058	0.057	0.056	0.055	0.054	0.053	0.052	0.051	0.050
75	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.060	0.059	0.058	0.057	0.056	0.055
76	0.073	0.072	0.071	0.070	0.069	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0.060
77	0.080	0.079	0.077	0.076	0.075	0.074	0.073	0.072	0.070	0.069	0.068	0.067	0.066
78	0.087	0.086	0.085	0.083	0.082	0.081	0.080	0.079	0.077	0.076	0.075	0.074	0.073
79	0.095	0.094	0.092	0.091	0.090	0.089	0.088	0.086	0.085	0.084	0.083	0.082	0.081
80	0.103	0.102	0.101	0.100	0.099	0.097	0.096	0.095	0.094	0.093	0.092	0.091	0.089
81	0.113	0.112	0.110	0.109	0.108	0.107	0.106	0.105	0.104	0.102	0.101	0.100	0.099
82	0.123	0.122	0.121	0.120	0.119	0.117	0.116	0.115	0.114	0.113	0.112	0.111	0.110
83	0.134	0.133	0.132	0.131	0.130	0.129	0.128	0.127	0.126	0.125	0.124	0.123	0.122
84	0.146	0.145	0.144	0.143	0.142	0.141	0.140	0.139	0.138	0.137	0.136	0.136	0.135
85	0.159	0.158	0.157	0.156	0.156	0.155	0.154	0.153	0.152	0.151	0.151	0.150	0.149
86	0.173	0.172	0.171	0.171	0.170	0.169	0.169	0.168	0.167	0.167	0.166	0.165	0.165
87	0.188	0.187	0.187	0.186	0.186	0.185	0.185	0.184	0.184	0.183	0.183	0.182	0.182
88	0.204	0.203	0.203	0.203	0.203	0.202	0.202	0.202	0.201	0.201	0.201	0.200	0.200
89	0.221	0.221	0.221	0.221	0.221	0.221	0.221	0.220	0.220	0.220	0.220	0.220	0.220
90	0.239	0.239	0.240	0.240	0.240	0.240	0.241	0.241	0.241	0.241	0.241	0.242	0.242
91	0.259	0.259	0.260	0.260	0.261	0.261	0.262	0.263	0.263	0.264	0.264	0.265	0.265
92	0.280	0.280	0.281	0.282	0.283	0.284	0.285	0.286	0.287	0.288	0.288	0.289	0.290
93	0.302	0.303	0.304	0.306	0.307	0.308	0.309	0.311	0.312	0.313	0.314	0.316	0.317
94	0.325	0.327	0.329	0.330	0.332	0.334	0.336	0.337	0.339	0.341	0.342	0.344	0.346
95	0.350	0.352	0.354	0.357	0.359	0.361	0.363	0.365	0.368	0.370	0.372	0.374	0.376
96	0.376	0.379	0.382	0.385	0.387	0.390	0.393	0.395	0.398	0.401	0.403	0.406	0.409
97	0.404	0.407	0.411	0.414	0.417	0.421	0.424	0.427	0.430	0.434	0.437	0.440	0.443
98	0.434	0.437	0.441	0.445	0.449	0.453	0.457	0.461	0.465	0.468	0.472	0.476	0.480
99	0.465	0.469	0.474	0.478	0.483	0.487	0.492	0.496	0.501	0.505	0.510	0.514	0.519
100	0.497	0.502	0.508	0.513	0.518	0.523	0.529	0.534	0.539	0.544	0.549	0.554	0.560
101	0.531	0.537	0.543	0.549	0.555	0.561	0.567	0.573	0.579	0.585	0.591	0.597	0.603
102	0.567	0.574	0.581	0.588	0.594	0.601	0.608	0.614	0.621	0.628	0.635	0.641	0.648
103	0.605	0.612	0.620	0.628	0.635	0.643	0.650	0.658	0.665	0.673	0.680	0.688	0.695
104	0.644	0.652	0.661	0.669	0.678	0.686	0.694	0.703	0.711	0.720	0.728	0.736	0.745
105	0.684	0.694	0.703	0.712	0.721	0.731	0.740	0.749	0.759	0.768	0.777	0.786	0.796
106	0.726	0.736	0.746	0.756	0.766	0.777	0.787	0.797	0.807	0.817	0.827	0.838	0.848
107	0.768	0.779	0.790	0.801	0.812	0.824	0.835	0.846	0.857	0.868	0.879	0.890	0.901
108	0.812	0.824	0.836	0.848	0.860	0.872	0.883	0.895	0.907	0.919	0.931	0.943	0.955
109	0.856	0.869	0.882	0.895	0.908	0.921	0.933	0.946	0.959	0.972	0.985	0.998	1.011
110	0.902	0.915	0.929	0.943	0.957	0.971	0.984	0.998	1.012	1.026	1.040	1.054	1.067

Table C.24: m_x, Whites, Females, 1985-1998

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
70	0.029	0.029	0.029	0.028	0.028	0.028	0.028	0.028	0.028	0.027	0.027	0.027	0.027	0.027
71	0.032	0.031	0.031	0.031	0.031	0.031	0.030	0.030	0.030	0.030	0.029	0.029	0.029	0.029
72	0.035	0.034	0.034	0.034	0.033	0.033	0.033	0.033	0.032	0.032	0.032	0.032	0.031	0.031
73	0.038	0.038	0.037	0.037	0.037	0.036	0.036	0.036	0.035	0.035	0.035	0.034	0.034	0.034
74	0.042	0.041	0.041	0.041	0.040	0.040	0.039	0.039	0.039	0.038	0.038	0.038	0.037	0.037
75	0.046	0.046	0.045	0.045	0.044	0.044	0.043	0.043	0.043	0.042	0.042	0.041	0.041	0.041
76	0.051	0.050	0.050	0.049	0.049	0.049	0.048	0.048	0.047	0.047	0.046	0.046	0.045	0.045
77	0.056	0.056	0.055	0.055	0.054	0.054	0.053	0.053	0.052	0.052	0.051	0.051	0.050	0.050
78	0.062	0.062	0.061	0.061	0.060	0.060	0.059	0.059	0.058	0.057	0.057	0.056	0.056	0.055
79	0.069	0.069	0.068	0.067	0.067	0.066	0.066	0.065	0.064	0.064	0.063	0.063	0.062	0.062
80	0.077	0.076	0.075	0.075	0.074	0.074	0.073	0.072	0.072	0.071	0.070	0.070	0.069	0.069
81	0.085	0.084	0.084	0.083	0.082	0.082	0.081	0.080	0.080	0.079	0.079	0.078	0.077	0.077
82	0.094	0.094	0.093	0.092	0.092	0.091	0.090	0.090	0.089	0.088	0.088	0.087	0.086	0.086
83	0.104	0.104	0.103	0.102	0.102	0.101	0.100	0.100	0.099	0.098	0.098	0.097	0.096	0.096
84	0.115	0.115	0.114	0.113	0.113	0.112	0.111	0.111	0.110	0.109	0.109	0.108	0.107	0.107
85	0.127	0.127	0.126	0.125	0.125	0.124	0.123	0.123	0.122	0.121	0.121	0.120	0.119	0.119
86	0.141	0.140	0.139	0.139	0.138	0.137	0.137	0.136	0.135	0.135	0.134	0.133	0.133	0.132
87	0.155	0.154	0.154	0.153	0.152	0.152	0.151	0.150	0.150	0.149	0.149	0.148	0.147	0.147
88	0.170	0.170	0.169	0.168	0.168	0.167	0.167	0.166	0.166	0.165	0.165	0.164	0.163	0.163
89	0.187	0.186	0.186	0.185	0.185	0.184	0.184	0.183	0.183	0.182	0.182	0.181	0.181	0.180
90	0.204	0.204	0.204	0.203	0.203	0.202	0.202	0.202	0.201	0.201	0.200	0.200	0.200	0.199
91	0.223	0.223	0.223	0.223	0.222	0.222	0.222	0.221	0.221	0.221	0.221	0.220	0.220	0.220
92	0.244	0.243	0.243	0.243	0.243	0.243	0.243	0.243	0.242	0.242	0.242	0.242	0.242	0.242
93	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265
94	0.288	0.288	0.289	0.289	0.289	0.289	0.289	0.290	0.290	0.290	0.290	0.290	0.290	0.291
95	0.313	0.313	0.313	0.314	0.314	0.315	0.315	0.315	0.316	0.316	0.317	0.317	0.317	0.318
96	0.338	0.339	0.340	0.340	0.341	0.342	0.342	0.343	0.343	0.344	0.345	0.345	0.346	0.346
97	0.366	0.367	0.367	0.368	0.369	0.370	0.371	0.372	0.373	0.373	0.374	0.375	0.376	0.377
98	0.394	0.396	0.397	0.398	0.399	0.400	0.401	0.402	0.404	0.405	0.406	0.407	0.408	0.409
99		0.426	0.428	0.429	0.430	0.432	0.433	0.435	0.436	0.438	0.439	0.440	0.442	0.443
100			0.460	0.462	0.464	0.465	0.467	0.469	0.471	0.472	0.474	0.476	0.477	0.479
101				0.496	0.498	0.500	0.503	0.505	0.507	0.509	0.511	0.513	0.515	0.517
102					0.535	0.537	0.540	0.542	0.545	0.547	0.550	0.552	0.554	0.557
103						0.576	0.579	0.582	0.584	0.587	0.590	0.593	0.596	0.599
104							0.619	0.623	0.626	0.629	0.632	0.636	0.639	0.642
105								0.665	0.669	0.673	0.676	0.680	0.684	0.688
106									0.714	0.718	0.722	0.726	0.730	0.735
107										0.764	0.769	0.774	0.778	0.783
108											0.816	0.822	0.827	0.832
109												0.870	0.876	0.882
110													0.926	0.932

Table C.25: m_x, Whites, Females, 1999-2011

Age	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
70	0.027	0.026	0.026	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025	0.024
71	0.029	0.028	0.028	0.028	0.028	0.028	0.027	0.027	0.027	0.027	0.026	0.026	0.026
72	0.031	0.031	0.030	0.030	0.030	0.030	0.029	0.029	0.029	0.028	0.028	0.028	0.028
73	0.033	0.033	0.033	0.033	0.032	0.032	0.032	0.031	0.031	0.031	0.030	0.030	0.030
74	0.037	0.036	0.036	0.036	0.035	0.035	0.035	0.034	0.034	0.033	0.033	0.033	0.032
75	0.040	0.040	0.039	0.039	0.039	0.038	0.038	0.037	0.037	0.037	0.036	0.036	0.036
76	0.044	0.044	0.044	0.043	0.043	0.042	0.042	0.041	0.041	0.041	0.040	0.040	0.039
77	0.049	0.049	0.048	0.048	0.047	0.047	0.046	0.046	0.045	0.045	0.045	0.044	0.044
78	0.055	0.054	0.054	0.053	0.053	0.052	0.052	0.051	0.051	0.050	0.050	0.049	0.049
79	0.061	0.060	0.060	0.059	0.059	0.058	0.058	0.057	0.057	0.056	0.056	0.055	0.054
80	0.068	0.067	0.067	0.066	0.066	0.065	0.065	0.064	0.063	0.063	0.062	0.062	0.061
81	0.076	0.075	0.075	0.074	0.074	0.073	0.072	0.072	0.071	0.071	0.070	0.069	0.069
82	0.085	0.084	0.084	0.083	0.082	0.082	0.081	0.080	0.080	0.079	0.079	0.078	0.077
83	0.095	0.094	0.094	0.093	0.092	0.092	0.091	0.090	0.090	0.089	0.088	0.088	0.087
84	0.106	0.105	0.105	0.104	0.103	0.103	0.102	0.101	0.101	0.100	0.099	0.099	0.098
85	0.118	0.117	0.117	0.116	0.116	0.115	0.114	0.114	0.113	0.112	0.112	0.111	0.110
86	0.132	0.131	0.130	0.130	0.129	0.128	0.128	0.127	0.127	0.126	0.125	0.125	0.124
87	0.146	0.146	0.145	0.145	0.144	0.143	0.143	0.142	0.142	0.141	0.140	0.140	0.139
88	0.162	0.162	0.161	0.161	0.160	0.160	0.159	0.159	0.158	0.158	0.157	0.156	0.156
89	0.180	0.179	0.179	0.179	0.178	0.178	0.177	0.177	0.176	0.176	0.175	0.175	0.174
90	0.199	0.199	0.198	0.198	0.197	0.197	0.197	0.196	0.196	0.196	0.195	0.195	0.194
91	0.219	0.219	0.219	0.219	0.218	0.218	0.218	0.218	0.217	0.217	0.217	0.216	0.216
92	0.242	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.240	0.240	0.240	0.240	0.240
93	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265
94	0.291	0.291	0.291	0.291	0.292	0.292	0.292	0.292	0.292	0.292	0.293	0.293	0.293
95	0.318	0.318	0.319	0.319	0.320	0.320	0.320	0.321	0.321	0.321	0.322	0.322	0.322
96	0.347	0.348	0.348	0.349	0.349	0.350	0.351	0.351	0.352	0.352	0.353	0.353	0.354
97	0.378	0.379	0.379	0.380	0.381	0.382	0.383	0.384	0.384	0.385	0.386	0.387	0.388
98	0.410	0.411	0.412	0.414	0.415	0.416	0.417	0.418	0.419	0.420	0.421	0.422	0.423
99	0.445	0.446	0.447	0.449	0.450	0.452	0.453	0.454	0.456	0.457	0.459	0.460	0.461
100	0.481	0.483	0.484	0.486	0.488	0.490	0.491	0.493	0.495	0.496	0.498	0.500	0.501
101	0.519	0.521	0.523	0.525	0.527	0.529	0.532	0.534	0.536	0.538	0.540	0.542	0.544
102	0.559	0.562	0.564	0.567	0.569	0.571	0.574	0.576	0.579	0.581	0.584	0.586	0.588
103	0.601	0.604	0.607	0.610	0.613	0.616	0.618	0.621	0.624	0.627	0.630	0.632	0.635
104	0.645	0.649	0.652	0.655	0.658	0.662	0.665	0.668	0.671	0.675	0.678	0.681	0.684
105	0.691	0.695	0.699	0.702	0.706	0.710	0.713	0.717	0.721	0.724	0.728	0.732	0.735
106	0.739	0.743	0.747	0.751	0.755	0.760	0.764	0.768	0.772	0.776	0.780	0.784	0.789
107	0.787	0.792	0.797	0.801	0.806	0.811	0.815	0.820	0.825	0.829	0.834	0.838	0.843
108	0.837	0.842	0.847	0.852	0.857	0.863	0.868	0.873	0.878	0.883	0.888	0.893	0.898
109	0.887	0.893	0.898	0.904	0.910	0.915	0.921	0.926	0.932	0.938	0.943	0.949	0.954
110	0.938	0.944	0.950	0.956	0.963	0.969	0.975	0.981	0.987	0.993	0.999	1.005	1.011

C.1.4 Best estimate mortality rates, 2011

Table C.26: m_x, Males, Ages 70-110, 2011

Age	African	Coloured	Indian	White	National
70	0.0665	0.0536	0.0487	0.0361	0.0549
71	0.0702	0.0582	0.0516	0.0390	0.0585
72	0.0742	0.0631	0.0548	0.0421	0.0624
73	0.0784	0.0684	0.0584	0.0457	0.0665
74	0.0828	0.0740	0.0627	0.0498	0.0710
75	0.0874	0.0801	0.0675	0.0546	0.0758
76	0.0923	0.0865	0.0731	0.0600	0.0809
77	0.0974	0.0933	0.0793	0.0661	0.0864
78	0.1027	0.1005	0.0864	0.0730	0.0923
79	0.1084	0.1082	0.0943	0.0808	0.0986
80	0.1143	0.1163	0.1032	0.0894	0.1053
81	0.1205	0.1249	0.1131	0.0991	0.1124
82	0.1270	0.1340	0.1241	0.1098	0.1200
83	0.1339	0.1437	0.1363	0.1216	0.1281
84	0.1411	0.1539	0.1497	0.1346	0.1366
85	0.1487	0.1647	0.1646	0.1489	0.1456
86	0.1567	0.1761	0.1810	0.1645	0.1552
87	0.1651	0.1882	0.1989	0.1816	0.1653
88	0.1739	0.2009	0.2186	0.2001	0.1759
89	0.1831	0.2143	0.2400	0.2201	0.1872
90	0.1929	0.2284	0.2634	0.2417	0.1990
91	0.2031	0.2433	0.2889	0.2651	0.2115
92	0.2139	0.2589	0.3165	0.2901	0.2246
93	0.2252	0.2754	0.3463	0.3170	0.2383
94	0.2371	0.2928	0.3786	0.3456	0.2528
95	0.2495	0.3110	0.4134	0.3762	0.2679
96	0.2626	0.3302	0.4508	0.4088	0.2838
97	0.2764	0.3503	0.4909	0.4434	0.3004
98	0.2908	0.3714	0.5338	0.4800	0.3177
99	0.3059	0.3935	0.5795	0.5187	0.3358
100	0.3217	0.4167	0.6279	0.5596	0.3548
101	0.3383	0.4410	0.6787	0.6027	0.3745
102	0.3557	0.4664	0.7316	0.6479	0.3951
103	0.3739	0.4930	0.7866	0.6954	0.4165
104	0.3929	0.5208	0.8436	0.7447	0.4388
105	0.4128	0.5498	0.9028	0.7956	0.4619
106	0.4336	0.5800	0.9641	0.8477	0.4859
107	0.4553	0.6114	1.0274	0.9010	0.5108
108	0.4780	0.6437	1.0928	0.9553	0.5364
109	0.5015	0.6769	1.1604	1.0108	0.5627
110	0.5259	0.7110	1.2300	1.0673	0.5896

Table C.27: m_x, Females, Ages 70-110, 2011

Age	African	Coloured	Indian	White	National
70	0.0415	0.0334	0.0347	0.0244	0.0397
71	0.0444	0.0368	0.0378	0.0260	0.0419
72	0.0474	0.0405	0.0412	0.0277	0.0442
73	0.0507	0.0446	0.0451	0.0299	0.0469
74	0.0542	0.0490	0.0494	0.0325	0.0499
75	0.0580	0.0539	0.0543	0.0356	0.0533
76	0.0621	0.0591	0.0598	0.0393	0.0570
77	0.0666	0.0648	0.0659	0.0436	0.0612
78	0.0713	0.0709	0.0727	0.0487	0.0657
79	0.0764	0.0775	0.0802	0.0545	0.0708
80	0.0819	0.0847	0.0886	0.0611	0.0762
81	0.0877	0.0923	0.0978	0.0688	0.0822
82	0.0940	0.1005	0.1080	0.0774	0.0887
83	0.1007	0.1092	0.1191	0.0872	0.0958
84	0.1078	0.1185	0.1313	0.0982	0.1034
85	0.1154	0.1285	0.1446	0.1104	0.1116
86	0.1235	0.1390	0.1591	0.1241	0.1205
87	0.1321	0.1502	0.1749	0.1393	0.1300
88	0.1413	0.1621	0.1919	0.1560	0.1402
89	0.1510	0.1746	0.2103	0.1743	0.1510
90	0.1613	0.1878	0.2302	0.1944	0.1626
91	0.1722	0.2017	0.2516	0.2162	0.1750
92	0.1837	0.2163	0.2745	0.2399	0.1881
93	0.1959	0.2317	0.2990	0.2654	0.2020
94	0.2087	0.2478	0.3252	0.2929	0.2167
95	0.2223	0.2647	0.3531	0.3224	0.2322
96	0.2366	0.2824	0.3828	0.3540	0.2485
97	0.2517	0.3008	0.4143	0.3876	0.2657
98	0.2675	0.3200	0.4477	0.4234	0.2837
99	0.2842	0.3400	0.4828	0.4613	0.3027
100	0.3017	0.3609	0.5196	0.5015	0.3225
101	0.3200	0.3825	0.5578	0.5438	0.3432
102	0.3392	0.4050	0.5972	0.5884	0.3648
103	0.3594	0.4283	0.6378	0.6352	0.3874
104	0.3804	0.4525	0.6796	0.6843	0.4109
105	0.4024	0.4774	0.7226	0.7355	0.4353
106	0.4254	0.5032	0.7666	0.7886	0.4605
107	0.4494	0.5298	0.8119	0.8430	0.4867
108	0.4743	0.5571	0.8583	0.8984	0.5135
109	0.5001	0.5850	0.9058	0.9545	0.5408
110	0.5268	0.6133	0.9544	1.0113	0.5685