FULL LIFE TABLES FOR SOUTH AFRICA FROM VITAL REGISTRATION DATA, 2006-2008

Simon H. Marandu

Thesis submitted to the Faculty of Commerce in partial fulfilment for the degree of Master of Philosophy in Demography, University of Cape Town

Centre for Actuarial Research (CARe)
November 2011

PLAGIARISM DECLARATION

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own. I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this report from the work(s) of other people has been attributed, and has been cited and referenced. This report is my own work. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signed

Simon Hlomayi Marandu

Date:

ABSTRACT

This research derives a set of full life tables for South Africa as a whole and by population group using vital registration data for the period 2006-2008. Given that not all deaths are registered, the research assesses the level of completeness of death registration for the national population and for all the population groups separately by using the deaths distribution methods.

Previous attempts at producing full life tables for South Africa were hampered by the absence of estimates of completeness of deaths for those below the age of 15. A study by Darikwa(2009) produced estimates of completeness of death registration for the under-fives using data from the 2001 national census and the 2007 community survey. Results from that study were used in this research to estimate completeness of death registration for the under-fives. The levels of completeness for the 5-14 age groups were estimated by assuming a linear trend between the levels of completeness for the under-fives and those above the age 15.

In addition, 26 per cent of the registered deaths do not indicate the population group of the deceased. In this research the population group was imputed using multinomial logistic 'hotdeck' imputation in Stata to assign population groups to all records without population group. This enabled levels of completeness and the mortality rates for all the population groups to be estimated. The central mortality rates were graduated using cubic spline smoothing in Matlab up to age 85 with knots at every age. Cubic spline smoothing was chosen ahead of other parametric methods as it does not impose a shape and profile on the crude rates and hence tends to produce graduated mortality rates that are consistent with the crude rates though it is less than ideal when crude rates fluctuate too heavily.

It has been observed that mortality rates for very old ages are affected by age misstatement and that the number of deaths are relatively small, which renders the mortality rates unstable, hence a method proposed by Coale and Kisker was used to extrapolate the central mortality rates from age 85 to age 109 for the national and the population group populations.

However, the completeness levels estimated after imputation do not tally with what is generally expected in the case of White and Coloured populations. Indications are that the imputation process allocated too many records to the black African population group and

hence less were allocated to the White and Coloured population groups. Further, it can be noted that the shapes of the mortality curves were not regular, more so for the small population groups.

Comparisons were made with estimates from the UNPD(2011) and the ASSA2008 (2010) demographic projection model with respect to the infant mortality, adult mortality, life expectancy and the general progressions of the mortality curves. The results of this study are generally consistent with the results produced by other researchers for the periods close to 2006-2008 for the national and black African population groups.

ACKNOWLEDGEMENTS

I would like to acknowledge the immense contributions made to this project by the following:

1. Professor Rob Dorrington for supervising the project and giving wise insights throughout.
2. The Centre for Actuarial Research (CARe) for putting together a challenging programme which pushed us to the precipice of our abilities as individuals and as a class.
3. The Hewlett Foundation for funding my studies and my stay in Cape Town.
4. My wife Enia and our children: Tatenda and Tamuka, for their relentless support and prayers during this often trying period of our lives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

A life table is produced from a set of mortality rates calculated from the number of deaths and estimates of the person-years of exposure derived from the estimates of the population. A life table can be presented as a complete/full life or in an abridged form. A complete life table presents all the functions in single years of age whereas an abridged life table presents the functions for certain pivotal ages usually spaced at five or ten year intervals. The reliability of estimates of mortality rates is dependent on the quality of the death data and the population estimates across all ages. The challenge of the above exercise in the context of a developing country like South Africa is that none of the two data sets are 100 per cent complete. Death registration is known to be incomplete and the censuses to undercount the population. Complete consensus among the different experts using different models for calculating population estimates is therefore difficult to achieve (Statistics South Africa 2010).

In South Africa, attempts have been made to construct official life tables but in all cases reasonably reliable tables were produced only for Whites, Coloureds and Indians/Asians to the exclusion of black Africans under the South African Life Tables (SALT) series (Dorrington, Moultrie and Timaeus 2004). The unavailability and incompleteness of death records for Africans made it impossible for researchers to produce life tables, including those for black Africans. Some of the earliest attempts to produce full life tables for black Africans were made by Dorrington (1989) for the period 1979-1981 and Dorrington, Bradshaw and Wegner (1999) for the periods 1984-1986 and 1989-91, which made use of data from the vital registration system.

Further estimates were produced which made use of vital registration data and population data from the 1996 and 2001 censuses (Dorrington *et al.* 2004). However, this attempt could not ascertain childhood mortality as there was no way of assessing the level of completeness of the vital registration data for 0-4 age group and the questions on children ever born and children surviving from the 2001 census failed to produce reasonable results. As a result, an estimate of ${}_5q_0$ was calculated for indicative purposes only, to allow the derivation of a complete life table.

The first set of official abridged life tables was published by Statistics South Africa in 2000 based on the 1996 census data, using survival of children ever born (CEB), survival of parents and vital registration data. Other researchers produced significantly different estimates of completeness of death registration from the same data, which made it difficult to choose estimates to rely on (Dorrington *et al.* 2004; Udjo 2008).

1.2 Aims and objectives of the research

This study seeks to produce a complete set of mortality rates and life tables for South Africa and each of the population groups separately, and in the process also estimate the levels of completeness of death registration for each of the population groups, for males and females and for the national population for the period 2006-2008. Incidental to this objective and arising out of the incompleteness of the vital registration system, this study also seeks to establish the impact of adjusting for incompleteness of the 'observed' number of deaths on the distribution and parameters of the 'true' numbers of deaths. The parameters of the new distribution: mean and variance, are useful in the process of assessing the goodness-of-fit of the graduated rates.

Around 26 per cent of the death notification forms do not indicate the population group of the deceased, therefore this study also attempts to impute population group to the unit records that do not indicate the population group of the deceased to allow the level of completeness for the different population groups to be assessed and the mortality rates to be estimated.

1.3 Statement of the problem

The process of producing a full life table in a developing country like South Africa is hampered by the absence of credible estimates of childhood mortality and methods of assessing the levels of completeness of death registration for persons under the age of 15.

The research seeks to produce life tables for all the four population groups in South Africa. However, 26 per cent of the death notification forms for the period 2006-2008 do not have an indication of the population group of the deceased. The absence of such information affects the process of assessing the level of completeness of death registration of the population groups and the process of estimating the central mortality rates for the different population groups.

Data errors in population estimates and the vital registration system lead to life tables with many sudden fluctuations from age to age, which tend to worsen in smaller populations. This study uses the spline smoothing method to graduate the mortality rates.

1.4 Significance of research

Currently, there is no up-to-date full set of life tables disaggregated by population group in South Africa derived from the vital registration system. Such a life table would be useful to insurance companies, and in particular the government, which administers a national insurance project for the aged. It would also assist in assessing the impact of social welfare and health interventions on longevity of people of different population groups and sex.

The process of building population projection models needs mortality estimates which are based on empirical data, therefore the mortality rates produced by this study can be used in building population projection models and in reviewing the current projection models.

1.5 Chapter outline

This study is divided into five chapters. The first chapter is the introduction which covers the background, objectives and relevance of the study. The second chapter reviews the literature relevant to the study, and the third chapter describes the methods used in the study. The fourth chapter provides the results of the study, and the fifth chapter concludes the study by a discussion of the results.

2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews previous efforts to produce life tables for South Africa. Attention is directed at methods and techniques employed to improve the quality of mortality estimates. In almost all the cases that deal with data from developing countries, there are challenges related to data quality that demographers and other researchers face, and this chapter attempts to put these into perspective.

One of the challenges facing demographers in developing countries like South Africa is that of missing and incomplete survey data. In this research the data from the vital registration system is known to be incomplete and some of the death notification forms have missing information relating to the population group of the deceased. Therefore, a discussion of techniques for imputing values for missing fields is important to deal with the aspect related to missing data fields. The methods that are used for estimating and adjusting for incompleteness will also be reviewed.

The chapter concludes by reviewing the graduation and extrapolation techniques that are of interest to this research, that is, literature related to the extrapolation method proposed by Coale and Kisker (1989) and cubic spline smoothing technique.

2.2 History of life tables in South Africa

In the history of South Africa, estimates of mortality rates and life tables were produced under the series called 'The South African Life Tables' (SALT) from 1920 to 1986 (Bah 1997). This entire series of life tables, only included Whites, Coloureds and Indians/Asians, excluding black Africans (Dorrington *et al.* 1999). Black Africans were not included in the vital registration system until 1979 and by that time some of them were not eligible as they were deemed to be citizens of the 'independent' homelands created for black Africans by the Apartheid government of the time (Dorrington *et al.* 2004). This explains in part the failure by demographers to estimate life tables for Black Africans over the period. People who lived in the Transkei, Bophuthatswana, Venda and Ciskei were not included in the vital registration system as they were not considered by the Apartheid government of the time to be citizens of the Republic of South Africa.

In computing the life tables in the SALT series, it was assumed that the vital registration system was 100 per cent complete in respect of deaths of Whites, Coloureds and Indians/Asians (Bah 1997). Bah (1997) contends that mortality was underestimated as a result of this assumption. The last set of full official tables was produced using deaths reported for the period from 1984 to 1986. The tables produced in this instance did not include black Africans, although they included Whites, Coloureds and Indians/Asians. At the time of the 1991 census, no attempt was made to construct life tables by population groups because by that time death certificates did not indicate the population group of the deceased. This arrangement of not including population group on the death certificate was maintained from 1991 to 1997. As a result of this limitation, Statistics South Africa produced only an abridged life table by population group covering the period 1985-1994 using data on the survival of children ever born (CEB) and survival of parents collected during the 1996 census (Statistics South Africa 2000). They produced life tables for 1996 by province using data from the same census in addition to the data from the vital registration system (Dorrington *et al.* 2004).

An effort to include black Africans was made by Dorrington, Bradshaw and Wegner (1999). In this instance, the level and shape of the mortality curve for black Africans was estimated using indirect techniques on the reported deaths for the period covering 1984-1986 and then combined with the official tables of the other population groups by assigning different weights to the different population groups to produce a set of full national life tables. Dorrington, Moultrie and Timaeus (2004) made another attempt to produce a full set of life tables by population group and by province using the 1996 and 2001 censuses, as well as the reported deaths in the intermittent period. The numbers of deaths were adjusted using the level of completeness estimates from the General Growth Balance method proposed by Hill (1987), which took into account the intercensal migration estimates. The biggest challenge faced was in the calculation of childhood mortality as the level of completeness of the vital registration system could not be assessed for the under-fives. Hence, they calculated childhood mortality for the 12 months prior to the 2001 census using the census estimates of the number of children at each age under five and the number of deaths at each age of the under-fives reported by the households for indicative purposes only (Dorrington *et al.* 2004).

2.3 Imputation

When data are collected in a survey, on occasion there are non-responses which fall into two broad categories: unit non-response; and item non-response. Unit non-response is when a respondent does not respond to any of the questions; and item non-response is when a respondent responds to some of the questions and not others (Nordholt 1998). There are ways of correcting for unit non-response. These will be discussed in the next section, which will cover issues surrounding incompleteness of death registration.

The data recorded by the death notification system in South Africa are not only incomplete in the sense that not all deaths are recorded but also because about 26 per cent of the death notification forms do not indicate the population group of the deceased. There are many ways of dealing with missing values. The units with missing values could be removed or one of the many techniques to impute missing values could be applied to create complete data units for analysis. In the case at hand, the missing values constitute a high enough proportion to warrant imputation.

The imputation techniques that are available fall in two broad categories: single imputation; and multiple imputations. Single imputation will provide one value in the place of the missing value whereas the multiple imputation will provide a series of possible values for a single missing response and the probability for each possible replacement value (Andridge and Little 2010). The most commonly used single imputation techniques are 'hot decking' and 'cold decking'.

In the cold deck imputation method, data from a unit in another sample is used to impute the missing value. In other words the missing item can be imputed by taking a value from the same unit from another survey conducted previously. In addition, a set of predetermined rules can be used to impute missing values (Acock 2005). In the case of the hot deck imputation method, data from a different unit with similar specified covariates is used to impute the value of the missing variable. In this instance both the donor unit and the accepting unit are in the same sample or survey data set (Altermayer nd).

Hot decking can either be stochastic or deterministic. When a stochastic hot decking method is applied, the process will involve identifying all the potential donor units, which have similar characteristics to the receiving unit and then picking a random unit from that pool. In the case of the deterministic model, the same process is followed but the difference is that the 'nearest neighbour' to the receiving unit will be nominated as the donor unit if the units are arranged in a particular order (Andridge and Little 2010).

The process of imputing fields in large data sets is cumbersome; hence the use of software has become widespread. Computer based software can be used to create 'donor pools' constituted by units which have similar covariates to the 'recipient' units and then to pick the specific donor unit using specified rules. Mander and Clayton (1999) created software called 'hotdeck6 version 1.65', which can be used in Stata for hot deck imputation. Alternatively, one can build a code using imputation commands that are available in Stata that can perform the imputation.

2.4 Completeness: Death distribution methods

In South Africa, as is the case in most countries in the developing world, not all deaths are captured by the death notification system. However, when compared with other countries in Sub-Saharan Africa, South Africa has made significant strides towards achieving a reasonable level of completeness of registration of deaths. It is generally accepted that the proportions of deaths that are not recorded are higher in the rural areas when compared with the urban areas. Further, in South Africa, different population groups might also have different levels of completeness. In order to estimate reasonably accurate mortality rates, the levels of completeness of death registration have to be computed by population group. The methods that can be used to do that are: the General Growth Balance method (GGB); the Synthetic Extinct Generations method (SEG); and a method that combines both the GGB and the SEG methods (GGB+SEG) proposed by Hill Choi and Timaeus (2005). Hill You and Choi (2009) applied all the three methods on simulated data and concluded that the GGB+SEG method is the best method to use. However, a similar study by Dorrington, Moultrie and Timaeus (2004) concluded that the SEG method (with adjustment for relative coverage of the censuses) is as good if not better when using data with unknown errors. It is apparent that there is not unanimity among researchers on the choice of a method that may be considered to be the best. However, the researchers concluded that all the methods work well when the relevant assumptions are met. Further, Hill, You and Choi (2009) concluded that the GGB+SEG method is suitable when the population is not affected significantly by migration. In the same analysis they proposed that for populations where migration was not insignificant the average of GGB and SEG give the best completeness estimates.

2.4.1 General Growth Balance Method (GGB)

The GGB method was proposed by Hill (1987) as an extension of the original Growth Balance method proposed by William Brass (1975). It was developed to estimate the level of completeness of death registration in comparison to a population count (census). The Growth Balance method is based on the demographic balance equation assuming a closed and stable population. Hill later generalised the method to allow its use without having to assume stability. Therefore, the General Growth Balance method can be applied in populations that are not stable (Hill 1987). This method (GGB) assumes that the population is closed to international migration. This assumption would reduce the population balance equation:

$P_2 = P_1 + B - D + M_n$ to the following, $P_2 = P_1 + B - D$,

where P_2 and P_1 are the true populations at time t_2 and t_1 respectively, B, D and M_n are the true numbers of births, deaths and net international migration respectively between the two censuses. If this equation applies to a population of persons above a certain age x, then:

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

where $P_1(x+)$ and $P_2(x+)$ are the true numbers of persons aged x and over at time t_1 and t_2, $N(x)$ is the true number of persons reaching exact age x in the period, $D(x+)$ is the true number of deaths of people aged x and over. Dividing the above equation throughout by $N(x+)$ which denotes the person years lived by people aged x and over in the intercensal period, yields the following equation:

$$\frac{N(x)}{N(x+)} - r(x+) = \frac{D(x+)}{N(x+)} \qquad (1)$$

where $r(x+)$ is the growth rate at all ages above a certain age x of the population, assuming that it is stable. The term on the right hand side of the equation represents a direct calculation of death rates and if some deaths were not included, then the right hand side will be lower than the left hand side, assuming that the censuses were complete. This is so because the left hand side represents an indirect calculation of the death rate as the growth rate is subtracted from the 'birth rate' (i.e. those becoming age x). The indirect calculation is deemed to be correct as long as the census numbers are accurate.

However, in the event of differential coverage of the two censuses, where k_1 and k_2 represent the level of coverage of the first and second censuses, respectively, and c, the level of completeness of death registration, and assuming that the level of completeness of death registration above a certain age does not vary by age, then the following equations will come into effect:

$$P_1(x+) = \frac{P_1*(x+)}{k_1} \qquad (2a)$$

$$P_2(x+) = \frac{P_2*(x+)}{k_2} \qquad (2b)$$

$$D(x+) = \frac{D*(x+)}{c} \qquad (2c)$$

where *P*(x+)* denotes the enumerated number of persons aged x and over at the time of the census and *D*(x+)* denotes the recorded number of deaths in the vital registration system.

The true annual growth rate *r*(*x*+) between the two censuses *t* years apart can be estimated by assuming that the population grows exponentially and at a constant rate per year, and therefore:

$$r(x+) = \frac{1}{t}\ln\left[\frac{P_2(x+)}{P_1(x+)}\right], \qquad (3)$$

Substituting equations (2a) and (2b) will result in the following:

$$r(x+) = r*(x+) + \frac{1}{t}\ln\left[\frac{k_1}{k_2}\right] \qquad (4)$$

where *r *(x+)* is the growth rate calculated from the observed population figures aged *x* and over.

Since we are assuming that the population grows exponentially, the person years lived by a person aged *x* and over can be approximated by the geometric mean of the initial and final population figures for the number of people aged *x* and over, multiplied by the time *t* between the two censuses, i.e.

$$N(x+) = t\left[P_1(x+)P_2(x+)\right]^{0.5}. \qquad (5)$$

This equation can be simplified by substituting equations (2a) and (2b) to get the following:

$$N(x+) = \frac{N*(x+)}{\left[k_1k_2\right]^{0.5}}. \qquad (6)$$

In the same way the number of people reaching age *x* in a period denoted by *N(x)* can be shown to be

$$N(x) = \frac{N*(x)}{k_1k_2}. \qquad (7)$$

Thus substituting equations (2c), (4), (6) and (7) into equation (1) will result in the following:

$$\frac{N^*(x)}{N^*(x+)} - r^*(x+) = \frac{1}{t}\ln\left[\frac{k_1}{k_2}\right] + \frac{(k_1k_2)^{0.5}}{c}\frac{D^*(x+)}{N^*(x+)}.$$

This can be re-written in simpler terms as:

$$n^*(x) - r^*(x+) = a + b.d^*(x+) \qquad (8)$$

where:

$$n^*(x) = \frac{N^*(x)}{N^*(x+)}, a = \frac{1}{t}\ln\left[\frac{k_1}{k_2}\right], \quad b = \frac{(k_1k_2)^{0.5}}{c} \text{ and } \quad d^*(x+) = \frac{D^*(x+)}{N^*(x+)}.$$

The above equation can be interpreted as a straight line with *b* being the slope and *a,* the intercept, which can be estimated by fitting the line using the observable quantities; (*n*(x)*-*r*(x+)*)*,* *d*(x+)*). The process of fitting the straight line through the above points can be done using a variety of methods: the group mean method; weighted trimmed means method; the least squares regression method; and the orthogonal regression method. The presence of data errors means that not all the points lie on a straight line. As a result, these methods were developed to minimise the effect of age misstatement and other data errors. Most of the methods of fitting the straight line listed above rely on individual judgement and hence different researchers using similar methods may produce different estimates of levels of completeness. The methods based on the application of regression techniques (Least Squares Regression (LSR) and Orthogonal Sums of Squares Regression (OR)) have less room for subjective judgement. Bhat (2002) pointed out that the least squares regression technique only minimises the vertical distances of the residuals and hence proposes the orthogonal regression method as a better option as it minimises the perpendicular distance of the residuals.

It can be noted that the values of k_1 and k_2 cannot be calculated individually but only relative to one another. By assuming that the bigger of the two is 1 and the other will therefore be less than one. From the slopes of the line we get that

$$\frac{k_1}{k_2} = \exp(ta),$$

and the level of completeness of death registration can be estimated by the following equation:

$$c = \frac{(k_1 k_2)^{0.5}}{b}.$$

The completeness of death registration for all the ages above childhood age will be denoted by the value c assuming that completeness does not vary by age (UN Population Division 2002).

In the context of countries which experiences significant migration activity, it is be important to take into account migration statistics over the period between the two censuses. Bhat (2002) proposed a generalisation of the GGB method for countries open to migration. The balancing equation as indicated in equation (8) is generalised to include the net migration rate, as follows:

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

where $nm^*(x+)$ is the partial net intercensal migration rate for persons age x and over. However, it must be noted that this method can only be applied if data on intercensal migration are available.

The General Growth Balance method was subjected to simulated tests to ascertain its performance in the event that the assumptions are violated (Hill and Choi 2004). These tests showed that it allows for additional systematic error introduced by a change in census coverage. However, it is affected by age misreporting in the censuses and in the deaths and by differential coverage by age. Therefore, differential coverage of deaths by age will have an adverse effect on GGB (Hill and Choi 2004). Further tests by Hill, You and Choi (2009) showed that the violation of assumptions related to international migration greatly affected the results from the GGB method. It was shown that when emigration rates are very high the GGB method underestimates mortality. Tests by Dorrington, Timaeus and Moultrie (2008) concurred that if emigration is not accounted for, the GGB method does not perform well.

2.4.2 Synthetic Extinct Generations Method (SEG)

Vincent (1951) developed a method called the Extinct Generations method designed to estimate the number of persons in the old ages using the number of reported deaths. Preston, Coale, Trussel *et al* (1980) extended the Extinct Generations method to allow for the estimation of completeness of death registration for stable populations by calculating the ratio of the enumerated population to the population estimated using the number of reported deaths by age.

The Synthetic Extinct Generations method (SEG) is a further extension of the above methods. It was designed to estimate the completeness of death registration for non-stable populations (Bennett and Horiuchi 1981). Preston, Coale, Trussel *et al* (1980) proposed that the true number of people in a cohort aged *a* at some previous time *t*, *N(a,t)*, can be estimated by the following equation for a stable population:

$N(a,t) = \int_a^\infty D(x,t)\exp\left[\int_a^x r(z,t)dz\right]dx$, where $D(x,t)$ denotes the true number of deaths aged *x* at time *t* and *r*(*z*,*t*) is the growth rate of the population aged *z* at time *t*. The time variable, *t,* can be dropped from the equation as all the functions are at time *t*, in which case the above simplifies to:

$$N(a) = \int_a^\infty D(x)\exp\left[\int_a^x r(z)dz\right]dx \quad . \qquad (1)$$

If the level of completeness of death registration is assumed to be equal to *c*, then the recorded deaths $D^*(x)$, will be such that

$$D^*(x) = cD(x) \quad .$$

Therefore,

$$N(a) = \frac{1}{c}\int_a^\infty D^*(x)\exp\left[\int_a^x r(z)dz\right]dx ,$$

and it can be noted that the integral term represents the 'observed' number of persons in the cohort and can be denoted by $\hat{N}(a)$ such that

$$\hat{N}(a) = \int_a^\infty D^*(x)\exp\left[\int_a^x r(z)dz\right]dz \quad ,$$

and the level of completeness, *c*, can then be estimated by applying the following equation:

$$c = \hat{N}(a)/N(a).$$

Reviewing equation (1) in this section, which has integral limits *a* and infinity, the same equation can be broken into two components: integrals from *a* to *a*+*n* and from *a*+*n* to infinity such that

$$N(a) = \int_a^{a+n} D(x)\exp\left[\int_a^x r(z)dz\right]dx + \left\{\int_{a+n}^\infty D(x)\exp\left[\int_{a+n}^x r(z)dz\right]dx\right\}\exp\left[\int_a^{a+n} r(z)dz\right].$$

The above equation can be simplified to:

$$N(a) = \int_a^{a+n} D(x)\exp\left[\int_a^x r(z)dz\right]dx + N(a+n)\exp\left[\int_a^{a+n} r(z)dz\right].$$

If we define $\int_a^{a+n} D(x)dx = {}_nD_a$ and $\int_a^{a+n} r(z)dz = n.{}_nr_a$, then

$$N(a) = N(a+n)\exp(n.{}_nr_a) + {}_nD_a\exp(y.{}_nr_a),$$

where y represents the time period from age a to the mid-point of the interval $(a, a+n)$ and ${}_nr_a$ is the growth rate experienced by those in the age range from a to $a+n$. Assuming a five-year interval, the above equation can re-written as:

$$N(a) = N(a+5)\exp(5{}_5r_a) + {}_5D_a\exp(2.5{}_5r_a) \qquad (2)$$

where ${}_5D_a$ denotes the number of persons dying in the age group from a to $a+5$. This equation is iterative as the value of $N(a+5)$ is needed to calculate $N(a)$ and this means the value corresponding to the highest age group (the open interval) must be established independently. Bennett and Horiuchi (1981) proposed the following approximation for the open interval:

$$\hat{N}(a) = D^*(a+)\left[\exp(r(a+)e(a) - r(a+)e(a)/6\right]$$

where $e(a)$ denotes the life expectancy at the beginning of the open interval, which value must be taken from elsewhere. It must be noted that the values of the age specific growth rates require two population censuses. However, differential coverage by age between the two censuses will affect the values of the age specific growth rates. If k_1 and k_2 stand for coverage of the first and second censuses respectively, then:

$${}_5r_a = {}_5r_a^* + \ln\left(\frac{k_1}{k_2}\right).$$

The last term on the right hand side is denoted by delta, δ, and the method is referred to as "SEG" in this research.

As in the case of the GGB method discussed in the previous section, the SEG method was also subjected to simulated tests by Hill and Choi (2004) to ascertain its sensitivity to errors that are not covered by the assumptions associated with the method. Hill and Choi (2004) concluded that the SEG method (without delta) was less sensitive to age misreporting than the GGB method. In addition, the SEG method performed better than the GGB method in the cases involving differential coverage by age. However, they concluded that it could not deal with differential census coverage. This assertion was later shown to be incorrect by Dorrington, Timaeus and Moultrie (2008), who pointed out that the method developed by Bennett and Horiuchi had an added capacity to adjust for differential census coverage. The $-\ln(k_1/k_2)$ in the SEG method allows for the quantification and adjustment of differential coverage between two censuses.

2.5 Population Estimates

Population censuses and surveys are meant to collect, primarily, information relating to the number of people at different ages within a particular geographical area. The quality of such information depends on the ability and willingness of people to respond truthfully and correctly to the questions on the census or survey questionnaire. The deaths distribution methods are affected by poor quality data, particularly age misreporting, digit preference (age heaping) and undercounting of persons in some age groups (Hill and Choi 2004). In this research, the 2001 population census data, 2007 Community survey data, mid-year population estimates for 2006, 2007 and 2008 are used. Specifically, the 2001 Population Census data and the 2007 Community Survey data are used to ascertain level of completeness using the deaths distribution methods.

Age misstatement can be assessed using visual inspection or using the various indices: the Bachi Index, Myers Index, Zelnik's Index and Whipple's Index for age accuracy. The Whipple's Index of age accuracy is the most widely used on account of its simplicity (Spoorenberg and Dutreuilh 2007). The method was initially designed to detect age heaping on ages ending with 0 and 5 and was later improved to separate these and then was further improved to allow it to detect the possibility of age heaping on all the ten digits (Spoorenberg and Dutreuilh 2007).

In the case of periods where there are no population counts, estimates from models can be used to estimate the relevant population. In the case of South Africa, there is the ASSA2008 demographic projection model, and projection models produced by the United States Census Bureau and the United Nations Population Division.

2.6 Graduation and Extrapolation

Empirical survey data may be defective in many ways. In the context of demographic data like population census data and reported deaths, there are likely to be cases of age misreporting due to digit preference and age exaggeration and cases of undercounts. Some of the data errors distort the smooth progression of mortality rates over time and some introduce bias.

The graduation process seeks to establish mortality rates that progress smoothly by age but at the same time retain the underlying mortality pattern exhibited by the empirical data (Heligman and Pollard 1980). There are mathematical and statistical ways of smoothing crude data. Broadly, the methods can be grouped as either parametric or non-parametric. The most used parametric equations for graduating mortality rates are the Heligman and Pollard equation (for the whole age range), the Makeham equation, Generalised Linear Models and the Gompertz-Makeham equation (Debon, Montes and Sala 2005). In this research, we concentrate on the cubic spline smoothing method which is a non-parametric method.

The older ages are affected by poor data quality caused by age misstatement. The smaller population sizes at these ages tend to amplify the fluctuations of mortality rates to an extent that smoothing may not yield reasonably smooth mortality rates progression. This limitation can be avoided by extrapolating the mortality curve after a certain cut-off age. These graduation and extrapolation methods are discussed in turn in the next sections.

2.6.1 Cubic Spline Smoothing

Given a curve *g* on an interval [a, b], its roughness can be measured by $\int_a^b \{g''(t)\}^2 dt$, if $g(t)$ is twice differentiable. This function $g(t)$ can be defined as a cubic spline if it is a cubic polynomial on all the sub-intervals (a,t_1), (t_1,t_2),.......,(t_n,b) satisfying $a<t_1<t_2<...<t_n<b$, and its first and second derivatives are continuous at each t_i and on the entire interval. The polynomials intersect smoothly at the t_i points, which are called knots.

If there are *n* observations on the interval *(a, b)* denoted by $(x_1, x_2,....., x_n)$ at times $t_1...t_n$, which are subject to random error, then $g(t)$, a smooth function, will estimate the rough curve produced from the observed data by minimizing the penalised sum of squares function:

$S = \sum_{i=1}^{n} w_i \{x_i - g(t_i)\}^2 + \alpha \int_a^b \{g''(y)\}^2 dy$, where α is a positive smoothing parameter and w_i are strictly positive weights that can either be calculated as the reciprocals of the variances of x_is, assuming that the number of deaths follow a Poisson distribution or can be allocated by the graduator's judgment to achieve closer fits in particular age ranges. The first option gives more emphasis to the data points with smaller variances. The first term in the above function is the weighted residual sum of squares, which quantifies the goodness-of-fit to the observed data, and the smoothing parameter in the second term indicates the degree of roughness tolerated by the graduator. A large value of the smoothing parameter will result in a smooth curve and a small value will allow the curve to follow the observed data more closely (Green and Silverman 1994).

The relative importance of goodness-of-fit to each observation, x_i, is controlled by the values assigned to a vector of weights, w_i. The number and position of the different knots, the weights and the smoothing parameter are flexible and can be set and adjusted subjectively. The knots can be evenly spaced to give uniform splines or optimally chosen to give non-uniform splines. In deriving the English Life Table (ELT) 15, the Government Actuaries (2005) used the 'classical cubic spline regression' technique, which spaces the knots evenly over the age range (Government Actuaries Department 1995). This technique uses all the ages as knots. However, in deriving ELT 14, they used a variable-knot method which requires optimality in establishing the position and number of knots over the age range and hence does not require the knots to be evenly spaced (Haberman 1996).

In ELT 15, they used a model which is similar to the one introduced by Green and Silverman (1994), albeit with a few adjustments. They introduced an extra weight, 1-α which works reciprocally with the smoothing parameter, α, such that the two weights add up to unity and that 0<α<1:

$$S = \alpha \sum_{i=1}^{n} w_i \{x_i - g(t_i)\}^2 + (1-\alpha) \int_a^b \{g''(y)\}^2 dy .$$

The value of the smoothing parameter can be estimated by using the algorithm proposed by Reinsch (1967). It seeks to establish the value of the smoothing parameter, α, which gives a spline function, $g(t)$ such that the first part of the spline model given above, $S_1 = \sum_{i=1}^{n} w_i \{x_i - g(t_i)\}^2$, lies in the interval, $[n - \sqrt{2n}, n + \sqrt{2n}]$. This procedure gives a range of values of the smoothing parameter and hence a series of tests can be conducted to establish the value that gives a satisfactory fit within that range (Haberman 1996).

2.6.2 Parametric Graduation

Parametric graduation seeks to establish mortality rates that are based on empirical mortality data by using models that are age dependent and whose parameters are established from the crude mortality rates. The procedure involves applying regression techniques to obtain the parameters that give the best possible fit with the least number of parameters. Some of the most used models are: Makeham model, Gompertz model, Heligman-Pollard model (H-P), generalised Gompertz-Makeham model, Logit Gompertz-

Makeham model (LGM) and Generalised Linear Models (GLM) (Renshaw, Haberman and Hatzopoulos 1997). In applying these models, the challenge is to balance the competing objectives of achieving a good fit and maintaining parsimony, that is, avoiding over-parameterisation. Over-parameterisation can give rise to solutions that are not unique and this affects the model's predictive capability. The Heligman-Pollard model has eight or nine parameter which often means that the model is over-parameterised. In addition, complex models sometimes fail to converge on a solution when the iterative procedures are used to fit the curve when the starting values are not close to the final parameter estimates (Debon *et al.* 2005).

2.6.3 Coale and Kisker Method

Direct calculation of mortality rates for the old ages faces serious limitations, which are a result of two key challenges. Very few people survive to ages above 85 and this leads to high random fluctuations in the mortality rates. The second challenge is related to age misreporting, which is a result of either digit preference (age heaping) or age exaggeration. Age exaggeration is particularly difficult to correct and will also lead to underestimation of the true level of mortality. Hence the need to employ extrapolation techniques to estimate mortality rates for the old ages (Wilmoth 1995).

Coale and Kisker (1989) proposed a method for estimating mortality rates for old ages based on two assumptions. The first is that the rate of change of age specific mortality declines linearly after the age of 85. The second is that the central rates of mortality for males and females at age 110 are 1 and 0.8 respectively.

The age specific rate of mortality change at age x, k_x, is calculated by:

$$k_x = \ln m_x - \ln m_{x-1} .$$

This equation can be re-arranged to make m_x the subject:

$m_x = \exp(k_x + \ln m_{x-1}) = m_{x-1} \exp(k_x)$. This is a recursive algorithm and it can be shown that mortality rates for older ages can be calculated by: $m_{x+n} = m_{x-1} \exp(\sum_{x}^{x+n} k_y)$.

Assuming that the rate of change in mortality is linear above age 85 then:

$$k_x = k_{85} + (x - 85)g$$

where g is the slope in the change in mortality above age 85 and it follows that for any age x above age 85:

$$m_x = m_{84}\exp(\sum_{85}^{x} k_{85} + (y-85)g).$$

In practice the slope, g, is estimated by assuming that m_{110}=1, which leads to the following formula: $g = [\log(\frac{m_{84}}{m_{110}}) + 26k_{85}]/325$.

The value of k_{85} is estimated by the geometric average rate of increase in mortality from the 7 years of age preceding the age of 88, such that:

$$k_{85} = \ln(\frac{m_{88}}{m_{81}})/7.$$

In a review of the Coale and Kisker method, Wilmoth (1995) verified the plausibility of the first assumption that k_x is linear but was not able to verify the second assumption that the values of m_{110} for males and females are equal to 1 and 0.8 respectively. He proposed a method of estimating mortality above age 85 without fixing the central mortality rates at age 110. However, the method requires accurate data, which are generally not available in developing countries. In the same review, the conclusion was drawn that the assumed value of m_{110} has minimal effect on the mortality rates produced by the model (Wilmoth 1995).

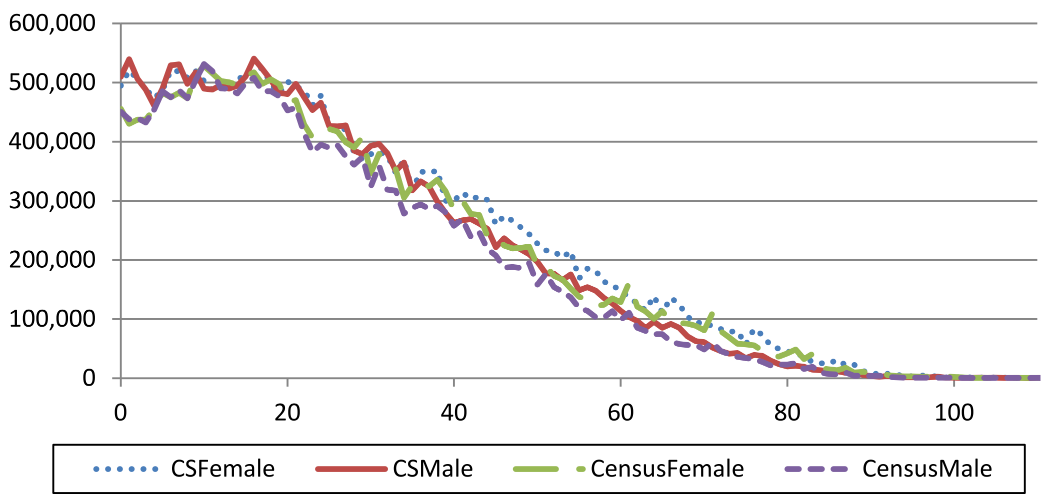
3.1 Data Sources

3.1.1 2001 Census and 2007 Community Survey Data

This research uses data from the national census conducted in 2001 in South Africa and the nationally representative Community Survey conducted in 2007. The national census was conducted on the night of 9-10 October 2001 and the Community Survey can be deemed to have been conducted on the night of 14-15 February 2007 (Machemedze 2009).

The 2007 Community Survey sampled 947 331 people in 250 348 households and collected household information similar to that in the census, namely: population group, age distribution, fertility, mortality, migration, and information that can be used to estimate the national population. The 2001 national census data are available in the form of a 10 per cent sample of the total units enumerated. Both the national census data and the Community Survey data were sourced from Data First database in Stata format.

Figure 3.1 Age distribution of the population estimates from the 2001 Census and the 2007 Community Survey (CS), weighted to reflect the total population



The age distributions of the population estimates from the 2001 census and the 2007 Community Survey show that there was a substantial undercount of children under the age of 10 in the 2001 census. As shown in Figure 3.1 the number of persons in the 0-10 year age group are markedly lower in the 2001 census when compared with the 2007 Community survey. It was also noted that the population estimates from the 2007 Community Survey appears to have overstated the number of women in the age groups above age 54 and hence a decision was made to re-estimate the numbers by adding two thirds of the estimates from the 2007 Community Survey to a third of the estimates derived from the ASSA2008 model for the same period.

3.1.2 Vital Registration Data

In South Africa, the process of registering deaths is administered by the Department of Home Affairs. When a person dies, a death notification form is completed and a death certificate issued. The death notification form contains information about, *inter alia*, the deceased's age at last birthday, sex, population group, date of birth and date of death. The Department of Home Affairs first registers all deaths of people on the population register, recording the date of death and whether the cause of death is natural, unnatural or unknown and then provides copies of these death notification forms to Statistics South Africa for processing. Statistics South Africa compiles the information from the death notification forms and produces annual cause of death mortality reports under the series entitled, 'Mortality and causes of death in South Africa' (Statistics South Africa 2010). The process of death registration is an ongoing process as deaths are not always registered in the year of death. For this reason, the mortality reports are produced by Statistics South Africa for the year two years prior to the date of the report. However, not all deaths are registered, particularly in the rural areas and among children (Statistics South Africa 2009).

Ordinarily the reports and data made available by Statistics South Africa to the public do not indicate population group and hence a special request was made to obtain death data at unit record level. The data were provided in Stata files (Professor Rob Dorrington personal communication). About 26 per cent of the records for the period 2006-2008 do not indicate the population group of the deceased (see Table 3.1).

Table 3.1 The distribution of deaths by population group in the period 2006-2008

Population Group	Frequency	Percentage
Black	1,158,227	62.6%
White	108,238	5.9%
Indian/Asian	23,670	1.3%
Coloured	79,355	4.3%
Missing	481,689	26.0%
Total	1,851,179	100%

3.1.3 Mid-Year Population

This research makes use of mid-year population estimates for 2006, 2007 and 2008 to calculate central rates of mortality. The United States Census Bureau (USCB), United Nations Population Division (UNPD), Statistics South Africa and the Actuarial Society of South Africa (ASSA) independently produce annual mid-year estimates of the population in South Africa. Unlike the other agencies, ASSA produces mid-year estimates both in single ages and for the national population and all the population groups.

Statistics South Africa releases the mid-year estimates in 5-year age groups and hence one has to interpolate to convert to single year age groups. There are several ways of interpolating but Statistics South Africa recommend the use of Sprague interpolation coefficients to estimate the single-year population estimates (Statistics South Africa 2007b).

3.2 Imputation of vital registration data

The process of imputing the 478,460 unit records without population group was conducted in Stata using the hotdeck imputation method proposed by Stata Corp (2009). This method identifies unit records with similar values to the one with the missing variable for certain pre-determined variables (predictor covariates) and then takes the missing attribute from the records that match as the assumed or imputed value.

The first stage involved analysis of data before imputation to identify predictor covariates that can be used in the imputation process. Variables with high non-response were excluded from the imputation process. The second stage involved the actual imputation using the predictor covariates identified in the first stage. Stata uses the 'multinomial logistic imputation' (mlogit) which is suitable for imputing categorical data. It first analyses the variables for suitability and identifies variables with a high proportion of missing fields and gives the user the option to proceed or stop the process. It then proceeds to impute the missing data points in the specified variable for all the unit records with the variable missing using the predictor variables chosen by the user.

After this the data were investigated to ascertain the plausibility of the imputed result. Population distribution by age and population group were examined and compared with that before imputing. In addition, the levels of completeness for the different population groups after imputation were estimated and subjected to reasonability checks. The code for this imputation process is listed in Appendix A3.

3.3 Completeness

The General Growth Balance Method and the Synthetic Extinct Generations methods were applied to the data in line with the recommendation by Hill, You and Choi (2009) for populations where migration is not insignificant. The 2001 national census population estimates were used as the starting population (P_1) and the 2007 Community Survey population estimates were used as the closing population (P_2). The numbers of deaths in the inter-survey period were compiled from those enumerated by Statistics South Africa in the annual Mortality and Causes of Death in South Africa: Findings from Death Notification reports from 2001 to 2007 (Statistics South Africa 2001; 2006; 2005; 2002; 2003; 2004; 2007c) . However, in the case of population groups, the numbers of deaths were estimated using unit record data from Statistics South Africa provided by Prof Rob Dorrington (personal communication).The non-registrations, *i.e.*, deaths that remain unregistered after two years, are assumed to be independent of age and hence will not distort the age distribution of deaths exhibited by the enumerated deaths.

In the case of 2001 and 2007 the estimates of numbers of deaths included are the deaths that occurred between the night of 9 October 2001 to the end of the year and those that occurred from the beginning of the year 2007 to the night of 14 February 2007. These deaths in partial years were estimated by pro-rating using the number of days in the year included in the relevant period, assuming that deaths occurred uniformly in both years.

In calculating the level of completeness for the four population groups, the 2001 census population estimates and the 2007 Community Survey population estimates for the different population groups are used as the starting (P_1) and closing (P_2) populations respectively. The numbers of deaths for each population group are estimated after imputing population group to the unit records without population group.

3.3.1 General Growth Balance Method

In applying the GGB method, the slope was estimated by using the Orthogonal Sum of Squares Regression (OR) method proposed by Bhat (2002), which establishes the slope of the straight line which minimises the vertical and horizontal residuals. The slope was estimated on the age range starting from age 5 up to 85 to compensate, in part, for net migration. The robust method introduced by the United Nations (2002) was also applied for comparison.

3.3.2 Synthetic Extinct Generations Method

In applying the SEG method, it is necessary to estimate the life expectancy at the base age of the open age interval. The ASSA2008 model estimates were used to estimate the life expectancy for the last age group. The open ended age group starts at age 85 and hence the ASSA2008 model was used to estimate the life expectancy at age 85. The level of completeness was estimated over the 5-85 age range.

3.3.3 Completeness for the 15-85 age groups for the year 2007.5

The level of completeness of death registration calculated for the 15-85 age group using the 2001 census and the 2007 Community Survey corresponds to the period between 2001 and 2007. However, it must be noted that even though the level of completeness of death registration estimated was for the 15-85 age range, the 5-85 age range was used in the actual estimation process to compensate for the absence of migration estimates. The 5-14 age range was included to compensate for the absence of net migration estimates. The level of completeness for the period of interest (2006-2008), which can be represented by the mid-point, (2007.5), was assumed to be at the same level as the average for the period 2001-2007.

3.3.4 Completeness for the under-fives and 5-14 age groups

In calculating the level of completeness for the under-fives, results of estimates of mortality rates from a study by Darikwa (2009), which covered the period 1996-2006, were used to estimate the level of completeness for the period covering 1996-2006 (see Table 3.2). However, the results did not report the levels of completeness for the age range 1-4 in single ages and hence these were estimated from the information on number of deaths and completeness for the infants and the under-fives. The completeness estimates for the infants and the 1-4 age group were extrapolated to estimate the level of completeness for the years 2007 and 2008 by applying a logistic curve. The process of extrapolating was done using the Logistic spreadsheet from PASEX spreadsheets, which were developed by the United States Bureau of Statistics (Arriaga, Johnson and Jamison 1994).

The levels of completeness for the under-fives for the different population groups were estimated using the same figures estimated for the entire population. Completeness of death registration for the 5-14 age groups was estimated by assuming a linear trend between the completeness for the 1-4 age group and that of the 15-85 age group for males and females separately.

Table 3.2 Completeness of reported deaths for the under-fives and the infants,1996-2006.

	Infants		1-4		Under-Fives	
Year	Male	Female	Male	Female	Male	Female
1996	43.3	43.6	48.1	49.1	44.4	44.8
1997	45.5	44.4	41.0	43.1	44.3	44.1
1998	45.9	47.7	44.3	41.1	45.5	45.9
1999	49.2	49.0	41.8	45.5	47.0	48.0
2000	53.8	52.4	44.0	47.4	50.8	51.0
2001	59.6	57.7	41.5	44.6	53.2	53.4
2002	65.8	65.0	45.5	45.8	59.0	58.7
2003	73.9	74.0	46.2	47.4	63.9	64.7
2004	80.6	81.0	50.6	51.3	69.3	69.6
2005	87.8	86.4	53.9	54.5	75.7	75.4
2006	92.3	91.7	57.9	58.3	80.0	80.1

Source: Table Reconstructed from Darikwa (2009): Table 4.4 page 46, Table A8 page 68

3.4 Central rates of mortality

In order to calculate the age specific crude central rates of mortality, an estimate of the true number of deaths and an estimate of the 'central exposed to risk of death' are needed. The deaths as recorded in the death notification system are adjusted to account for the deaths that are not recorded by using the level of completeness estimated in section 3.3. The mid-year population estimates are used as the estimates of the annual 'central exposed to risk of death' for the different years. The crude central rates of mortality are estimated by applying the following method proposed by the Government Actuaries Department (2005):

$$m^o{}_x = \frac{(d_{2006,x} + d_{2007,x} + d_{2008,x})}{(P_{2006.5,x} + P_{2007.5,x} + P_{2008.5,x})} = \mu^o{}_{x+\frac{1}{2}}.$$

The method assumes that $P_{t+0.5,x}$ is a linear function of *t* for each calendar year.

where:

- $d_{t,x}$ represents the aggregate number of deaths (after adjustment) of people aged x last birthday in year t.
- $P_{t+0.5,x}$ represents the mid-year population of people aged *x* last birthday in year *t*.
- $\mu^o{}_{x+1/2}$ represents the crude force of mortality for persons aged *x*+1/2.

In the case of the different population groups, the same method was applied after imputing population groups to the unit death records without population groups.

3.5 Graduation of mortality rates

3.5.1 Cubic Spline Smoothing

The cubic spline smoothing technique was implemented to graduate the crude central rates of mortality using Matlab curve fitting software, see (Appendix A1). Cubic spline smoothing takes uniformly spread knots at all ages from age 1 to age 85, creating a total of 85 knots. The cubic spline smoothing process gives a piecewise polynomial form of the cubic spline function *g*(*t*) between all the ages (knots) by minimising the function *S* given below:

$$S = \alpha \sum_{i=1}^{n} w_i \{x_i - g(t_i)\}^2 + (1-\alpha) \int_{a}^{b} \{g''(y)\}^2 dy .$$

This function allows for the allocation of weights to the data at individual ages and a smoothing parameter α. Weights were estimated by finding the reciprocal of the variance of each data point:

$W_x = \dfrac{1}{\mathrm{var}(D_x)}$. The weights, W_x, were scaled up by dividing individual weights by the sum of the weights to produce w_x.

The smoothing parameter, α, represents the trade-off between having a very smooth spline and having a fitted curve that is very close to the empirical data. When the smoothing parameter, α=0, the fitted curve will be a least-squares straight-line and when the smoothing parameter, α=1, the fitted curve will be a natural cubic spline interpolant (Mathworks 2010). Matlab allows the graduator to adjust manually the smoothing parameter until a desired fit is achieved. In this research, a fit is considered acceptable if it exhibits a smooth progression of mortality rates over the entire age range on a logarithmic scale and follows the profile of the empirical data closely. Further, the value of the smoothing parameter should lie in the interval estimated using Reinsch's algorithm (Haberman 1996). Visual inspection method was used to inspect the curve to ensure that there are no sudden breaks and bumps as the rates progress from age to age.

3.6 Graduation Tests

The graduated mortality rates were tested to check for smoothness and goodness-of-fit. Smoothness tests were applied to ascertain if there is a gradual progression of the fitted mortality rates and that there are no rapid fluctuations in the mortality rates by age. This was done by visual inspection of the curves and an analysis of the third differences.

Goodness-of-fit tests are applied to check whether the fitted mortality rates are close enough to the crude estimates of the mortality rates. The following tests were applied: sign test, Steven's grouping of signs test, chi-squared test, standardised deviations test, and the cumulative deviations test.

The sign test is a non-parametric test that was applied to test if the graduated mortality rates do not consistently over or under-estimate the observed mortality rates for the age range 1 up to 84. The number of times that the fitted rates exceed the crude rates follows a binomial distribution. Since the sample size of 84 is greater than 30, the normal approximation to the binomial distribution was used to calculate the test statistic and to establish the critical region.

The Steven's grouping of signs test was applied to test whether the graduated mortality rates do not over or underestimate the observed rates for large spans of ages by checking the number of times that there is a switch from overestimation to underestimation of the crude rates by the fitted rates and vice versa.

In applying the standardised deviations test, the cumulative deviations test and the chi-squared test, the variance introduced by adjusting for incompleteness was taken into account by simulating the levels of completeness and the numbers of deaths to give 'random completeness' and 'random deaths'. This was done by assuming that the level of completeness follows a uniform distribution and that the levels for the different ages are +/-5 percentage points from the calculated level in absolute terms and calculating the random levels in excel between the lower bound and the upper bound. The simulated random number of deaths were calculated by assuming that the number of deaths follow the normal distribution as an approximation of the Poisson distribution. This allowed for the adjusted number of deaths for each age to be recalculated using the following formula:

$$Adjusted\_random\_deaths = \frac{RandomDeaths * Completeness}{RandomCompleteness} \ .$$

It is noted that the distribution of the adjusted random deaths is a convolution of the product of the uniformly distributed completeness level ($f(v) = \frac{1}{b-a}$) and the normally distributed random deaths ($f(y)=\frac{1}{\sigma\sqrt{2\pi}}\exp-\frac{(y-\mu)^2}{2\sigma^2}$). The result is a distribution which is not a normal distribution and it can be shown that the distribution of the new composite variable can be reduced to the following assuming that the level of completeness and the number of deaths are independent:

$f_w(w)=\frac{1}{b-a}\frac{1}{\sigma_y\sqrt{2\pi}}\int_a^b \exp(-\frac{1}{2}(\frac{y-\mu_y}{\sigma})^2).\frac{1}{v}dv$ (Professor MacDonald personal communication).

The variance of the product can be shown to be equal to the following by applying a result proposed by Rohtgi(1976) in Glen Leemis and Drew (2003) :

$$Var(\frac{D'_x}{P_x})=\frac{1}{P_x^2}V(D'_x)=\frac{12\sigma_y{}^2+(b-a)^2(\mu_y{}^2+\sigma_y{}^2)}{12P_x^2},$$

where D'_x stands for the random 'true' deaths of persons aged x last birthday, P_x stands for the number of persons aged x last birthday, variable Y stands for the number of deaths and variable V stands for the level of completeness (see Appendix A3).

The chi-squared test was applied to test the extent of the deviation of the graduated rates from the crude rate. This procedure was done for male and female populations at national and population group levels with 83 degrees of freedom at the 5 per cent level of significance.

The standardised deviations test was applied to all the data points for the national populations and the different population groups to ascertain if the distribution of the standardised deviations follows the shape expected under the distribution from the simulation of the product of the uniform distribution and normal distribution. This was done by establishing the distribution of the standardised deviations in the ranges ($-\infty$,-1), (-1,0), (0,1) and (1,∞) and comparing the frequencies with the expected frequencies under the simulated distribution by applying the chi-squared test of goodness-of-fit test with 3 degrees of freedom at the 5 per cent level of significance. The intervals were chosen in such a way that the expected frequency was greater than 5 for all the categories.

The cumulative deviations test was applied to the same data as above and unlike in the case of the standardised deviations test, which was applied for individual data points, this test was applied to check the cumulative bias in the curve by summing the deviations from age 1 up to age 84 for the national population and all the population groups.

3.7 Extraplation

3.7.1 Coale and Kisker Method

The Coale and Kisker (C-K) method can be used to extrapolate central rates of mortality above the age of 85. In this research, this method was applied to extrapolate the mortality rates to age 110 for males and 115 for females by assuming that the central rates of mortality at the age of 110, m_{110} is 1 for males and 0.8 for females. The value of k_{85} was calculated using the formula: $k_{85} = \ln\left(\frac{m_{88}}{m_{81}}\right)/7$. The slope, g, of the change in mortality rates between the ages 84 to 110 was calculated using the formula: $g = -[\ln(\frac{m_{84}}{m_{110}}) + 26k_{85}]/325$.

Coale and Kisker (1989) showed that the central rates of mortality above the age of 84 can be calculated by applying the formula: $m_x = m_{84}\exp(\sum_{y=85}^{110}(k_{85} + (y-85)g)$. Wilmoth (1995) showed that the same equation can be simplified to:

$m_x = m_{84}\exp[(x-84)k_{85} + \frac{(x-84)(x-85)}{2}g]$. The simplified version of the formula was used to calculate the central rates of mortality above the age 84. The value of m_{84} was estimated by calculating the average of the central rates of mortality from age 82 to 86 to ensure a smooth blending of the extrapolated rates and the rates from 0 to 84 Coale and Kisker (1989).

3.7.2 Blending

When the extrapolated rates are joined to the graduated rates, a kink or a sudden and rapid change in rates may occur. The process of blending of different series of mortality rates was conducted to join smoothly the rates from the Cubic Spline Smoothing process to the Coale and Kisker rates.

The blending process was conducted using a function, which allows the adjustment of one or both series by introducing two continuous functions to the blending function:

$U_{r+t} = K_{r+t}U^a_{r+t} + (1-K_{r+t})U^b_{r+t}$, where:

$$K_r = 1$$
$$K_s = 0.$$

U^a and U^b are the two series to be joined and U is the blended function. K_{r+t} is a continuous function which is estimated over the overlapping range from t=0 onwards, r+1 represents the extrapolation age such that K_r=1 and K_s=0, s-1 represents the last age in the overlapping range such that K_s=0 (Benjamin and Pollard 1993).

The function K_{r+t} was estimated by applying a method referred to in Benjamin and Pollard (1993) as 'The curve of sines':

$K_{r+t} = \frac{1}{2}(1+\cos\frac{t\pi}{\text{s - r}})$ (Benjamin and Pollard 1993).

The success of the blending process was tested by calculating and inspecting the third differences from age 80 to 90 and by visually inspecting the curves in the relevant interval.

3.8 Calculating q_0

Mortality rates fall rapidly with age in the first few years of life. Hence, deaths cannot be assumed to occur uniformly in the first year of life and hence an approximation of the actual age at death for infants that die before reaching the first birthday must be used to calculate q_0. The Government Actuaries Department (1995) proposed calculating q_0 by applying the following formula: $q_0 = \frac{m_0{}^o}{1+(1-\phi)m_0{}^o}$, where $m_0{}^o$ represents the crude central rate of mortality for age 0 last birthday, ϕ represents for the average age at death for all infant deaths. The value of ϕ was calculated by taking the actual number of hours lived by infants that died in aged 0 and calculating the average for all the cases. The calculations were done for the males and females for the national data and for all the population groups. The records without an indication of sex were not included in this analysis. The data made available by Statistics South Africa included still births and hence measures were taken to exclude the still births from the above calculations.

3.9 Life Table Construction

The graduated rates from the cubic spline smoothing output are in the form of the force of mortality at age x+0.5 ($\mu_{x+0.5}$) and hence must be converted into mortality probabilities at age x. The conversion process was conducted for all the ages from age 1 to 110 for both males and females by applying the following:

$$q_x = 1 - \exp\left[-\int_{-0.5}^{0.5}\mu_{x+0.5+t}\,dt\right].$$

Age 0 was not included in this conversion as q_0 was estimated separately in another section. Further, when applying the cubic spline smoothing technique, the first data point related to age 0 is not included in the graduation process and hence q_1 was estimated separately by assuming that l_x is quadratic over the age interval (1, 3) and hence the following equation was applied:

$$q_1 = m_1 \left[\frac{1 + \frac{1}{2} m_2}{1 + \frac{1}{12}(7m_1 + 5m_2) + \frac{1}{3} m_1 m_2} \right]$$ (Government Actuaries Department 2005).

The rest of the mortality probabilities were estimated by numerical integration with 4 divisions over age range from 2 up to 109. Simpson's rule was applied over all the unit intervals and it can be shown that with 4 divisions the rule reduces to:

$$q_x = \frac{1}{8}[\mu_x + 4\mu_{x+0.25} + 2\mu_{x+0.5} + 4\mu_{x+0.75} + \mu_{x+1}] \text{ for } x=2,3,\ldots\ldots$$

4 DATA ANALYSIS AND RESULTS

4.1 Imputation

The population group variable was imputed to all the unit record data without this variable. A total of 470,728 records, representing 26 per cent of the total, had no indication of the population group of the deceased (see Table 4.1). These included deaths classified as unspecified, unknown and other. The predictor covariates that were used were ascertainment (method of ascertaining the cause of death), sex, age, province of death, underlying cause of death (natural or unnatural) and institution of death (place of death). Other variables that would have been included as predictor covariates were excluded because of high rates of non-response; these include province of birth, province of residence, smoking habits and level of education.

Table 4.1 Population distribution by population group before and after imputation

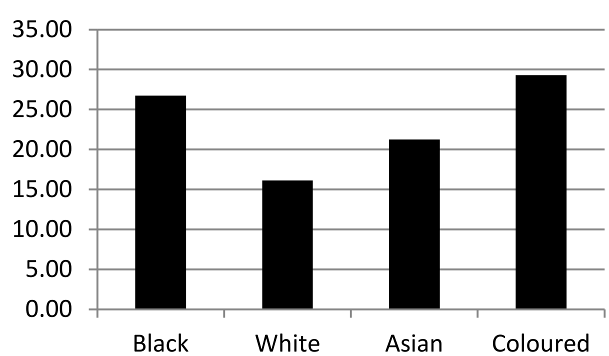
Before Imputation

Pop Group	Frequency	Proportion
Black	1,128,845	62.5
White	107,583	6,0
Indian/Asian	23,296	1.3
Coloured	76,891	4.3
Missing	470,728	26.0
Total	1,851,179	100

After Imputation

Pop Group	Frequency	Proportion
Black	1,540,734	85.2
White	128,275	7,1
Indian/Asian	29,583	1,6
Coloured	108,751	6,0
Total	1,851,179	100.00

Figure 4.1 Proportion of imputed units per population group (%)



After imputing the population group to those records where it was missing, the numbers of deaths increased for all population groups but not proportionally. It appears that, relatively, Coloureds were more likely to have population group missing than any of the other population groups. The proportion of Coloureds increased more by 41.4 per cent, followed by Blacks at 36.5 per cent, Indians/Asians at 27.0 per cent and Whites at 19.2 per cent after imputation. This is shown in Figure 4.1, which shows that Coloureds have the highest proportion with missing population group at 29.3 per cent followed by Blacks at 26.3 per cent, Asians at 21.25 per cent and Whites at 16.3 per cent. However, the levels of completeness estimated for the different population groups after imputation give unexpected rankings of the population groups in terms of mortality rates as explained in section 4.2.

4.2 Completeness of death registration

4.2.1 Under-five

The death distribution methods that were developed to estimate levels of completeness of death registration cannot be used to estimate the level of completeness of the under-fives. In this research the levels of completeness were estimated by reconstructing the completeness implied by mortality rates from a study by Darikwa (2009) covering the period 1996-2006. A logistic curve was fitted to the results and then extrapolated to the years 2007 and 2008. The averages of the years 2006-2008 were calculated and used as estimates of completeness for both males and females. In all the cases the infants were dealt with separately from the 1-4 age group. In the absence of population group specific data, these estimates of completeness were initially also used as estimates of levels of completeness for all the population groups. Table 4.2 shows the estimated levels of completeness for the infants and the 1-4 age group.

Table 4.2 Estimates of levels of completeness for the under-fives

	Infants		1-4	
Year	**Male**	**Female**	**Male**	**Female**
2006	88.5	87.7	52.7	53.7
2007	91.0	90.3	53.9	54.8
2008	93.0	92.4	55.1	55.9
Average	90.8	90.1	53.9	54.8

4.2.2 Application of the GGB and SEG methods

In estimating the level of completeness for the 15-84 year age group, population estimates from the 2001 national census and the 2007 Community Survey were used. The numbers of deaths in the inter-survey period as captured by the vital registration system were used. Both the population estimates and the deaths were collated into five-year age groups. The General Growth Balance (GGB) method and the Synthetic Extinct Generations (SEG) methods were applied to the data over different age ranges between age 5 and 90. The age range considered in estimating the completeness for the 15-85 age range was stretched to include ages below 15 to reduce the impact of ignoring migration. In this case, all the points in the diagnostic plots as shown in Figure 4.2 lie close to the straight-lines (male and female) and hence changing the range over which the slope is calculated would not have a significant impact on the completeness estimates. In the case of the population groups, the points in the scatter plots are close to straight lines for all the population groups with the exception of Coloured and Indian/Asian males (see Appendix A12).

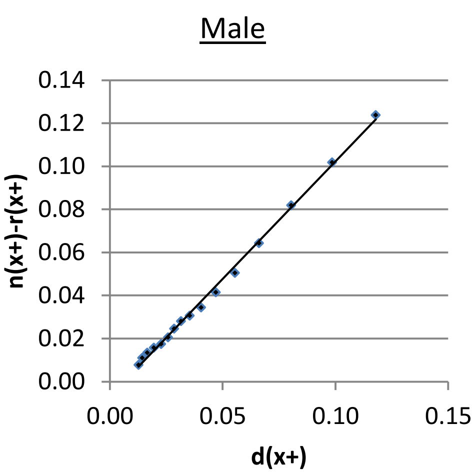
In the case of the GGB method, two methods were used to estimate the slopes: the Orthogonal Regression Sum of Squares method proposed by Bhat (2002) and the 'Robust line' method proposed by the United Nations (2002). Figure 4.2 below shows the scatter-plots used to estimate levels of completeness for both males and females for the period from the 2001 National Census to the 2007 Community Survey. Estimates of completeness were calculated over different age ranges starting between ages 5 to 15 and ending at ages 60 to 90 and found to be broadly consistent.

In the cases of both the robust line and ORSS methods, the estimated level of completeness was very high (above 100 per cent) when the age range ended at 75 and below, decreased progressively above 75 and stabilised when the ranges extended to 85 and above. Given the unavailability of reliable migration data in this research, the estimated level of completeness was estimated over the age range 5 to 85 at 92 per cent for males and 89 per cent for females relative to the 2001 Census and 2007 Community Survey population estimates after compensating for differential coverage of the two population estimates.

When the ORSS method was applied on the same data, the level of completeness was estimated at 91 per cent for males and 87 per cent for females relative to the 2001 census and 2007 Community Survey population estimates. The results from the 'Robust line' technique suggest that there was differential coverage in the enumeration process between the two population estimates. The relative under-coverage was around 3 percent for males and around 1 percent for females in the 2001 Census relative to the 2007 Community Survey.

In applying the SEG method, life expectancy at age 85 was estimated using the ASSA2008 model for both males and females. The diagnostic plots in Figure 4.3 show that the population estimates are not without problems as the estimate of completeness fluctuates by age for both males and females. The drop off under age 25 is probably due to unaccounted for migration. The level of completeness for males was estimated at 88.26 per cent and 88.51 per cent for females relative to the 2001 Census and 2007 Community Survey population estimates. The estimates from the SEG method were calculated by estimating the median age specific completeness for the ages ranging from 5 to 85 in five year age groups. The diagnostic plots for the population groups are included in Appendix A12.

Figure 4.2 Application of the GGB method to male and female national populations



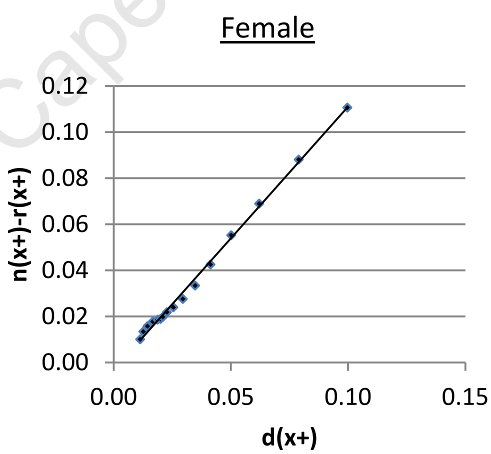


Figure 4.3 Application of the SEG method on males and females nationally

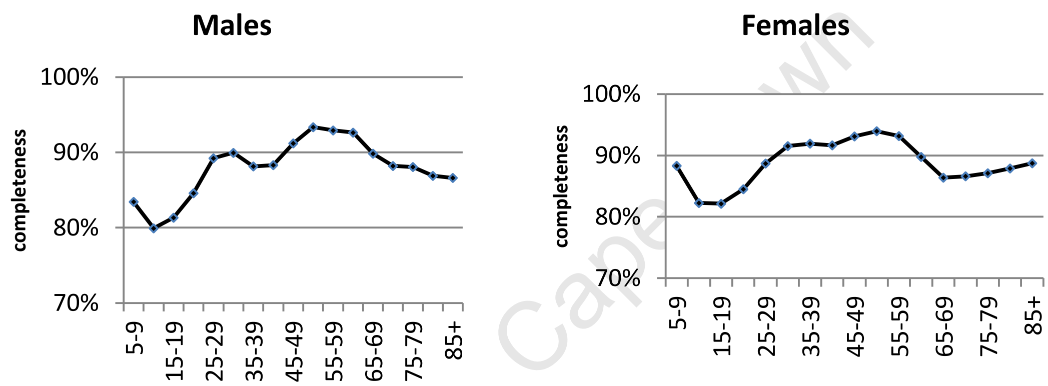


Table 4.3 GGB and SEG summary for national data

	SEG	GGB-Robust	GGB-Orthogonal	Mean
Male	88 %	92 %	91 %	90 %
Female	89 %	89 %	87 %	88 %

Table 4.4 The average of the GGB and SEG for population groups

	Blacks	Whites	Coloureds	Indians/Asians
Male	94 %	80 %	88 %	100 %
Female	89 %	82 %	93 %	100 %

The sense that one gets from the completeness estimates is that nationally for males the level is between 88 and 92 per cent, and for females between 87 and 89 per cent. These estimates were based on the age range 5 to 85 the final estimates for this research for both males and females were estimated by calculating the mean of the two methods for both males and females (see **Table 3.1**).

The same methods were applied to the data for population groups to establish estimates of completeness for the different population groups during the same period for males and females. The estimates are reasonably close to the national estimates except for the case of Whites which are implausibly lower than expected. At 80 per cent and 82 per cent for males and females respectively, the level of completeness for Whites is lower than for all the other population groups and lower than the national estimates, which were estimated at 91 percent and 88 per cent. In the case of Indians/Asians the average completeness estimates were implausibly high (well above 100 per cent) for both males and females and hence it was assumed that the vital registration for Indians/Asians is 100 per cent complete (see Table 4.4). It can be pointed out that, the fact that completeness was assessed to be above 100 per cent might suggest that the results of the imputation of records without population group cannot be relied upon.

The levels of completeness calculated above correspond to the period covered by the 2001 Census and the 2007 Community Survey whose mid-point is 2004.45. In this research the period of interest lies between 2006 and 2008 with a mid-point of 2007.5. According to Statistics South Africa (2010), the level of completeness of death registration increased between 2004 and 2006 and decreased in 2007 and 2008. A closer look at the actual numbers of deaths registered between the period 2001 and 2008 (Figure 4.4) shows that the number of deaths registered has grown by about 2 per cent per year between 2004 and 2007. If the natural increase in population is taken into account, it can be assumed that the level of completeness of death registration remained stagnant between 2004 and 2007. Therefore in this research, the level of completeness in the period 2006-2008 was assumed to be the same as that of the period between 2001 and 2007.

4.2.3 Completeness: 5-14 year age group

The levels of completeness for the ages 5-14 were estimated by assuming a linear trend between the level of completeness for the 1-4 age group and the level estimated for the 15-85 age group. The levels of completeness estimated using the above method are tabulated in Table 4.5. The same method was applied in the case of the population groups.

Figure 4.4 Total number of registered deaths since 2000

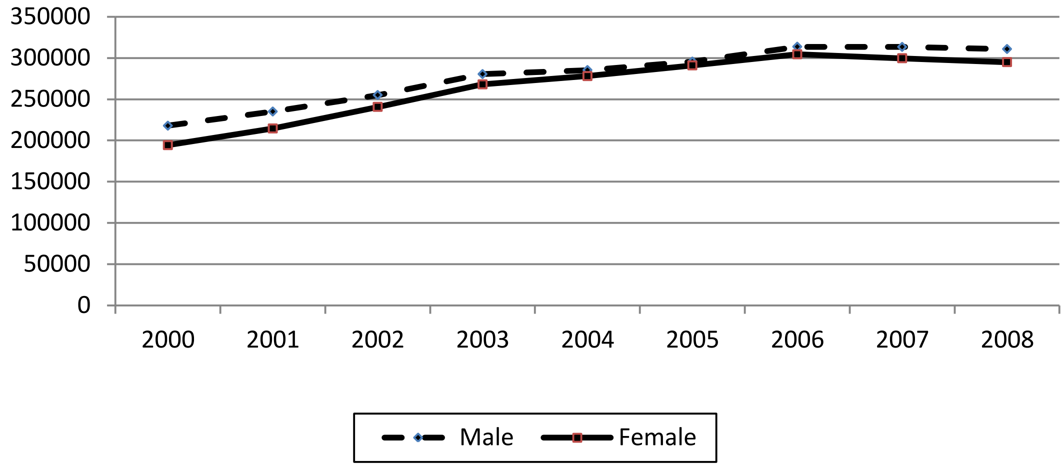


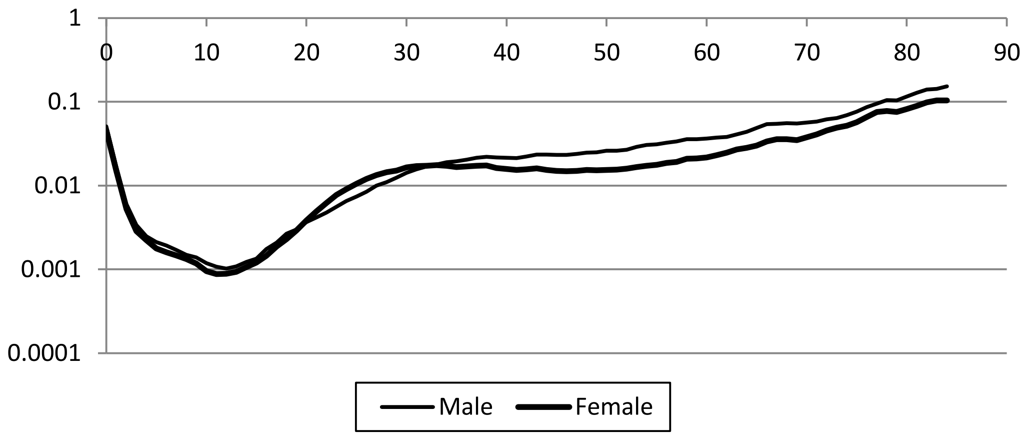
Table 4.5 Estimates of completeness for ages 5-14 for national data

Age	Male	Female
5	57	58
6	61	61
7	64	64
8	67	67
9	70	70
10	74	73
11	77	76
12	80	79
13	84	82
14	87	85

4.3 Mortality rates and graduation

The recorded deaths for the period 2006-2008 were adjusted for incompleteness using the levels of completeness established in the previous section. The adjustment was done for the ages ranging from 0 to 84. The forces of mortality were estimated by dividing the total adjusted deaths for the three years by the sum of the mid-year populations for the three years for both males and females. The mid-year population estimates from the ASSA2008 model were used as they follow the population counts closely and adjust for the undercount in the younger age groups. Care was taken not to include still births in the calculations. The curve for the crude force of mortality for males and females for national data are shown in Figure 4.5.

Figure 4.5 Crude central mortality rates for males and females nationally on a log scale**.**



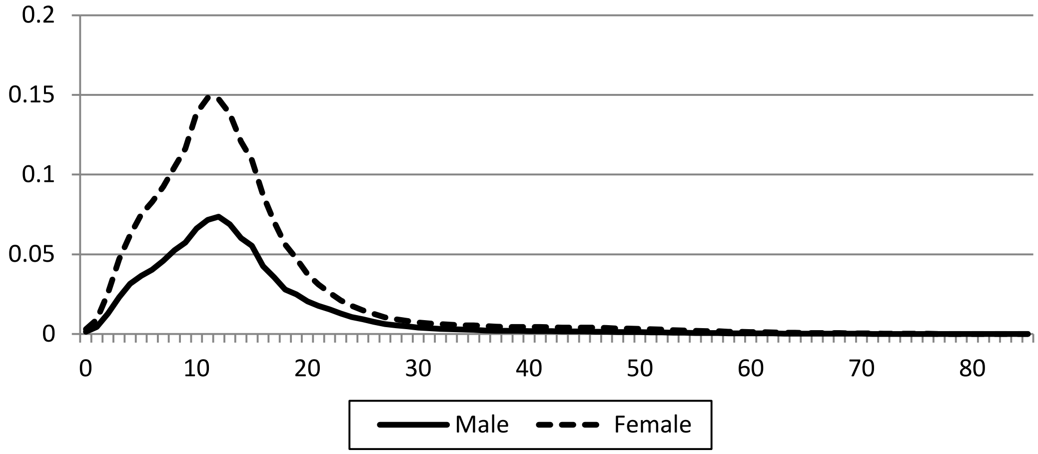
4.3.1 Extrapolation age

At very high ages the quality of data deteriorates as a result of age 'heaping' or digit preference and age exaggeration. Further, at very high ages the mortality rates tend to fluctuate rapidly by age as a result of small numbers of people who survive to those ages (Coale and Kisker 1989). Poor quality data compromises the application of smoothing techniques including the spline based methods (Marszalek 2010). Coale and Kisker (1989) recommended age 85 as the age beyond which graduation should not be applied but rather that extrapolation methods be applied. Assessment of South African old age data by Machemedze (2009) concluded that the problems of age misreporting are more prevalent after the age of 85 and therefore in this research the age 85 was the extrapolation age for national data for both males and females and also for all the population groups.

4.3.2 Weights

Age specific weights were calculated using the variance of the numbers of deaths for every age for both males and females. The weights calculated using this method were significantly bigger than the mortality rates and so had to be scaled down by dividing throughout by the sum for all ages. As shown in Figure 4.6 this gives more weight to the ages in the middle of the age spectrum than the younger and the older age groups. This reflects the level of reliability of the data in the different age groups. The data for persons aged around 10 have the highest weights and ages older than 30 have very low weights.

Figure 4.6 Scaled weights for males and females nationally**.**

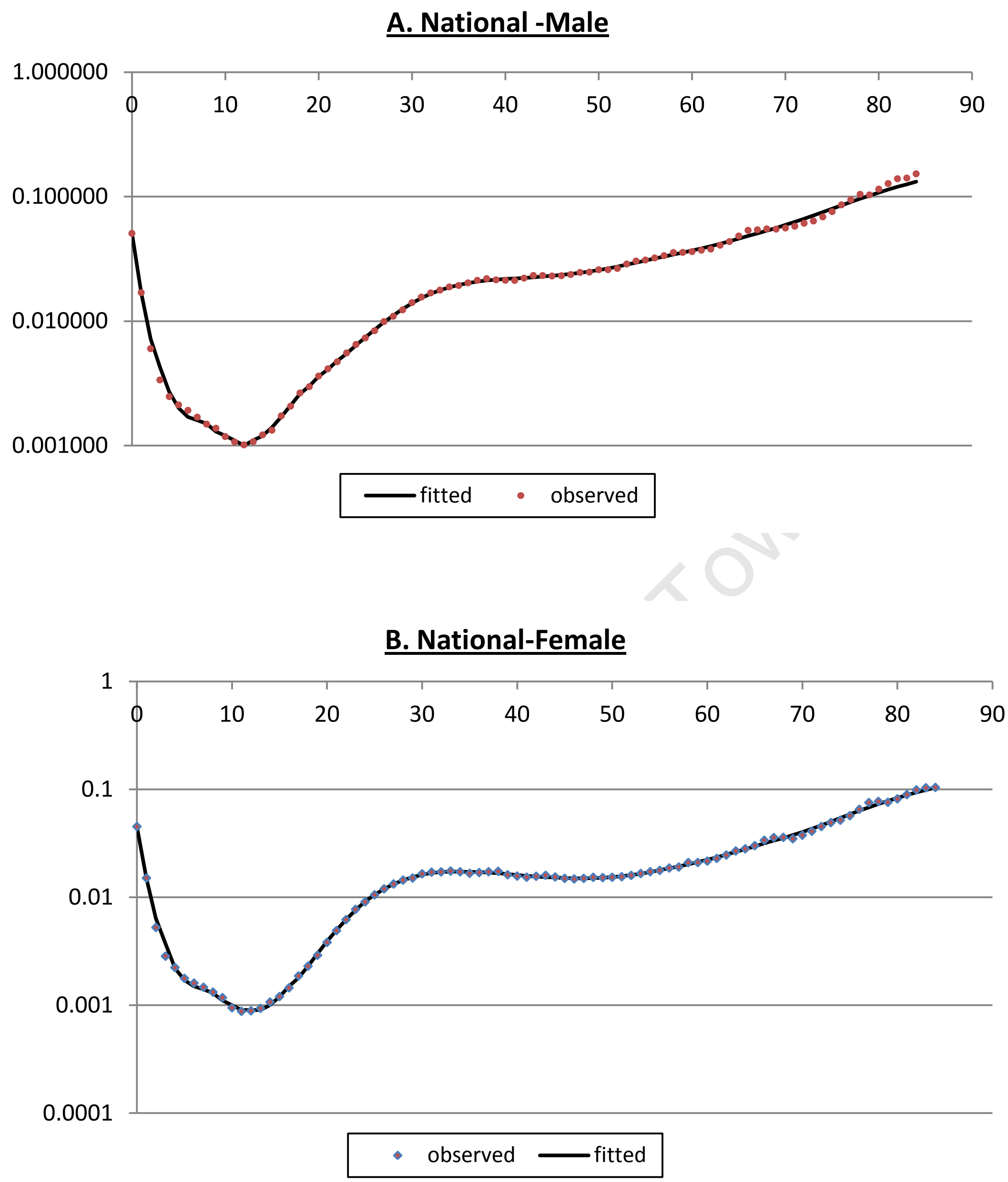


4.3.3 Graduation using the cubic splines for males and females nationally

The outputs given by the cubic spline technique in Matlab for different values of the smoothing parameter, α, were examined for goodness-of-fit and smoothness by visual inspection and by calculating the 'unsmoothness' (λ) of the graduated force of mortality. The value of the quantity that can be called 'unsmoothness' (λ) was calculated by summing the squares of the second difference of the graduated force of mortality and was estimated as 0.00000505 for males and 0.00000546 for females (Marszalek 2010).

The values of α that give a good balance between goodness-of-fit and smoothness were estimated as the average of the lower and upper estimates from the Reinsch algorithm for males and females. The smoothness parameter value was 0.98 for males and 0.99 for females (see Figure4.7).

Figure 4.7 Graduated rates and crude central mortality rates for males and females nationally on log scales.



4.3.4 Graduation using the cubic spline method for population groups

The crude mortality rates for the four population groups were graduated using the cubic spline smoothing technique. In the case of Blacks, the crude central rates of mortality were graduated by applying different smoothing parameter values. The Reinsch algorithm was applied to find an optimum smoothing parameter. When this method was applied to the entire age range (1-84) the algorithm did not converge in the appropriate interval,

, and hence the highest smoothing parameter 0.99, which gives the least smoothing, was assumed for both males and females.

The cubic spline smoothing procedure was applied to the central mortality rates for Whites using different smoothing parameter values. The Reinsch algorithm was applied over the entire age range and it converged in the interval between 0.98 and 0.99 for females and between 0.96 and 0.98 for males. The values of the mid points of the intervals were taken as the smoothing parameters. Thus the smoothing parameters assigned are 0.99 for females and 0.97 for males. However, the curves produced after graduating were visibly not smooth. The rapid fluctuations were quite visible in the younger age groups, particularly for the females. Whereas the expectation is that Reinsch algorithm must give a smoothing parameter that gives a smooth progression of force of mortality, in this case it clearly failed. The estimates of the variance calculated could be the source of the problem if the variance estimates underestimate the actual variance. However, it can be noted that smoothness improves noticeably up until $\alpha=0.9$, thereafter it appears that there is not much change in smoothness. Hence the output presented is based on $\alpha=0.9$ for both males and females. It should be noted that in the higher age groups, the smoothing process worked well as the progression is quite smooth, as evidenced by the third differences (Appendix A5) and also by inspecting the graphs visually.

The cubic spline smoothing procedure was applied to data for Coloureds for a range of smoothing parameter values. Reinsch algorithm was applied and converged on p=0.99 and 0.98 for males and females respectively. However, in the case of males the progression of the central mortality rates from the resulting output is not smooth in the younger ages. Further inspection showed that p=0.94 gives much improved smoothing and so it was used.

A similar procedure was applied to the data for Indians/Asians and the Reinsch's algorithm converged on p=0.98 and 0.99 for both males and females. However, the graduated central rates are again not satisfactorily smooth. Further inspection of progressively heavier smoothing showed that p=0.95 and 0.8 for males and females respectively offered better smoothing (see Table 4.6).

Table 4.6 Smoothing parameters for population groups

	Blacks	Whites	Coloureds	Asians
Males	0.99	0.90	0.94	0.95
Females	0.99	0.90	0.98	0.80

4.4 Graduation Test Results

Test results: National population

The results of the statistical graduation tests applied to the graduated force of mortality from the cubic spline method were acceptable for both males and females at the 5 per cent level of significance for the signs test, Steven's test and the cumulative deviations test.

Table 4.7 Test results for the national population

National		Chi-squ	Cum.Dev	Sign	Steven's	Std.Dev
National	Male	1,548.5	-1.12	39	-0.65	7.81
	Female	1,411.3	-1.27	40	-0.27	11.61
Critical Values		105.3	+-1.96	[33,51]	-1.65	7.81

In the case of the chi-squared test both graduations failed as the calculated chi-squared statistics, at 1,549 and 1,411 for males and females respectively, were substantially higher than the critical value of 105. However, the chi-squared test has limitations in that it tends to fail large population data as it is very sensitive to large data sets. Further, an inspection of the chi-squared values for individual ages shows that the younger ages had very high deviations, particularly at age two. It is also evident that in the age range from 70 to 79 the graduated rates are visibly different from the observed rates, as can be seen in Figure 4.7 (panels A and B) for both males and females as the observed rates fluctuated rapidly whereas the graduated rates followed a course that captures the trend of the change in force of mortality. Although the graduated rates deviate from the observed rates, the trend in the graduated rates is preferable.

In the case of the individual standardised deviations test, the distributions of the standardised deviations were classified into the four categories: $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$ and

(1, ∞), and a chi-squared test at the 5 per cent level of significance with three degrees of freedom was applied. The expected frequencies of standard deviations were estimated from a simulation of the 'true' random deaths. The test statistic for males at 7.81 fell inside the acceptance region as defined by the critical value of 7.81 but the females at 11.61 was in the rejection region. Although the female graduated rates failed the test, the test statistic was reasonably close to the acceptance region. Given that the graduations for both males and females passed most of the tests except the chi-squared test, which is a very sensitive test, it can be concluded that the graduated central mortality rates follow the empirical rates close enough to be accepted.

4.4.1 Test results: Black African population

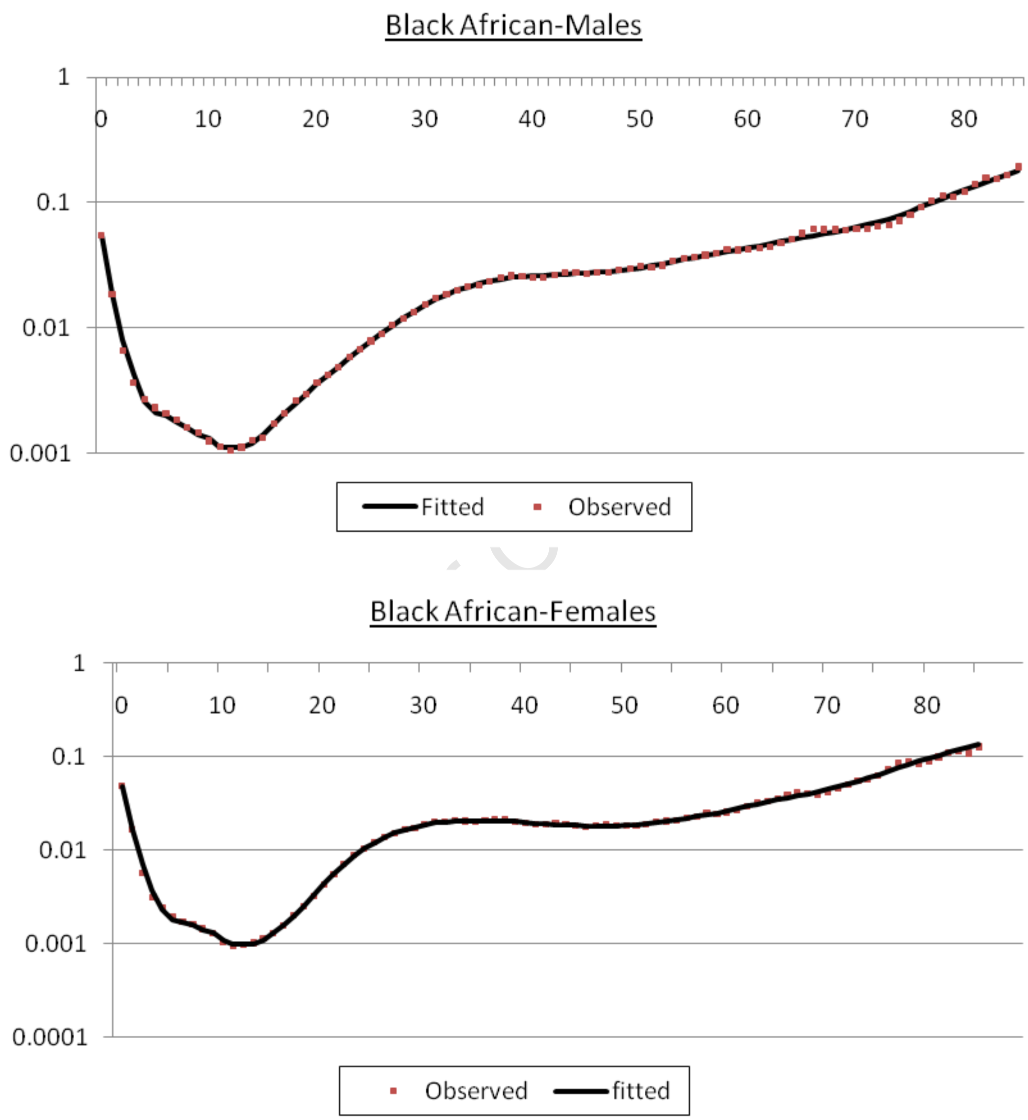
The graduated rates were satisfactory according to the Steven's test and the sign test. In the case of the Steven's test the test statistics were calculated as -0.28 and -1.02 for males and females against the critical region below -1.65. In the case of the sign test the acceptance region was such that the number of positive or negative signs had to be between 33 and 51 and the test statistics were 41 and 38 respectively.

Both the males and females failed the cumulative deviations test, the chi-squared test and the standardised deviations test. In the case of the cumulative deviations test, the removal of one data point related to age 2 lowers the value of the test statistic significantly. In the case of males the test statistic drops from -2.8 to -0.40, which falls in the acceptance region. It is evident that the cumulative deviations test was unduly affected by one data point, for both males and females.

In the case of the standardised deviations test, the standardised deviations were grouped into intervals such that every interval had an expected frequency of more than 5. A chi-squared test was applied to the data and tested against the chi-square distribution with 3 degrees of freedom at the 5 per cent level of significance. The graduated rates failed the test mainly because there were too many standardised deviations in the extreme upper and lower tails. A look at Figure 4.8 shows that the younger age groups and the older age groups contributed to the failure as they are more visible disparities between the fitted and the observed curves.

The chi-squared test was applied with 83 degrees of freedom and failed because it is sensitive to large populations. The test statistics were 1,315 and 1,275 for males and females respectively. The critical value which corresponds to 83 degrees of freedom at the 5 per cent level of significance is 105. Further, data points related to ages 2 to 5 contributed significantly to the failure of the test. However, it should be noted that the number of degrees of freedom should be lower than 83 to reflect the loss of freedom due to the graduation process and hence the critical value should have been even lower than was used.

Figure 4.8 Graduated and crude central mortality rates for black Africans on a log scale



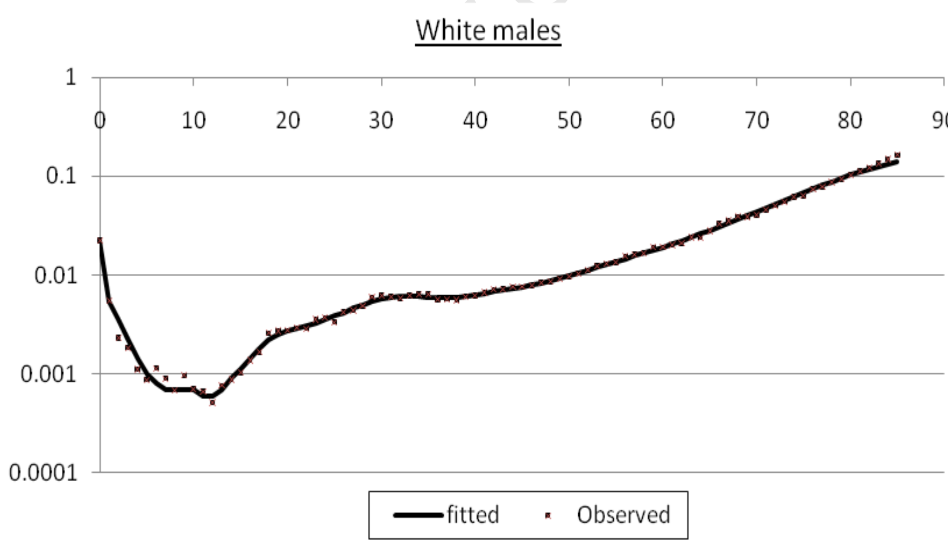
Visual inspection of the curves shows that the tests that failed, failed as a result of the rapid fluctuations in the empirical rates and given that the spline fit tends to follow the trend as opposed to the sudden changes it can be concluded that the graduated rates provide a plausible picture of the changes of the force of mortality in relation to age. (Test results listed Appendix A4).

4.4.2 Test results: White population

The goodness-of-fit test applied to the data for Whites produced satisfactory results in the case of the sign test, Steven's test and standardised deviations test for both males and females. The test statistics for the standardised deviations test were 1.79 and 1.78 for males and females respectively and they both fall in the acceptance region as the critical region is defined as being any values higher than 7.8.

The same data failed the chi-squared test for both males and females because of the sensitivity of the test when applied to large population data. In the case of the cumulative deviations test, the data for males gave a satisfactory test statistic of -1.0, whereas the females gave -3, which fell in the critical region given by a critical value of +-1.96. The pattern of the central mortality rates for females in the 70-80 age range forced the spline estimates to overestimate the observed central mortality rates (Figure 4.10). This resulted in a very high negative sum of the residuals and a test statistic outside the acceptance region. Although the test statistic lies in the rejection region, the trend shown by the graduated rate is preferable (Test results listed in Appendix A4).

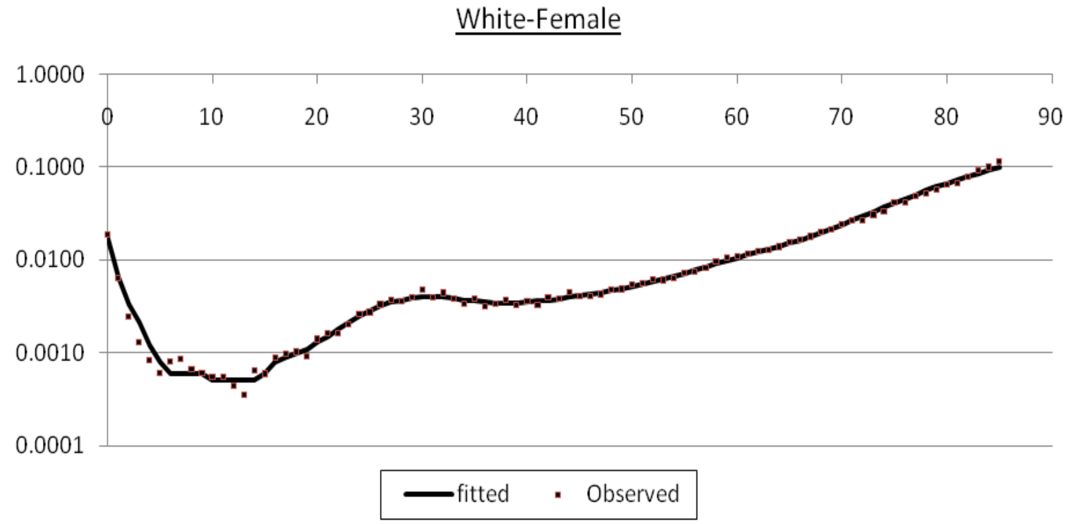
Figure 4.9 Graduated and crude central mortality rates for White males



4.4.3 Test results: Coloured population

The output for Coloured population data produced satisfactory results for the sign test, standardised deviations test, Steven's test and cumulative deviations test. However, the same output failed the chi-square goodness-of-fit test because of its sensitivity to large population data. Therefore the graduated central mortality rates can be accepted as being representative of the empirical data as they passed all the other tests (Test results listed Appendix A4).

Figure 4.10 Graduated and crude central mortality rates for White females



4.4.4 Test results: Indian/Asian population

The graduated central mortality rates gave satisfactory rates for the sign test, standardised deviations test, Steven's test and the cumulative deviations test. However, the same output failed the chi-squared test because of its sensitivity to large population data. Hence, the graduated rates can be accepted (Test results are listed in Appendix A4).

4.5 Extrapolation and blending

Blending was done by using the forces of mortality for ages 1-84 from the cubic spline output with the outputs of the Coale and Kisker extrapolation.

The Coale and Kisker method was applied to the graduated central rates of mortality. The process of blending was applied using the 'sine curve' method proposed by Benjamin and Pollard (1993) in the 85 to 90 age range. The quality of the blending was measured by analysis of the progression third differences and by inspecting visually over the range in which the two series join. In the case of national data, the analysis of progression of third differences was done from age 81 to age 89 for both males and females.

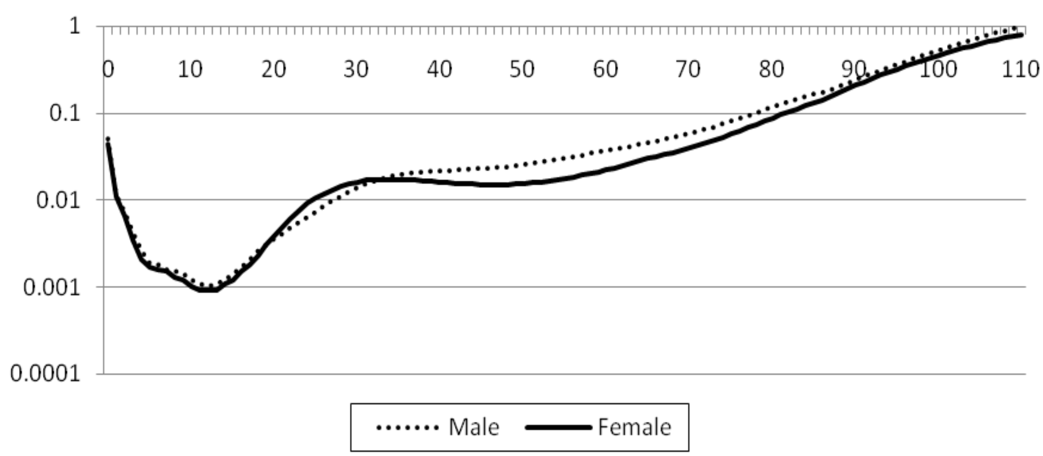
The blending process was considered successful, as the sizes of the third differences over the blending range, 81 to 89 were very small as compared to the actual central mortality rates (see Appendix A5). The Coale-Kisker blended output keeps the central mortality rates for males higher than for females as is generally expected (see Figure 4.11). An analysis of the third differences of the graduated central rates of mortality for all the population groups did not show any increased variation with increasing age including the ages close to and up to age 85.

4.6 Mortality Rates

The national life tables were checked for internal consistency by calculating and comparing them with the weighted sum of the mortality rates of the population groups for differences. This was done by calculating the expected number of deaths by multiplying the mortality rates by the 'central exposed to risk of death' estimates for both the national mortality rates and the sum of the weighted mortality rates. The chi-squared difference test yielded test statistics of 1,352 and 824 for males and females respectively and would therefore fail the test given that the critical region was defined by a critical value of 106 from 84 degrees of freedom at the 5 per cent level of significance. However, the curves of mortality rates for the national population fall in between the curves of the population groups for both males and females, as shown in

Figure **4.13** and Figure 4.14

Figure 4.11 Blended central mortality rates: Coale-Kisker output for national population



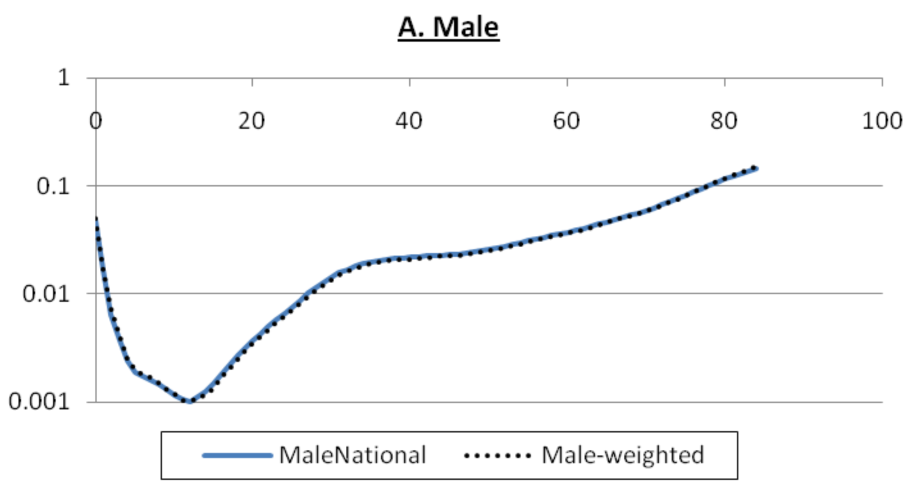
. Further, the comparison of the national mortality rates and the weighted mortality rates on the same axes show consistency in the cases of both males and females (see

Figure **4.12**). The differences in the q_x estimates of the national and the sum of weighted mortality rates were very small and averaged 3 per cent for males and 2 per cent for females. Given that the chi-squared test is sensitive to large population data, it can be concluded that the mortality rates are internally consistent.

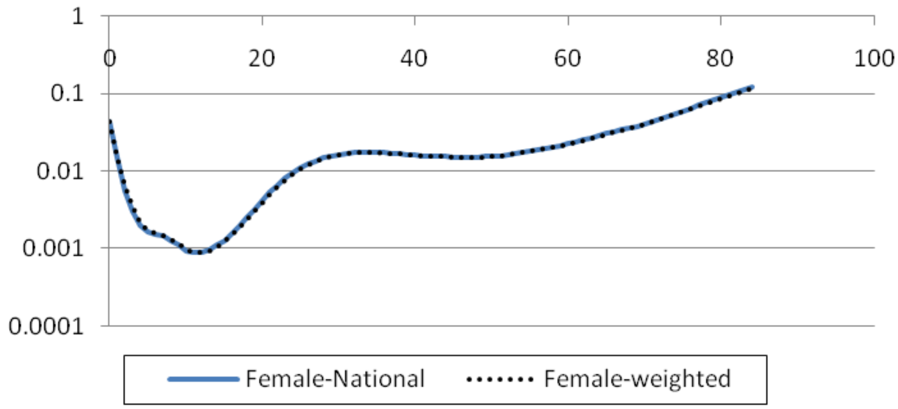
However, the curves produced by the graduated mortality rates for the smaller population groups are irregular in the younger ages. Visual inspection of the curves for White males, White females, Indian/Asian males and Indian/Asian females shows that there are rapid fluctuations in the younger ages. The bigger populations appear to have visually acceptable curves with lighter fluctuations in the younger age groups when compared with the smaller populations.

A look at the profiles of the male and female mortality curves shows that the 'hump' which other demographers call the 'AIDS' hump is more pronounced in the black males and females than in other population groups and the effect of the black population on the national curves is evident in that the 'accident' hump is also visibly more pronounced than in the population groups. When one compares the males and the females it is evident by visual inspection that the AIDS hump is more pronounced in the females than in males. The male AIDS hump is spread over a larger age span than the female hump.

Figure 4.12 Comparison of national mortality rates and the weighted mortality rates



B. Female



The calculated infant mortality rates for Whites and Coloureds were heavily affected by imputation of the unassigned deaths. The differences in scale of numbers between the different population groups at age 0 caused problems after imputing, as the numbers of deaths increased by widely differing margins.

The number of African infants who died aged 0 increased after imputation by 35 per cent as compared to 147 per cent for Whites, 95 per cent for Indians/Asians and 40 percent for Coloureds. This resulted in implausibly high infant mortality rates for Whites and Indians/Asians and therefore the infant mortality rates presented for Whites and Indians/Asians are averages from the data with imputed units and the data without the imputed units. The numbers of deaths for the ages 1 and 2 were also affected but not significantly.

Figure 4.13 Male mortality rates curves on a log scale

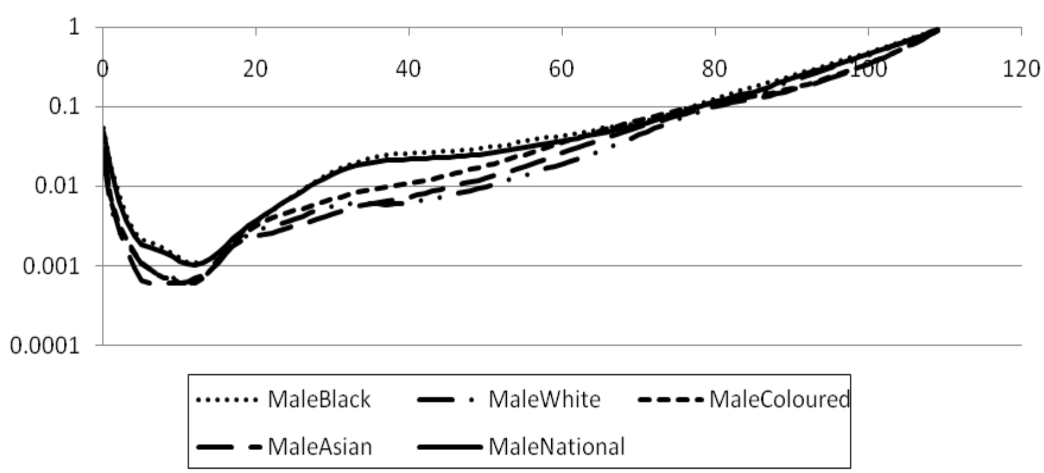
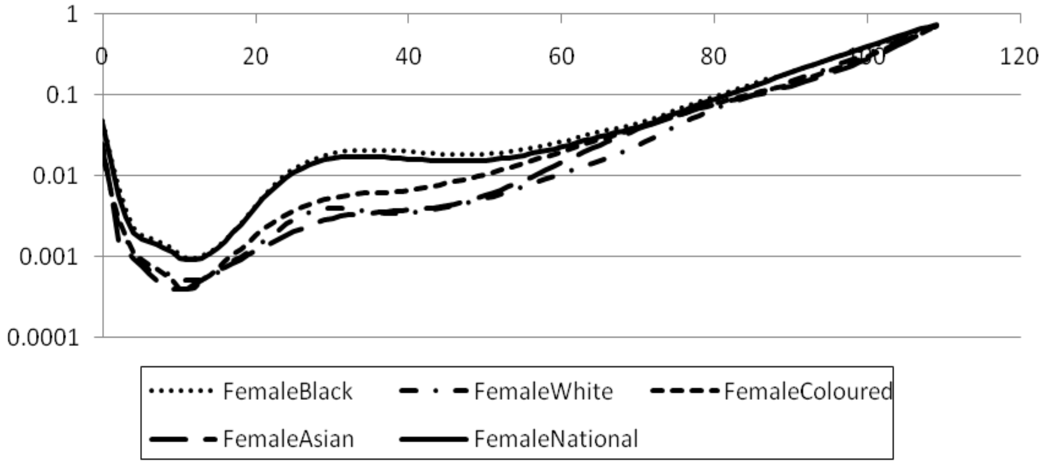


Figure 4.14 Female mortality rates curves on a log scale



Hence, the numbers of deaths for ages 1-4 were not adjusted. However, the uncertainty around the imputation process affects the credibility of various mortality indicators like the infant mortality rate, under-five mortality rate, adult mortality rates and life expectancies for the population groups. Further, the adjustments alter the implied levels of completeness for infant populations for both Whites and Indians/Asians, as shown in Table 4.8. The implied completeness estimate for the White infants are higher than 100 per cent while those for the Indians/Asian have hardly changed. This raises further questions on the reasonability of the results of the imputation process with regards to the infants. Clearly some of the infants with missing units were wrongly allocated to the White population. This has the impact of over-estimating the infant mortality rates for the Whites while under-estimating, slightly, the same for the other population groups.

Table 4.8 Implied completeness after adjustments for White and Indian/Asian infants

	White	Indian/Asian
Male	132%	91%
Female	132%	91%

The infant mortality rates for the country as a whole at 49 and 44 deaths per one thousand infants for males and females respectively are lower than the estimated rates for Blacks. The rates for the other population groups imply that Blacks have the highest at 55 and 48, followed by Coloureds at 28 and 25 then Indians/Asians at 22 and 16 and Whites with the lowest at 15 and 13 for males and female infant deaths per one thousand infants respectively. The rates for Blacks have a heavier effect on the national rates as might be expected than the smaller population groups **(see Table 4.9).**

Table 4.9 Infant mortality per 1,000

	National	Blacks	Whites	Coloureds	Indians/Asians
Male	49	55	15	28	22
Female	44	48	13	25	16

Table 4.10 Under-five Mortality rates per 1,000

	National	Blacks	Whites	Coloureds	Indians/Asians
Male	76	86	20	42	28
Female	68	76	19	36	21

The under-five mortality rates show a similar trend to that exhibited by the infant mortality rates. The mortality rates for the females are lower than for males for the national population and the population groups as is widely expected (see **Table 4.10**).

Table 4.11 Adult mortality: ${}_{45}q_{15}$

	National	Blacks	Whites	Coloureds	Indians/Asians
Male	0.561	0.613	0.260	0.414	0.311
Female	0.461	0.518	0.166	0.275	0.173

In the case of adult mortality, estimates of the probability of a person aged 15 surviving to reach the age of 60 were calculated for males and females for the national population and all the population groups. The pattern shows that the national averages are pushed up by the mortality rates for black Africans, which are higher than the rest of the population groups (see
Table 4.11).

Table 4.12 Life Expectancy at birth by sex and population group

	National	Blacks	Whites	Coloureds	Asians
Male	52	49	66	59	64
Female	56	53	72	66	71

The life expectancy at birth estimates for the national are consistent with the weighted average of the population groups. The national estimates are at 52 years for females and 56 years for males and fall in between those for Whites with the highest at 66 and 72 and for Blacks with the lowest with 49 and 53 years for males and females respectively (see **Table 4.12**).

5 DISCUSSION AND CONCLUSION

5.1 Introduction

The aim of this research project was to produce a set of full life tables for South Africa at both national and population group levels using deaths captured by the vital registration system in the period 2006-2008. In the process the study also estimates the levels of completeness of the death registration for the national population and population groups for males and females separately.

This chapter discusses the quality of the data used in the study, the estimates of completeness and the impact of the imputation process on completeness estimates. The graduation methods applied and the mortality rates produced are also discussed and the mortality rates are compared with estimates from other relevant studies.

The chapter concludes with an outline of the limitations of the study, the scope for future research and the main conclusion which will set out to explain the importance and relevance of the results and the methodology applied.

5.2 Data quality

The research made use of the 2001 national census population estimates, 2007 Community Survey population estimates, deaths captured by the vital registration system and mid-year population estimates from ASSA2008 demographic projection model.

The study noted that the 2001 national census population estimates underestimated male and female children. However, these population estimates were used for the purpose of estimating completeness using the deaths distribution methods which exclude the 0-4 age group population estimates. The study also noted that digit preference as a result of mass registrations is evident in the older age groups in both the 2001 national census and the 2007 community survey (Machemedze 2009). This introduced fluctuations in mortality rates for the older age groups and underestimation of old age mortality. In this study, the cubic spline graduation technique was applied to 'smooth' the mortality rates up to age 85 for the national population. Beyond the age 85, extrapolation techniques were applied to estimate mortality rates and these produce smooth mortality rates.

It was noted that in the case of data from the vital registration system, not all deaths are registered and for those that are registered up to 26 per cent did not indicate the population group of the deceased. To estimate the 'true' number of deaths, the death distribution methods were applied and an imputation technique was used to impute the population group of the deceased in the cases of the unit records that had no indication of the population group of the deceased. While the results of the imputation process looked plausible, the resulting estimates of completeness for the population groups raise some questions on the accuracy of the process. The results of the completeness estimation will be discussed in the next section.

5.3 Imputation and Completeness

As discussed in earlier sections, not all deaths are reported and captured in the vital registration system. Further, the level of coverage of different censuses and surveys might not necessarily be the same. Therefore, the population estimates from the censuses and surveys and the death data from the vital registration system cannot be applied as they are. In this research the SEG and the GGB methods were applied first to estimate the level of completeness of death reporting and then to assess the extent of differential coverage between the 2001 national census and the 2007 Community Survey. The results of these assessments were then taken into account in the process of estimating the 'true' number of deaths.

In the process of estimating the completeness of the different population groups, imputation of population group for the relevant unit record data was done using the hot deck method in Stata. Thereafter, the levels of completeness of the different population groups were estimated using the GGB and the SEG methods. The methods gave consistent estimates for the national population and the population groups.

In the case of Indians/Asians, after imputing for missing population group, the estimates of the levels of completeness, exceeded 100 per cent for both males and females and was lower than expected for Whites at 80 and 82 per cent for males and females respectively. While the completeness levels for Whites are similar to the estimates from studies by Dorrington Moultrie and Timaeus (2004) based on reported deaths without imputation in the period 1996-2001, which estimated completeness to be 77 and 79 per cent for males and females respectively, the expectation was that after including the imputed units the levels would increase to levels close to 100 per cent. The same was expected for the Coloured population.

Table 5.1 Completeness estimates without imputation

	Blacks	Whites	Coloureds	Indians/Asians
Male	63%	76%	61%	89%
Female	59%	78%	66%	87%

Table 5.2 Percentage-point increase in level of completeness after imputation

	Blacks	Whites	Coloureds	Indians/Asians
Male	31%	4%	27%	11%
Female	40%	4%	27%	13%

When one looks at Figure 5.1, the current estimates for Blacks, Coloureds and Indians/Asians are significantly higher than the estimates from previous studies which were done without imputation. Dorrington Moultrie and Timaeus (2004) placed the estimates for Blacks at 63 and 68 percent for male and female black populations without imputation respectively, whereas in this research the estimates were placed at 94 and 89 percent for males and females with imputation, respectively. In the case of Coloureds, Dorrington Moultrie and Timaeus (2004) placed the estimates at 70 and 80 percent without imputation whereas the current research placed the same at 88 and 94 per cent for males and females with imputation, respectively.

Tables 5.1 and 5.2 show that the imputation process resulted in a 4 percentage points increases in completeness for both male and female White populations. In the case of black Africans the levels of completeness increased by 31 and 30 percentage points for males and females respectively.

While the levels of completeness increased significantly for all the population groups except for Whites, it can be concluded that the imputation process did not produce the expected results for Whites and Coloureds. It appears that the imputation process over-allocated the imputed units to the Indian/Asian and black African population groups at the expense of White and Coloured population groups. As a result, the levels of completeness for Whites and Coloureds were underestimated.

Figure 5.1 Completeness estimates for males: Comparisons

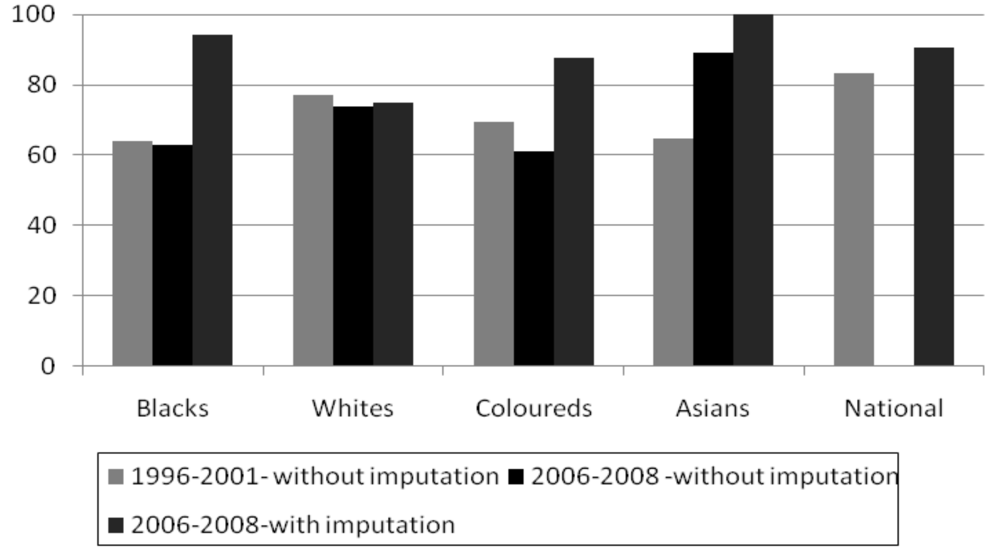
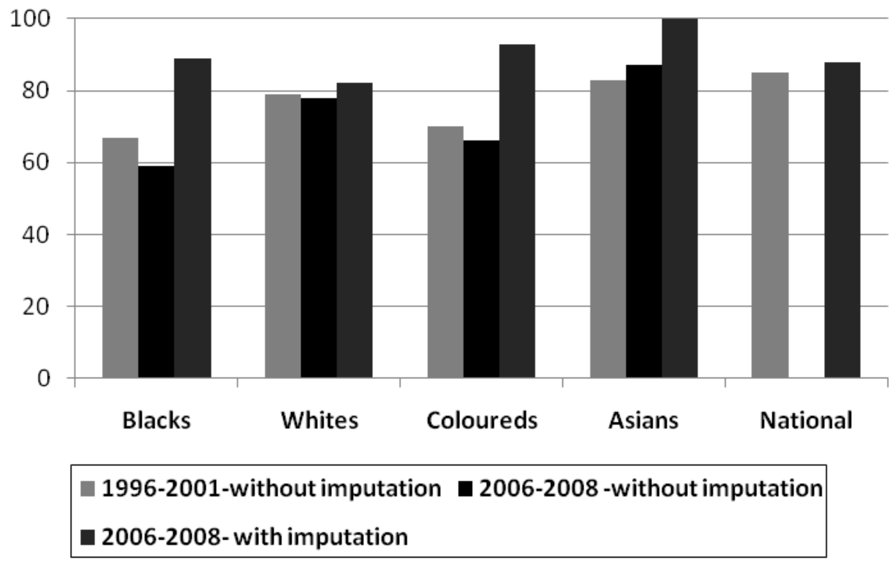


Figure 5.2 Completeness estimates for females: Comparisons



5.4 Graduation and mortality rates

The graduation of central mortality rates was undertaken using the cubic spline smoothing technique which fitted cubic polynomials between single ages with knots at every age using Matlab. This technique has the advantage of producing graduated mortality rates that follow the empirical data very closely as it does not impose any predetermined shape on the final shape and profile of the mortality curve. Therefore, a cubic spline fitted mortality curve is more likely to fit the data more closely compared to parametric formula like the Heligman and Pollard equation.

However, the cubic spline smoothing output did not give smooth and regular curves in the case of small population data like the Indian/Asian and the White populations (see Figures 5.3 and 5.4). Further, the cubic spline smoothing technique does not have an automatic way of calculating the smoothing parameter value that strikes an optimum balance between the goodness-of-fit objective and smoothness. The graduator has to exercise judgement in choosing the smoothing parameter or use Reinsch algorithm to narrow down the range within which the graduator can exercise judgement. The application of the Reinsch algorithm depends to a large extent on the estimates of variance of the number of deaths. In a developing country context the estimation of variance of the number of deaths is affected by the incompleteness of vital registration and hence the estimates of variance will depend on the assumptions that the graduator chooses to make regarding the distribution of the level of completeness.

In this research, the Reinsch algorithm did not converge in the required range of values for some of the data sets and in some cases where it converged, the smoothing parameters in the suggested ranges did not give optimum smoothing and hence the graduator had to adjust the smoothing parameters by using judgment.

Figure 5.3 Graduated and crude central mortality rates for Indian/Asian males

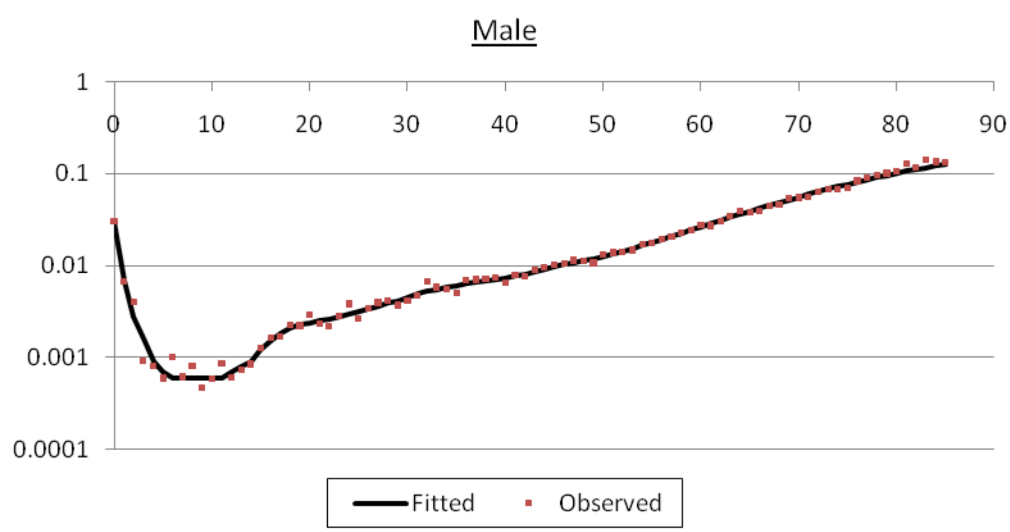
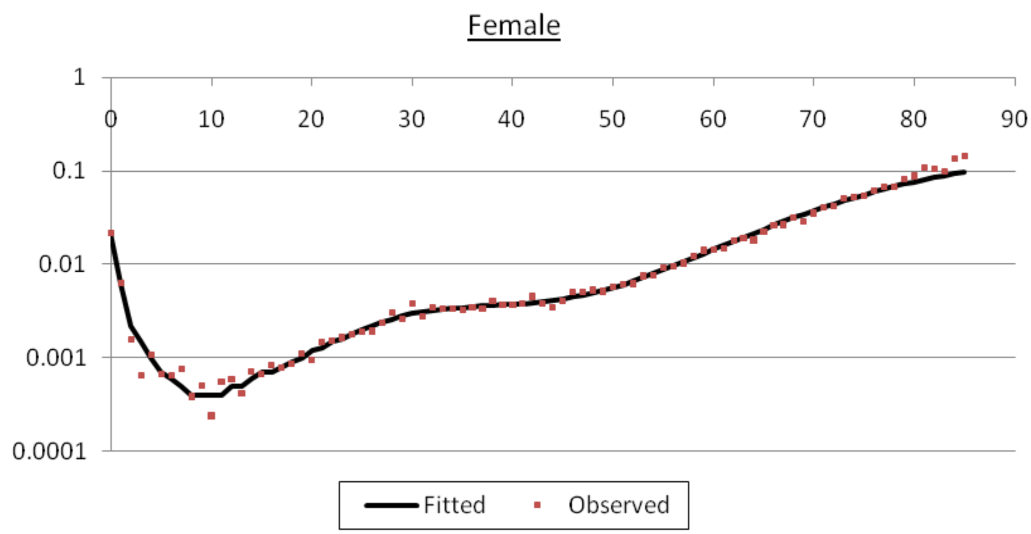


Figure 5.4 Graduated and crude central mortality rates for Indian/Asian females



The national mortality rates for both males and females were very close to the weighted sum of all the population group mortality rates (see Figures 4.12). The age specific population proportions for the different population groups were used as the weights. The national mortality rates were checked for internal consistency by comparing them with the weighted sum of the population groups. While the Chi-squared test for differences gave unsatisfactory results, other tests indicated that the national mortality rates are consistent with the weighted sum of the mortality rates of all the population groups.

The infant mortality estimates for the national population are consistent with the rates calculated in previous studies. Marszalek (2010) estimated the infant mortality rates at 51 and 45 deaths per one thousand male and female infants respectively, using vital registration data for the period 2006-2007. In this research, which is centered in the middle of 2007 the estimates are at 49 and 44 deaths per one thousand male and female infants. On the other hand, the ASSA2008 model places the infant mortality rates at 43 and 38 male and female deaths per one thousand infants (Actuarial Society of South Africa 2010). Statistics South Africa projected the infant mortality rate at 45 deaths per one thousand male and female births combined in the year 2007. The UNPD (2011) places the same estimates for the period 2005-2010 at 55 and 43 deaths. The infant mortality estimates produced in this research are in the same range as the estimates from other studies although there is a lot of variation in the estimates produced by different researchers.

The full version of the ASSA2008 model can be used to project mortality rates for all the four population groups separately. The infant mortality rates estimated in this research for the population groups are substantially higher than the estimates from the ASSA2008 projection model. In the case of black Africans (Table 4.8), the ASSA2008 model places the estimates at 47 and 42 per one thousand male and female infants as compared to 55 and 48 male and female deaths in this research. The same pattern persisted for all the other population groups.

In the case of adult mortality as estimated by calculating ${}_{45}q_{15,}$ for the national population, ASSA2008 places the ${}_{45}q_{15}$ at 56 and 42 per cent for males and females as compared to the estimates in this research of 56 and 46 per cent. The UNPD places the same estimates at 58 and 51 per cent for males and females for the period 2005-2010 (UNPD 2011). Clearly the male rates are consistent but there is wide variation in the female estimates with the estimate from this research in the middle of the ASSA2008 estimate and the UNPD estimate. A study by Dorrington Moultrie and Timaeus(2004), which covered the period between the 1996 census and the 2001 census placed ${}_{45}q_{15}$ at 51 and 34 per cent for males and females respectively. One can conclude that adult mortality between 1998 and 2007 has increased. Of note is the increase in female adult mortality over the period from 34 per cent to between 41 and 52 percent as compared with the male increase from 51 to 56 per cent. It appears that the effect of HIV/AIDS was greater on the female population than on the male population during this period.

A comparison of the life tables produced in this study in Figure 5.4 for the national male and female populations shows that the male life tables from the ASSA2008 model, the UNPD and this research are fairly more consistent, particularly from age 10 upwards, compared to the females.The male mortality estimates from this research flatten out faster than the other estimates, giving higher mortality estimates in the 40-60 age range.

In the case of the female life tables, the estimates from this research overestimate the level of mortality when compared to the ASSA2008 model estimates. There are differences in the sizes of the AIDS humps of the three life tables. The UNDP life table has a more pronounced AIDS hump than the other life tables and that is reflected in the estimate for ${}_{45}q_{15}$ for the UNPD which is higher than the estimates from this research and from the ASSA2008 model. The ‘hump’ as shown in Figure 5.6 is between age 20 and 40.

Figure 5.5 ASSA2008, UNPD life tables and estimated life table for national male populations

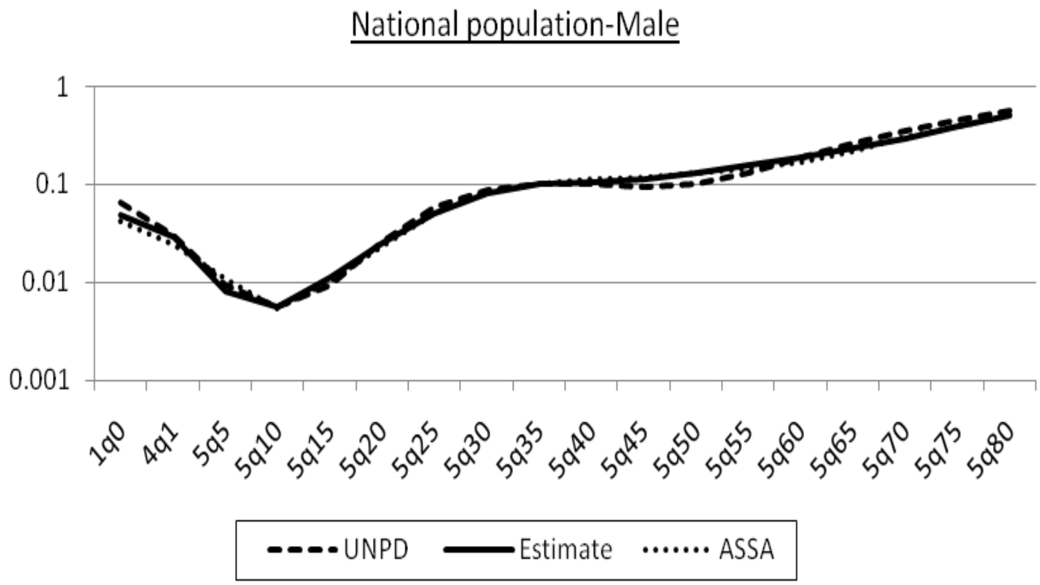
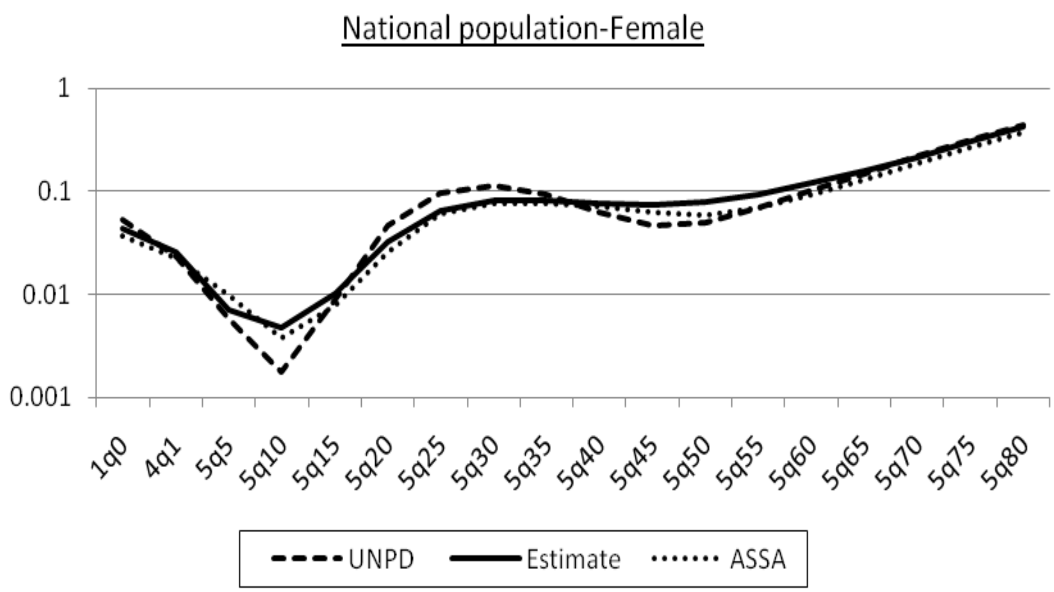


Figure 5.6 ASSA2008, UNPD life tables and estimated life table for national female populations



The mortality estimates for the population groups follow the expected ranking with the black Africans recording the highest mortality followed by Coloureds, Indians/Asians and lastly by Whites who had the lowest mortality rates for both males and females. However, at the highest ages the ranking changes as the White mortality rates increased at a faster rate around age 85 settling in second position after black Africans. It is also evident that the AIDS humps are more pronounced among the black African males and females when compared to other population groups. The black African mortality patterns resemble the national mortality patterns for both males and females. This is because the black Africans constitute a significant majority of the national population when compared with the other population groups (see Figures 4.13 and 4.14).

The life tables for Indians/Asians, Whites and Coloureds do not progress smoothly at the younger ages. These are smaller populations with even smaller numbers of people in the younger age groups and hence the mortality rates tend to be unstable.

Externally the smaller population groups' mortality estimates were generally not consistent with the rates from the ASSA2008 model, particularly for the younger age groups for both males and females. It appears that the ASSA2008 produced lower mortality estimates for younger age groups when compared to estimates from this research. However, the black Africans' mortality rates were generally consistent with the rates from the ASSA2008 model (Appendix A11).

5.5 Limitations of the study

The fact that not all deaths are registered implies that researchers have to estimate the 'true' number of deaths. In this research, the level of completeness was estimated using the GGB and SEG methods by assuming that the level of completeness does not vary with age in the age range 15-85, an assumption which may not be sustained for the entire age range. Further, the GGB and SEG methods cannot be used to estimate the levels of completeness for the deaths of the under-fifteens. Consequently, results from research by Darikwa (2009) were used to estimate the level of completeness only for the under-fives. In the case of the 5-14 a different approach was used which entailed assuming a linear trend between the level of completeness of the under-fives and the level of completeness of the population above 15.

The fact that not all registered deaths indicated the population group of the deceased impacted on the process of estimating mortality rates for the different population groups. The missing population groups had to be imputed to the unit records that had no indication of the population group. The hot deck imputation method which uses predictor covariates from the same data set was used. However, in this instance some of the potentially useful covariates like level of education, province of birth and smoking habits also had missing information and hence could not be used in the imputation process. This weakened the ability of the imputation process to predict the most correct and most likely population group. Comparison of the numbers of imputed deaths by population group with the numbers based on simple apportionment within each age group suggests that implementation of the hot deck imputation produced results that are so implausible as to undermine the life tables produced for the population groups.

The estimation of mortality rates requires population counts for the relevant periods but these are not available for the 2006-2008 period as censuses are now held every 10 years. Hence, population estimates derived from the ASSA2008 model were used. In estimating the level of completeness of death registration in the period 2001-2007, the population estimates calculated from a sample survey, the 2007 Community Survey were used as opposed to a proper full population count. This introduces sampling errors and the risk of possible bias in the population estimates (Statistics South Africa 2007a). Even though the extent to which the sampling errors affect the completeness estimates and the mortality rates cannot be assessed, the expectation is that they should not have a significant impact.

In estimating the levels of completeness of vital registration and in estimating the mortality rates a no-migration assumption was made. This was a result of the unavailability of migration data for the relevant periods.

The Reinsch algorithm which narrows down the range within which a graduator can locate the optimum smoothing parameter did not produce plausible ranges in some instances. In the cases where the algorithm produced ranges in the acceptable region, the mortality curves produced in some cases were not the best and hence the graduator had to use judgment to settle for smoothing parameters. The failure of the Reinsch algorithm could have been a result of the uncertainty in the estimation of the variance of the number of deaths.

5.6 Scope for future research

There is need for research on how the level of completeness for the 5-14 age group can be estimated. The assumptions made in this research regarding the levels of completeness for the 5-14 age group cannot be sustained.

The imputation of population group to unit records is an area that clearly needs further research. The results of the imputation conducted in this research did not produce the expected results in terms of the ranking by population group of the levels of completeness. The method used in this research in Stata rejected some useful covariates as they had missing information. Further research could be directed at finding ways of making use of such covariates even when they are not 100 per cent complete.

The use of the Reinsch algorithm in narrowing the range within which the smoothing parameter value is located is an important aspect of the cubic spline technique (Reinsch 1967). However, this approach depends on how the variance of level of completeness of death registration is estimated. The fact that in some instances the Reinsch algorithm converged outside the possible range (0 to 1) points to the need for further research on how the variance of the level of completeness can be estimated.

5.7 Conclusion

The weighted sum of the mortality rates of the population groups are reasonably close to the national mortality rates for males and females, as shown in

Figure 4.12.

Externally, the national life tables and life tables for the black Africans produced in this research are generally consistent with results from other related studies. However, the mortality rates from the smaller population groups were generally not consistent with the rates from the ASSA2008 model. The differences might be due to the fact that migration was ignored in estimating the completeness of reporting of deaths. The absence of migration estimates in the context of a country like South Africa, which has a lot of migration activity, is bound to affect the mortality rates. Further, this research uses results from the imputation to estimate mortality rates for the population groups. As discussed in earlier sections, the imputation process did not produced plausible completeness estimates for the population groups. The expectation was that completeness would rise to close to 100 per cent for Whites and Coloureds. Further, the infant mortality rates for Whites and Indians/Asians were considerably higher than expected when calculated including the imputed units. As a result the infant mortality rates were estimated by averaging the mortality rates calculated before and after imputation only for the Whites and the Indians/Asians. The other population groups were not affected and hence the mortality rates were estimated after imputation.

The method of graduation employed preserves the shape and profile of the crude mortality rates and worked reasonably well for the bigger populations. The mortality curves for the national and the black African male and female populations exhibit reasonably smooth progressions. However, in the case of the smaller populations: Whites, Coloureds and Indians/Asians, the mortality curves are not smooth, particularly for the younger age groups.

The assessment of the level of completeness for the national population was consistent with results from other researchers but for the population groups the failure of the imputation process affected the estimated levels of completeness for the population groups. Therefore, only the life tables and the mortality rates for the national and possibly the black African population group can be considered to be reliable. The life tables and mortality rates for White, Coloured and Indian/Asian populations were affected by the results of the imputation process and the unavailability of migration data (in the case of the White population) and hence it is not advisable that they be used at all. It must be noted, however, that there are few researchers who include population groups in their analyses of mortality patterns in South Africa and hence other than ASSA, there are no other estimates available for further comparison of the population groups' life tables. However, the heterogeneity of mortality patterns of the different population groups justify the analysis of mortality at population group level in South Africa.

6 REFERENCES

Acock, A.C. 2005. "Working with missing values", *Marriage and Family* **67**(.):1012-1028.

Actuarial Society of South Africa (2010) *ASSA2008 Aids and Demographic Models* Cape Town.

Altermayer, L. nd. *Hot Deck Imputation: A simple data step*. Washington D.C.:US Census Bureau.

Andridge, R.R. and R.J.A. Little. 2010. "A Review of the Hot Deck Imputation for Survey Non-response", *International Statistical Review* **78**(1):40-64.

Arriaga, E., P. Johnson and E. Jamison. 1994. *Population Analysis with microcomputers: Volume 1*. United States Census Bureau.

Bah, S.M.1997. "A critical analytical and historical review of South African published life tables.," Paper presented at Annual Conference of the South African Statistical Association. Brits, November.

Benjamin, B. and J.H. Pollard (1993) *The Analysis of Mortality and Other Actuarial Statistics* The Institute of Actuaries and The Faculty of Actuaries,U.K.

Bennett, N.G. and S. Horiuchi. 1981. "Estimating the Completeness of Death Registration in a Closed Population ", *Population Index* **47**(2):207-221.

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

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

Coale, A.J. and E. Kisker. 1989. "Defects in Data on Old Age Mortality in The United States: New procedure for calculating mortality schedule and life tables at the highest ages ", *Asian and Pacific Population Forum* **4**(1):1-31.

Darikwa, T.B. 2009. "Estimating the level and trends of child mortality in South Africa, 1996-2006 " Unpublished MPhil Thesis, Cape Town: University of Cape Town.

Debon, A., F. Montes and R. Sala. 2005. "A comparison of parametric models for mortality graduation. Application to mortality data for the Valencia region (Spain)", *SORT* **29**(2):269-288.

Dorrington, R. 1989. "African mortality estimates: an initial estimate", *Transactions of the Actuarial Society of South Africa* **7**(1):169-202.

Dorrington, R., 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:Medical Research Council Technical Report.

Dorrington, R., T.A. Moultrie and I.M. Timaeus. 2004. *Estimation of Mortality using the South African Census 2001 Data*. Cape Town:Centre for Actuarial Research.

Dorrington, R., I.M. Timaeus and T.A. Moultrie. 2008. *Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors, Revisited*. Cape Town:Centre for Actuarial Research, University of Cape Town.

Glen, A.G., L.M. Leemis and J.H. Drew. 2003. *Computing the distribution of the product of two continuous random variables*. http://www.math.wm.edu/~leemis/2003csada.pdf. Accessed: 29 June 2011.

Goodman, L.A. 1960. ""On the exact variance of products"", *American Statistical Association* **55**(292):708-713.

Government Actuaries Department. 1995. *English Life Table No. 15*. London:Office of National Statistics.

Government Actuaries Department. 2005. *English Life Table No. 16*. London:Office for National Statistics
Green, P.J. and B.W. Silverman (1994) *Nonparametric Regression and Generalized Linear Models,* London SE1 8HN, Chapman and Hall.
Haberman, S. 1996. *Smoothing spline graduation of the crude data for the English Life Tables No. 15*. London:Government Actuaries Department.
Heligman, L. and J.H. Pollard. 1980. "The age pattern of mortality ", *Journal of the Institute of Actuaries* **107**(.):49-80.
Hill, K. and Y. Choi.2004. "Death distribution methods for estmating adult mortality: Sensitivity analysis with simulated data errors," Paper presented at Adult mortality in developing countries workshop. Marconi Center, Marin County, Carlifornia, 8-11 July.
Hill, K., Y. Choi and I.M. Timaeus. 2005. "Unconventional approaches to mortality estimation", *Demographic Research* **13**(12):281-300.
Hill, K., D. You and Y. Choi. 2009. "Death distribution methods for adult mortality: Sensitivity analysis with simulated data errors", *Demographic Research* **21**(9):235-254.
Hill, K.H. 1987. "Estimating Census and Death Registration Completeness", *Asian and Pacific Population Forum* **1**(3):8-13.
Machemedze, T. 2009. "Old Age Mortality in South Africa." M Phil Demography, Cape Town: University Of Cape Town.
Mander, A. and D. Clayton (1999) Hot Deck: Stata module for imputing the missing values using the hotdeck method. *Statistical Software Components.*
Marszalek, S. 2010. "South African Life Table (2006-2007)." Honours Project, Cape Town: University of Cape Town.
Mathworks. 2010. *R2011a Documentation: Users Guide*.
Nordholt, E.S. 1998. "Imputation: Methods, Simulation Experiments and Practical Examples", *International Statistical Review* **66**(2):157-180.
Preston, S.H., A.J. Coale, J. Trussell*, et al.* 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index* **46**(2):179-202.
Reinsch, C.H. 1967. "Smoothing by Spline Functions", *Numerische Mathematik* **10**(177-183.
Reinsh, C.H. 1967. "Smoothing by Spline Functions", *Numerische Mathematik* **10**(177-183.
Renshaw, A., S. Haberman and P. Hatzopoulos. 1997. "On the duality of assumptions underpinning the construction of life tables", *Astin Bulletin* **27**(27):5-22.
Rohatgi, V.J. (1976) *An introduction to probability theory and mathematical statistics,* New York, Wiley.
Spoorenberg, T. and C. Dutreuilh. 2007. "Quality of age reporting: Extension and application of the modified Whipple's Index", *Population* **62**(4):729-742.
Stata Corp. 2009. *Statistical Software. Release 11*. College Station, Texas:Stata Corp, LP.
Statistics South Africa. 2000. *South African life tables, 1985-1994 and 1996*. Pretoria:Statistics South Africa.
Statistics South Africa. 2001. *Mortality and Causes of Death in South Africa: Findings from the death notification* Pretoria:Statistics South Africa.
Statistics South Africa. 2002. *Mortality and Causes of Death in South Africa: Findings from the death notification*. Pretoria:Statistics South Africa.
Statistics South Africa. 2003. *Mortality and Causes of Death in South Africa: Findings from the death notification*. Pretoria:Statistics South Africa.
Statistics South Africa. 2004. *Mortality and Causes of Death in South Africa: Findings from the death notification*. Pretoria:Statistics South Africa.
Statistics South Africa. 2005. *Mortality and Causes of Death in South Africa, 1997-2003: Findings from Death Notification* Pretoria:Statistics South Africa.

Statistics South Africa. 2006. *Mortality and Causes of Death in South Africa, 2003 & 2004: Findings from death notification* Pretoria:Statistics South Africa.

Statistics South Africa. 2007a. *Community Survey, 2007 (Revised version)*. Pretoria:Statistics South Africa.

Statistics South Africa. 2007b. *Mid-year Population Estimates, 2007: Statistical Release P0302*. Pretoria:Statistics South Africa.

Statistics South Africa. 2007c. *Mortality and Causes of Death in South Africa: Findings from the death notification*. Pretoria:South Africa.

Statistics South Africa. 2009. *Mortality and Causes of Death in South Africa, 2007: Findings from death notification*. Pretoria:Statistics South Africa.

Statistics South Africa. 2010. *Mortality and Causes of Death in South Africa, 2008: Findings from Death Notification*. Pretoria:Statistics South Africa.

Udjo, E. 2008. "A re-look at recent statistics on mortality in the context of HIV/AIDS with particular reference to South Africa", *Current HIV Estimate* **6**(143-151.

UN Population Division. 2002. *Methods for estimating adult mortality*. New York:United Nations, Department of Economic and Social Affairs.

UNPD. 2011. *World Population Prospects: The 2010 Revision*. CD-ROM Edition:

Vincent, P. 1951. "La Mortalite' des Vieillards", *Population* **6**(2):181-204.

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

A1 Code for cubic spline smoothing in Matlab

```
>x= [age]; y= [mortality rates], p=value,w=[weights]
>pp=csaps(x,y,p,[ ],w)
>xx=linspace(0,85,86)
>yy=csaps(x,y,p,xx,w)
```

A2 Code for hotdeck imputation in Stata

Data/Predictor variable exploration

```
tab age popgroup, missing
tab  mstatus,missing
tab  educode,missing
tab  smoker, missing
tab  deathinst, missing
tab death_prov, missing
tab  birth_prov, missing
tab  birth_prov, missing
tab  res_prov, missing
tab  res_prov, missing
tab  ascertainment, missing
tab  underlyingcause, missing
tab  underlyingcause, missing
tab popgroup, missing
codebook popgroup
# Marking the unit records with missing population group
replace popgroup=. if popgroup==9
replace popgroup=. If popgroup==8
replace popgroup=. If popgroup==5
# Setting and registering the data set and variables for imputation
mi set mlong
mi register imputed popgroup
mi register regular  ascertainment sex death_prov deathinst
# Imputing missing population group
mi impute mlogit popgroup  sex deathinst death_prov ascertainment, add(1)
#Analysis of imputed data set
tab popgroup
tab age popgroup
tab age popgroup if sex==1
tab age popgroup if sex==2
tab sex
# Calculating the number of still births
tab age popgroup if  ageminute==0 & agehour==0 & ageday==0& agemonth==0 & age==0 &
sex==1
tab age popgroup if  ageminute==0 & agehour==0 & ageday==0& agemonth==0 & age==0 &
sex==2
```

A3 Adjustment of variance

The adjustment of the number of deaths for incompleteness alters the variance and therefore the new variance has to be calculated by randomizing the level of completeness and the number of deaths as a way of simulating the empirical numbers of deaths. The simulated number of deaths can be expressed as follows:

$$D'_x = \frac{c_x D_x}{c}.$$

D_x stands for the number of deaths for persons aged x last birthday adjusted for incompleteness, P_x stands for the number of persons aged x in the mid-year population, the ratio c_x /c follows a uniform distribution U(a ,b) and D_x follows normal $N(\mu,\sigma^2)$. We can define a new variable $W=YV$ given that the other quantities are defined as follows:

$$W = D'_x, V = \frac{c_x}{c}, Y = D_x, E(V) = \mu_x, Var(V) = \sigma_x^2, E(Y) = \mu_y, Var(Y) = \sigma_y^2 .$$

Probability density function (pdf)

According to Rohatgi (1976) in Glen Leemis and Drew (2003) a convolution of two independent random variables V and Y where W=VY, will yield a p.d.f of the form:

$f_w(w) = \int_{-\infty}^{\infty} f_v(v).f_y(\frac{w}{v}).\frac{1}{|v|}dv$ which in this scenario will give the following p.d.f.

$f_w(w) = \frac{1}{b-a}\frac{1}{\sigma_y\sqrt{2\pi}}\int_a^b \exp(-\frac{1}{2}(\frac{y-\mu_y}{\sigma})^2).\frac{1}{v}dv$ (Professor MacDonald personal communication)

Expectation

The expected value of this variable with the given p.d.f. can be estimated by applying results which give the expected value and variance of two independent random variables:

E(W)=E(VY) and in this instance

$$E(\frac{c_x}{c}.D_x) = \frac{1}{c}E(c_x)E(D_x) = \frac{1}{c}.\frac{(a+b)}{2}.E(D_x) = \frac{1}{c}.c.E(D_x) = E(D_x) = \mu_y$$

*Assuming that E(c_x)=(a+b)/2=c

Variance

According to Goodman (1960) the variance of the product of two independent random variables can be estimated by applying the following:

$$Var(VY) = E(V)^2 Var(Y) + E(Y)^2 Var(V) + Var(V)Var(Y)$$

$$E(V) = E(\frac{c_x}{c}) = \mu_x = 1, \mathrm{var}(V) = \sigma_x^2 = \frac{(b-a)^2}{12}$$

$$V(D'_x) = 1^2.\sigma_y{}^2 + \frac{(b-a)^2}{12}.\mu_y{}^2 + \frac{(b-a)^2}{12}.\sigma_y{}^2 = \frac{12\sigma_y{}^2 + (b-a)^2(\mu_y{}^2 + \sigma_y{}^2)}{12}$$

The variance of the central mortality rate can then be expressed as follows:

$$Var(\frac{D'_x}{P_x}) = \frac{1}{P_x^2} V(D'_x) = \frac{12\sigma_y{}^2 + (b-a)^2(\mu_y{}^2 + \sigma_y{}^2)}{12P_x^2}$$

A4 Graduation test results after adjusting variance.

Black Population		Chi-squ	CumDev	Sign	Steven's	St.Dev
Blacks	Male	1315.2	-2.8	41	-0.28	17.66
	Female	1275.2	-3.7	38.00	-1.02	13.49
Critical Values		105	+-1.96	[33,51]	-1.64485	7.814728

White Population		Chi-squ	CumDev	Sign	Steven's	St.Dev
Whites	Male	255.9	-1	41	-0.66	1.65
	Female	230.3	-3.0	42	0.10	1.78
Critical Values		105	+-1.96	[33,51]	-1.64485	7.814728

Coloureds		Chi-squ	CumDev	Sign	Steven's	St.Dev
Coloureds	Male	157.1	0.2	40	-0.46	0.80
	Female	123.9	-0.4	43.00	0.10	1.72
Critical Values		105	+-1.96	[33,51]	-1.64485	7.814728

Asians		Chi-squ	CumDev	Sign	Steven's	St.Dev
	Male	116.6	1.0	46	-0.26	5.04
	Female	150.9	1.1	40.00	-0.08	5.98
Critical Values		105	+-1.96	[33,51]	-1.64485	7.814728

A5 Analysis of third differences – National population

Age	Mortality rate		First difference		Second Difference		Third Difference		% Third diff	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
82	0.13031	0.10317	0.00773	0.00798	0.00006	0.00041	-0.00037	-0.00113	-0.28	-1.09
83	0.13804	0.11115	0.00779	0.00839	-0.00031	-0.00072	0.00321	0.00238	2.33	2.14
84	0.14583	0.11954	0.00748	0.00767	0.00290	0.00166	0.00052	0.00065	0.36	0.55
85	0.15331	0.12721	0.01039	0.00933	0.00342	0.00231	-0.00097	-0.00068	-0.63	-0.54
86	0.16370	0.13654	0.01381	0.01164	0.00246	0.00163	-0.00078	-0.00013	-0.47	-0.10
87	0.17751	0.14818	0.01626	0.01326	0.00168	0.00149	-0.00343	-0.00172	-1.93	-1.16
88	0.19377	0.16144	0.01794	0.01476	-0.00175	-0.00022	0.00126	0.00044	0.65	0.27
89	0.21171	0.17620	0.01619	0.01453	-0.00050	0.00022				
90	0.22790	0.19073	0.01570	0.01475						
91	0.24360	0.20548								

A6 Life Tables – National population

Male

Age	q_x	l_x	L_x	e_x
0	0.04884	1.00000	0.98666	51.38349
1	0.01677	0.95116	0.94318	52.98470
2	0.00646	0.93521	0.93219	52.87973
3	0.00372	0.92917	0.92744	52.22022
4	0.00238	0.92572	0.92461	51.41316
5	0.00187	0.92351	0.92265	50.53480
6	0.00174	0.92179	0.92098	49.62837
7	0.00160	0.92018	0.91944	48.71408
8	0.00146	0.91871	0.91804	47.79135
9	0.00133	0.91737	0.91676	46.86042
10	0.00116	0.91615	0.91562	45.92231
11	0.00106	0.91508	0.91460	44.97499
12	0.00101	0.91412	0.91366	44.02211
13	0.00111	0.91319	0.91269	43.06604
14	0.00124	0.91218	0.91162	42.11327
15	0.00146	0.91105	0.91039	41.16500
16	0.00180	0.90972	0.90890	40.22439
17	0.00220	0.90808	0.90708	39.29602
18	0.00269	0.90609	0.90487	38.38156
19	0.00316	0.90365	0.90222	37.48380
20	0.00370	0.90079	0.89913	36.60098
21	0.00427	0.89746	0.89555	35.73505
22	0.00496	0.89363	0.89142	34.88603
23	0.00574	0.88920	0.88665	34.05738
24	0.00664	0.88409	0.88116	33.25116
25	0.00766	0.87822	0.87486	32.47014
26	0.00880	0.87150	0.86766	31.71687
27	0.01009	0.86383	0.85947	30.99402
28	0.01143	0.85511	0.85022	30.30489
29	0.01286	0.84533	0.83990	29.64960
30	0.01430	0.83446	0.82850	29.02930
31	0.01566	0.82253	0.81609	28.44318
32	0.01690	0.80965	0.80281	27.88769
33	0.01800	0.79597	0.78880	27.35850
34	0.01896	0.78164	0.77423	26.85081
35	0.01976	0.76682	0.75925	26.36003
36	0.02043	0.75167	0.74399	25.88128
37	0.02096	0.73631	0.72860	25.41073
38	0.02131	0.72088	0.71320	24.94399
39	0.02157	0.70552	0.69791	24.47619
40	0.02177	0.69030	0.68279	24.00467

Female

Age	q_x	l_x	L_x	e_x
0	0.04376	1.00000	0.98757	55.45501
1	0.01496	0.95624	0.94909	56.95996
2	0.00561	0.94194	0.93930	56.81731
3	0.00301	0.93665	0.93525	56.13494
4	0.00195	0.93384	0.93293	55.30281
5	0.00164	0.93202	0.93125	54.40989
6	0.00154	0.93049	0.92977	53.49853
7	0.00144	0.92905	0.92838	52.58037
8	0.00127	0.92771	0.92712	51.65556
9	0.00113	0.92654	0.92601	50.72044
10	0.00094	0.92549	0.92505	49.77742
11	0.00090	0.92462	0.92420	48.82386
12	0.00090	0.92378	0.92337	47.86739
13	0.00096	0.92295	0.92251	46.91006
14	0.00110	0.92207	0.92156	45.95458
15	0.00127	0.92105	0.92047	45.00464
16	0.00154	0.91989	0.91918	44.06108
17	0.00194	0.91847	0.91758	43.12834
18	0.00245	0.91668	0.91556	42.21127
19	0.00318	0.91444	0.91298	41.31372
20	0.00410	0.91153	0.90966	40.44406
21	0.00524	0.90779	0.90541	39.60850
22	0.00657	0.90303	0.90007	38.81457
23	0.00800	0.89710	0.89351	38.06784
24	0.00946	0.88993	0.88572	37.37080
25	0.01087	0.88151	0.87672	36.72287
26	0.01226	0.87193	0.86659	36.12082
27	0.01350	0.86124	0.85543	35.56289
28	0.01464	0.84961	0.84339	35.04271
29	0.01560	0.83717	0.83064	34.55599
30	0.01639	0.82411	0.81736	34.09568
31	0.01690	0.81061	0.80376	33.65555
32	0.01723	0.79691	0.79004	33.22551
33	0.01730	0.78317	0.77640	32.79937
34	0.01726	0.76962	0.76298	32.36799
35	0.01716	0.75634	0.74985	31.92763
36	0.01706	0.74336	0.73702	31.47629
37	0.01693	0.73068	0.72450	31.01387
38	0.01661	0.71831	0.71235	30.53947
39	0.01630	0.70638	0.70062	30.04680
40	0.01594	0.69487	0.68933	29.53639

Life Table- National population cont'd

Age	q_x	l_x	L_x	e_x
41	0.02204	0.67528	0.66784	23.52768
42	0.02231	0.66039	0.65303	23.04668
43	0.02260	0.64566	0.63837	22.56114
44	0.02289	0.63107	0.62385	22.07125
45	0.02319	0.61662	0.60947	21.57662
46	0.02350	0.60232	0.59525	21.07703
47	0.02396	0.58817	0.58112	20.57223
48	0.02447	0.57408	0.56705	20.06493
49	0.02507	0.56003	0.55301	19.55562
50	0.02580	0.54599	0.53895	19.04556
51	0.02664	0.53191	0.52482	18.53671
52	0.02763	0.51774	0.51058	18.03039
53	0.02870	0.50343	0.49620	17.52858
54	0.02981	0.48898	0.48169	17.03174
55	0.03100	0.47440	0.46705	16.53967
56	0.03220	0.45970	0.45230	16.05280
57	0.03340	0.44490	0.43747	15.57027
58	0.03460	0.43004	0.42260	15.09100
59	0.03581	0.41516	0.40772	14.61395
60	0.03717	0.40029	0.39285	14.13811
61	0.03870	0.38541	0.37796	13.66456
62	0.04035	0.37050	0.36302	13.19454
63	0.04221	0.35555	0.34804	12.72830
64	0.04426	0.34054	0.33301	12.26719
65	0.04643	0.32547	0.31791	11.81210
66	0.04869	0.31036	0.30280	11.36293
67	0.05104	0.29524	0.28771	10.91894
68	0.05355	0.28018	0.27267	10.47935
69	0.05631	0.26517	0.25771	10.04398
70	0.05935	0.25024	0.24281	9.61345
71	0.06285	0.23539	0.22799	9.18846
72	0.06685	0.22059	0.21322	8.77115
73	0.07131	0.20585	0.19851	8.36369
74	0.07631	0.19117	0.18388	7.96749
75	0.08185	0.17658	0.16935	7.58440
76	0.08785	0.16213	0.15501	7.21595
77	0.09425	0.14789	0.14092	6.86276
78	0.10097	0.13395	0.12718	6.52486
79	0.10800	0.12042	0.11392	6.20148
80	0.11527	0.10742	0.10123	5.89180
81	0.12271	0.09504	0.08920	5.59426
82	0.13031	0.08337	0.07794	5.30681
83	0.13804	0.07251	0.06750	5.02702
84	0.14583	0.06250	0.05794	4.75202
85	0.15331	0.05339	0.04929	4.47798
86	0.16370	0.04520	0.04150	4.19829
87	0.17751	0.03780	0.03445	3.92221
88	0.19377	0.03109	0.02808	3.66077
89	0.21171	0.02507	0.02241	3.42043
90	0.22790	0.01976	0.01751	3.20477

Age	q_x	l_x	L_x	e_x
41	0.01573	0.68379	0.67841	29.00678
42	0.01556	0.67303	0.66780	28.46246
43	0.01543	0.66256	0.65745	27.90438
44	0.01526	0.65233	0.64736	27.33395
45	0.01510	0.64238	0.63753	26.74974
46	0.01500	0.63268	0.62794	26.15219
47	0.01500	0.62319	0.61852	25.54283
48	0.01509	0.61384	0.60921	24.92419
49	0.01523	0.60458	0.59997	24.29844
50	0.01544	0.59537	0.59077	23.66658
51	0.01574	0.58618	0.58156	23.02992
52	0.01616	0.57695	0.57229	22.39025
53	0.01670	0.56763	0.56289	21.74977
54	0.01730	0.55815	0.55332	21.11067
55	0.01796	0.54849	0.54357	20.47351
56	0.01870	0.53864	0.53360	19.83876
57	0.01951	0.52857	0.52341	19.20729
58	0.02044	0.51826	0.51296	18.57950
59	0.02140	0.50766	0.50223	17.95678
60	0.02250	0.49680	0.49121	17.33853
61	0.02383	0.48562	0.47983	16.72611
62	0.02530	0.47405	0.46805	16.12228
63	0.02691	0.46205	0.45584	15.52778
64	0.02861	0.44962	0.44319	14.94334
65	0.03034	0.43676	0.43013	14.36871
66	0.03207	0.42351	0.41672	13.80267
67	0.03384	0.40992	0.40299	13.24338
68	0.03566	0.39605	0.38899	12.68974
69	0.03774	0.38193	0.37472	12.14048
70	0.04015	0.36752	0.36014	11.59704
71	0.04305	0.35276	0.34517	11.06122
72	0.04642	0.33757	0.32974	10.53634
73	0.05025	0.32190	0.31382	10.02487
74	0.05454	0.30573	0.29739	9.52881
75	0.05927	0.28905	0.28049	9.04967
76	0.06441	0.27192	0.26317	8.58830
77	0.06990	0.25441	0.24552	8.14512
78	0.07570	0.23663	0.22767	7.71968
79	0.08190	0.21871	0.20976	7.31097
80	0.08851	0.20080	0.19191	6.91854
81	0.09561	0.18303	0.17428	6.54180
82	0.10317	0.16553	0.15699	6.18052
83	0.11115	0.14845	0.14020	5.83397
84	0.11954	0.13195	0.12406	5.50098
85	0.12721	0.11618	0.10879	5.17997
86	0.13654	0.10140	0.09448	4.86210
87	0.14818	0.08755	0.08107	4.55190
88	0.16144	0.07458	0.06856	4.25674
89	0.17620	0.06254	0.05703	3.98000
90	0.19073	0.05152	0.04661	3.72431

A7 Life Table- Black Africans

Male					Female				
Age	q_x	l_x	L_x	e_x	Age	q_x	l_x	L_x	e_x
0	0.05447	1.00000	0.98513	49.05594	0	0.04748	1.00000	0.98651	52.51750
1	0.01852	0.94553	0.93678	50.83990	1	0.01647	0.95252	0.94468	54.09951
2	0.00821	0.92802	0.92421	50.78993	2	0.00722	0.93683	0.93345	53.99720
3	0.00438	0.92040	0.91839	50.20614	3	0.00382	0.93007	0.92830	53.38608
4	0.00266	0.91638	0.91516	49.42456	4	0.00232	0.92652	0.92545	52.58870
5	0.00214	0.91394	0.91296	48.55497	5	0.00184	0.92438	0.92352	51.70966
6	0.00196	0.91198	0.91109	47.65811	6	0.00169	0.92267	0.92189	50.80414
7	0.00183	0.91020	0.90936	46.75064	7	0.00159	0.92111	0.92038	49.88938
8	0.00163	0.90853	0.90779	45.83559	8	0.00144	0.91965	0.91898	48.96812
9	0.00143	0.90704	0.90639	44.90976	9	0.00127	0.91832	0.91774	48.03809
10	0.00126	0.90574	0.90517	43.97350	10	0.00107	0.91716	0.91667	47.09839
11	0.00111	0.90460	0.90410	43.02828	11	0.00096	0.91618	0.91574	46.14814
12	0.00110	0.90360	0.90310	42.07546	12	0.00096	0.91530	0.91486	45.19193
13	0.00111	0.90261	0.90211	41.12125	13	0.00101	0.91442	0.91396	44.23480
14	0.00124	0.90161	0.90105	40.16632	14	0.00114	0.91350	0.91298	43.27895
15	0.00137	0.90049	0.89987	39.21563	15	0.00133	0.91246	0.91185	42.32784
16	0.00170	0.89926	0.89849	38.26862	16	0.00160	0.91124	0.91051	41.38369
17	0.00210	0.89773	0.89678	37.33293	17	0.00200	0.90978	0.90887	40.44921
18	0.00254	0.89584	0.89470	36.41044	18	0.00254	0.90796	0.90681	39.52926
19	0.00304	0.89357	0.89221	35.50195	19	0.00331	0.90566	0.90416	38.62872
20	0.00360	0.89085	0.88924	34.60874	20	0.00431	0.90266	0.90072	37.75528
21	0.00424	0.88764	0.88576	33.73197	21	0.00557	0.89877	0.89627	36.91648
22	0.00491	0.88388	0.88171	32.87353	22	0.00710	0.89377	0.89060	36.12033
23	0.00580	0.87954	0.87699	32.03322	23	0.00876	0.88742	0.88354	35.37505
24	0.00677	0.87444	0.87148	31.21718	24	0.01046	0.87965	0.87505	34.68319
25	0.00784	0.86852	0.86511	30.42645	25	0.01210	0.87045	0.86518	34.04447
26	0.00906	0.86171	0.85781	29.66297	26	0.01369	0.85992	0.85403	33.45533
27	0.01044	0.85390	0.84944	28.92956	27	0.01516	0.84814	0.84172	32.91281
28	0.01191	0.84499	0.83995	28.22954	28	0.01650	0.83529	0.82840	32.41169
29	0.01351	0.83492	0.82928	27.56373	29	0.01770	0.82151	0.81424	31.94707
30	0.01524	0.82365	0.81737	26.93433	30	0.01876	0.80696	0.79940	31.51371
31	0.01690	0.81109	0.80424	26.34346	31	0.01956	0.79183	0.78408	31.10660
32	0.01849	0.79738	0.79001	25.78773	32	0.02010	0.77634	0.76854	30.71715
33	0.01993	0.78264	0.77484	25.26415	33	0.02043	0.76074	0.75296	30.33698
34	0.02124	0.76704	0.75889	24.76782	34	0.02060	0.74519	0.73752	29.95936
35	0.02249	0.75075	0.74230	24.29450	35	0.02070	0.72984	0.72229	29.57899
36	0.02356	0.73386	0.72522	23.84199	36	0.02079	0.71473	0.70730	29.19365
37	0.02450	0.71657	0.70779	23.40515	37	0.02076	0.69987	0.69261	28.80291
38	0.02516	0.69902	0.69022	22.98043	38	0.02056	0.68534	0.67830	28.40288
39	0.02556	0.68143	0.67272	22.56059	39	0.02020	0.67125	0.66448	27.98856
40	0.02584	0.66401	0.65543	22.13921	40	0.01980	0.65770	0.65118	27.55528

Life Table- Black Africans cont'd

Age	q_x	l_x	L_x	e_x
41	0.02610	0.64685	0.63841	21.71324
42	0.02640	0.62997	0.62166	21.28174
43	0.02679	0.61334	0.60512	20.84526
44	0.02710	0.59691	0.58882	20.40535
45	0.02741	0.58073	0.57277	19.95981
46	0.02780	0.56481	0.55696	19.50820
47	0.02827	0.54911	0.54135	19.05174
48	0.02887	0.53359	0.52589	18.59139
49	0.02960	0.51819	0.51052	18.12915
50	0.03041	0.50285	0.49520	17.66689
51	0.03137	0.48756	0.47991	17.20528
52	0.03250	0.47227	0.46459	16.74623
53	0.03376	0.45692	0.44920	16.29197
54	0.03513	0.44149	0.43374	15.84371
55	0.03653	0.42598	0.41820	15.40241
56	0.03793	0.41042	0.40263	14.96749
57	0.03930	0.39485	0.38709	14.53793
58	0.04066	0.37933	0.37162	14.11219
59	0.04203	0.36391	0.35626	13.68909
60	0.04346	0.34861	0.34104	13.26780
61	0.04501	0.33346	0.32596	12.84788
62	0.04677	0.31845	0.31101	12.42983
63	0.04870	0.30356	0.29617	12.01512
64	0.05070	0.28878	0.28146	11.60461
65	0.05270	0.27414	0.26691	11.19769
66	0.05470	0.25969	0.25259	10.79282
67	0.05669	0.24548	0.23853	10.38841
68	0.05861	0.23157	0.22478	9.98269
69	0.06077	0.21800	0.21137	9.57306
70	0.06324	0.20475	0.19827	9.16007
71	0.06621	0.19180	0.18545	8.74472
72	0.06988	0.17910	0.17284	8.32929
73	0.07431	0.16659	0.16040	7.91754
74	0.07952	0.15421	0.14808	7.51297
75	0.08561	0.14194	0.13587	7.11879
76	0.09242	0.12979	0.12380	6.73847
77	0.10001	0.11780	0.11191	6.37371
78	0.10821	0.10602	0.10028	6.02641
79	0.11701	0.09455	0.08901	5.69697
80	0.12631	0.08348	0.07821	5.38564
81	0.13601	0.07294	0.06798	5.09195
82	0.14610	0.06302	0.05841	4.81481
83	0.15647	0.05381	0.04960	4.55306
84	0.16711	0.04539	0.04160	4.30486
85	0.17792	0.03781	0.03444	4.06825
86	0.18904	0.03108	0.02814	3.84049
87	0.20096	0.02520	0.02267	3.61918
88	0.21429	0.02014	0.01798	3.40365
89	0.22937	0.01582	0.01401	3.19560
90	0.24523	0.01219	0.01070	2.99792

Age	q_x	l_x	L_x	e_x
41	0.01941	0.64467	0.63842	27.10179
42	0.01914	0.63216	0.62611	26.62831
43	0.01886	0.62006	0.61421	26.13821
44	0.01860	0.60837	0.60271	25.63099
45	0.01837	0.59705	0.59157	25.10729
46	0.01824	0.58609	0.58074	24.56770
47	0.01820	0.57539	0.57016	24.01489
48	0.01826	0.56492	0.55977	23.45080
49	0.01834	0.55461	0.54952	22.87763
50	0.01853	0.54444	0.53939	22.29574
51	0.01880	0.53435	0.52932	21.70732
52	0.01921	0.52430	0.51926	21.11366
53	0.01975	0.51423	0.50915	20.51737
54	0.02041	0.50407	0.49893	19.92067
55	0.02120	0.49379	0.48855	19.32527
56	0.02201	0.48332	0.47800	18.73301
57	0.02294	0.47268	0.46726	18.14332
58	0.02396	0.46184	0.45630	17.55759
59	0.02510	0.45077	0.44511	16.97630
60	0.02637	0.43946	0.43366	16.40050
61	0.02784	0.42787	0.42191	15.83110
62	0.02946	0.41596	0.40983	15.27016
63	0.03120	0.40370	0.39741	14.71847
64	0.03297	0.39111	0.38466	14.17638
65	0.03477	0.37821	0.37164	13.64261
66	0.03654	0.36507	0.35840	13.11600
67	0.03827	0.35173	0.34500	12.59449
68	0.04011	0.33827	0.33148	12.07572
69	0.04217	0.32470	0.31785	11.55940
70	0.04465	0.31101	0.30406	11.04627
71	0.04765	0.29712	0.29004	10.53917
72	0.05117	0.28296	0.27572	10.04147
73	0.05521	0.26848	0.26107	9.55600
74	0.05981	0.25366	0.24608	9.08519
75	0.06485	0.23849	0.23076	8.63131
76	0.07034	0.22302	0.21518	8.19520
77	0.07621	0.20734	0.19944	7.77745
78	0.08242	0.19154	0.18364	7.37780
79	0.08902	0.17575	0.16793	6.99556
80	0.09602	0.16011	0.15242	6.63027
81	0.10342	0.14473	0.13725	6.28140
82	0.11131	0.12977	0.12254	5.94826
83	0.11955	0.11532	0.10843	5.63065
84	0.12820	0.10153	0.09503	5.32731
85	0.13716	0.08852	0.08245	5.03717
86	0.14647	0.07638	0.07078	4.75840
87	0.15635	0.06519	0.06009	4.48916
88	0.16725	0.05500	0.05040	4.22844
89	0.17945	0.04580	0.04169	3.97723
90	0.19240	0.03758	0.03397	3.73767

A8 Life Table- Whites

Male

Age	q_x	l_x	L_x	e_x
0	0.01518	1.00000	0.99585	66.45759
1	0.00546	0.98482	0.98213	66.47080
2	0.00264	0.97944	0.97815	65.83314
3	0.00194	0.97685	0.97590	65.00618
4	0.00140	0.97496	0.97427	64.13167
5	0.00107	0.97359	0.97307	63.22088
6	0.00090	0.97255	0.97211	62.28786
7	0.00080	0.97168	0.97129	61.34352
8	0.00074	0.97090	0.97054	60.39223
9	0.00070	0.97018	0.96984	59.43668
10	0.00066	0.96950	0.96918	58.47797
11	0.00060	0.96886	0.96857	57.51616
12	0.00060	0.96828	0.96799	56.55039
13	0.00066	0.96770	0.96738	55.58404
14	0.00083	0.96706	0.96666	54.62033
15	0.00104	0.96626	0.96575	53.66547
16	0.00134	0.96525	0.96460	52.72091
17	0.00170	0.96396	0.96314	51.79106
18	0.00201	0.96232	0.96135	50.87841
19	0.00233	0.96038	0.95926	49.97979
20	0.00256	0.95814	0.95692	49.09551
21	0.00276	0.95569	0.95437	48.22015
22	0.00293	0.95306	0.95166	47.35214
23	0.00314	0.95026	0.94877	46.48998
24	0.00337	0.94727	0.94568	45.63492
25	0.00364	0.94409	0.94237	44.78739
26	0.00396	0.94065	0.93879	43.94926
27	0.00430	0.93692	0.93491	43.12193
28	0.00470	0.93290	0.93070	42.30599
29	0.00509	0.92851	0.92615	41.50341
30	0.00536	0.92378	0.92131	40.71326
31	0.00556	0.91883	0.91628	39.92989
32	0.00569	0.91373	0.91113	39.15028
33	0.00570	0.90853	0.90594	38.37153
34	0.00570	0.90335	0.90077	37.58863
35	0.00569	0.89820	0.89564	36.80125
36	0.00560	0.89309	0.89059	36.00904
37	0.00560	0.88808	0.88560	35.20902
38	0.00561	0.88311	0.88063	34.40448
39	0.00574	0.87816	0.87564	33.59570
40	0.00593	0.87312	0.87053	32.78682

Female

Age	q_x	l_x	L_x	e_x
0	0.01278	1.00000	0.99637	72.89707
1	0.00635	0.98722	0.98409	72.83111
2	0.00240	0.98096	0.97978	72.29304
3	0.00167	0.97861	0.97779	71.46575
4	0.00114	0.97698	0.97642	70.58423
5	0.00083	0.97586	0.97545	69.66433
6	0.00069	0.97505	0.97471	68.72202
7	0.00060	0.97437	0.97408	67.76924
8	0.00060	0.97379	0.97350	66.80962
9	0.00060	0.97320	0.97291	65.84943
10	0.00051	0.97262	0.97237	64.88867
11	0.00050	0.97212	0.97188	63.92141
12	0.00050	0.97164	0.97140	62.95314
13	0.00050	0.97115	0.97091	61.98438
14	0.00054	0.97067	0.97040	61.01514
15	0.00064	0.97014	0.96983	60.04794
16	0.00076	0.96952	0.96915	59.08617
17	0.00089	0.96878	0.96835	58.13063
18	0.00099	0.96792	0.96744	57.18207
19	0.00113	0.96696	0.96641	56.23833
20	0.00133	0.96586	0.96522	55.30157
21	0.00154	0.96458	0.96383	54.37474
22	0.00181	0.96309	0.96222	53.45792
23	0.00211	0.96135	0.96033	52.55386
24	0.00250	0.95932	0.95812	51.66384
25	0.00281	0.95692	0.95558	50.79207
26	0.00319	0.95424	0.95271	49.93371
27	0.00349	0.95119	0.94953	49.09199
28	0.00373	0.94787	0.94610	48.26225
29	0.00393	0.94433	0.94247	47.44123
30	0.00400	0.94062	0.93873	46.62659
31	0.00400	0.93685	0.93498	45.81184
32	0.00396	0.93311	0.93126	44.99382
33	0.00386	0.92941	0.92762	44.17064
34	0.00370	0.92583	0.92411	43.33979
35	0.00359	0.92240	0.92074	42.49888
36	0.00350	0.91909	0.91748	41.65027
37	0.00341	0.91587	0.91431	40.79480
38	0.00340	0.91275	0.91120	39.93261
39	0.00341	0.90965	0.90810	39.06714
40	0.00350	0.90655	0.90496	38.19904

Life Table – Whites cont'd

Age	q_x	l_x	L_x	e_x
41	0.00616	0.86794	0.86526	31.97953
42	0.00640	0.86259	0.85983	31.17460
43	0.00666	0.85707	0.85422	30.37218
44	0.00694	0.85136	0.84841	29.57241
45	0.00720	0.84545	0.84241	28.77563
46	0.00750	0.83937	0.83622	27.98069
47	0.00781	0.83307	0.82982	27.18836
48	0.00820	0.82657	0.82318	26.39839
49	0.00870	0.81979	0.81622	25.61251
50	0.00924	0.81266	0.80890	24.83291
51	0.00984	0.80515	0.80118	24.05988
52	0.01054	0.79722	0.79302	23.29405
53	0.01130	0.78882	0.78436	22.53690
54	0.01210	0.77990	0.77519	21.78876
55	0.01294	0.77047	0.76548	21.04951
56	0.01384	0.76050	0.75523	20.31895
57	0.01477	0.74997	0.74443	19.59712
58	0.01577	0.73890	0.73307	18.88335
59	0.01681	0.72725	0.72113	18.17784
60	0.01800	0.71502	0.70859	17.48005
61	0.01930	0.70215	0.69538	16.79130
62	0.02080	0.68860	0.68144	16.11191
63	0.02247	0.67428	0.66670	15.44353
64	0.02440	0.65913	0.65109	14.78698
65	0.02660	0.64305	0.63449	14.14430
66	0.02901	0.62594	0.61686	13.51716
67	0.03170	0.60778	0.59815	12.90604
68	0.03464	0.58852	0.57832	12.31219
69	0.03793	0.56813	0.55735	11.73607
70	0.04150	0.54658	0.53524	11.17909
71	0.04544	0.52389	0.51199	10.64146
72	0.04974	0.50009	0.48765	10.12425
73	0.05441	0.47521	0.46229	9.62803
74	0.05948	0.44936	0.43599	9.15325
75	0.06490	0.42263	0.40891	8.70053
76	0.07061	0.39520	0.38125	8.26968
77	0.07667	0.36730	0.35322	7.85996
78	0.08301	0.33914	0.32506	7.47108
79	0.08961	0.31098	0.29705	7.10212
80	0.09644	0.28312	0.26947	6.75196
81	0.10346	0.25581	0.24258	6.41926
82	0.11061	0.22935	0.21666	6.10233
83	0.11794	0.20398	0.19195	5.79905
84	0.12534	0.17992	0.16865	5.50760
85	0.13281	0.15737	0.14692	5.22521
86	0.14050	0.13647	0.12688	4.94885
87	0.14892	0.11730	0.10856	4.67607
88	0.15871	0.09983	0.09191	4.40681
89	0.17011	0.08399	0.07684	4.14381
90	0.18228	0.06970	0.06335	3.89074

Age	q_x	l_x	L_x	e_x
41	0.00359	0.90337	0.90175	37.33145
42	0.00370	0.90013	0.89846	36.46421
43	0.00384	0.89680	0.89507	35.59777
44	0.00396	0.89335	0.89158	34.73313
45	0.00413	0.88982	0.88798	33.86917
46	0.00427	0.88614	0.88425	33.00767
47	0.00445	0.88236	0.88039	32.14697
48	0.00467	0.87843	0.87638	31.28842
49	0.00494	0.87433	0.87217	30.43278
50	0.00521	0.87001	0.86775	29.58143
51	0.00554	0.86548	0.86308	28.73369
52	0.00590	0.86068	0.85814	27.89102
53	0.00630	0.85561	0.85291	27.05359
54	0.00671	0.85021	0.84736	26.22194
55	0.00721	0.84451	0.84147	25.39565
56	0.00777	0.83842	0.83517	24.57641
57	0.00840	0.83191	0.82842	23.76487
58	0.00906	0.82492	0.82119	22.96195
59	0.00979	0.81745	0.81345	22.16728
60	0.01050	0.80945	0.80520	21.38153
61	0.01130	0.80095	0.79642	20.60312
62	0.01210	0.79190	0.78711	19.83288
63	0.01297	0.78232	0.77724	19.06967
64	0.01401	0.77217	0.76676	18.31362
65	0.01515	0.76135	0.75559	17.56671
66	0.01647	0.74982	0.74365	16.82924
67	0.01800	0.73747	0.73084	16.10263
68	0.01974	0.72420	0.71705	15.38863
69	0.02174	0.70990	0.70218	14.68847
70	0.02404	0.69447	0.68612	14.00381
71	0.02668	0.67777	0.66873	13.33646
72	0.02961	0.65969	0.64992	12.68837
73	0.03295	0.64015	0.62961	12.06026
74	0.03671	0.61906	0.60770	11.45415
75	0.04081	0.59634	0.58417	10.87158
76	0.04531	0.57200	0.55904	10.31284
77	0.05021	0.54608	0.53238	9.77854
78	0.05545	0.51867	0.50429	9.26903
79	0.06104	0.48991	0.47495	8.78381
80	0.06694	0.46000	0.44460	8.32234
81	0.07311	0.42921	0.41352	7.88355
82	0.07954	0.39783	0.38201	7.46593
83	0.08616	0.36619	0.35041	7.06789
84	0.09291	0.33464	0.31909	6.68712
85	0.09982	0.30355	0.28840	6.32083
86	0.10693	0.27325	0.25864	5.96627
87	0.11469	0.24403	0.23003	5.62078
88	0.12360	0.21604	0.20269	5.28418
89	0.13389	0.18934	0.17666	4.95891
90	0.14484	0.16399	0.15211	4.64821

A9 Life Table-Coloureds

Male

Age	q_x	l_x	L_x	e_x
0	0.02839	1.00000	0.99225	58.77774
1	0.00736	0.97161	0.96804	59.47397
2	0.00331	0.96446	0.96287	58.91099
3	0.00221	0.96127	0.96021	58.10487
4	0.00150	0.95915	0.95843	57.23237
5	0.00110	0.95771	0.95718	56.31759
6	0.00090	0.95666	0.95623	55.37906
7	0.00080	0.95580	0.95541	54.42849
8	0.00071	0.95503	0.95469	53.47167
9	0.00070	0.95435	0.95402	52.50922
10	0.00061	0.95369	0.95340	51.54565
11	0.00060	0.95311	0.95282	50.57672
12	0.00061	0.95253	0.95225	49.60679
13	0.00071	0.95196	0.95162	48.63668
14	0.00087	0.95128	0.95087	47.67080
15	0.00116	0.95046	0.94991	46.71172
16	0.00154	0.94936	0.94862	45.76531
17	0.00200	0.94789	0.94694	44.83520
18	0.00240	0.94600	0.94486	43.92405
19	0.00280	0.94373	0.94240	43.02852
20	0.00320	0.94108	0.93958	42.14793
21	0.00360	0.93807	0.93638	41.28163
22	0.00400	0.93469	0.93283	40.42898
23	0.00431	0.93096	0.92895	39.58933
24	0.00464	0.92695	0.92479	38.75847
25	0.00494	0.92264	0.92036	37.93688
26	0.00530	0.91808	0.91565	37.12280
27	0.00570	0.91322	0.91061	36.31794
28	0.00610	0.90801	0.90524	35.52327
29	0.00650	0.90247	0.89954	34.73822
30	0.00690	0.89661	0.89351	33.96223
31	0.00739	0.89042	0.88713	33.19472
32	0.00786	0.88384	0.88037	32.43819
33	0.00830	0.87689	0.87325	31.69116
34	0.00879	0.86962	0.86579	30.95221
35	0.00920	0.86197	0.85800	30.22231
36	0.00954	0.85404	0.84997	29.49830
37	0.00989	0.84589	0.84171	28.77765
38	0.01019	0.83752	0.83326	28.06016
39	0.01049	0.82899	0.82464	27.34394
40	0.01080	0.82029	0.81586	26.62856

Female

Age	q_x	l_x	L_x	e_x
0	0.02465	1.00000	0.99300	66.23097
1	0.00610	0.97535	0.97238	66.88676
2	0.00295	0.96940	0.96797	66.29398
3	0.00176	0.96654	0.96569	65.48864
4	0.00112	0.96484	0.96431	64.60312
5	0.00090	0.96377	0.96333	63.67478
6	0.00080	0.96290	0.96251	62.73169
7	0.00071	0.96213	0.96179	61.78151
8	0.00061	0.96145	0.96115	60.82495
9	0.00050	0.96086	0.96062	59.86167
10	0.00040	0.96038	0.96019	58.89137
11	0.00040	0.96000	0.95981	57.91473
12	0.00041	0.95961	0.95942	56.93771
13	0.00050	0.95922	0.95898	55.96076
14	0.00059	0.95874	0.95846	54.98851
15	0.00070	0.95818	0.95784	54.02076
16	0.00087	0.95750	0.95709	53.05826
17	0.00104	0.95667	0.95618	52.10385
18	0.00123	0.95568	0.95509	51.15766
19	0.00146	0.95450	0.95380	50.22021
20	0.00180	0.95311	0.95225	49.29283
21	0.00220	0.95139	0.95035	48.38081
22	0.00256	0.94930	0.94808	47.48638
23	0.00290	0.94687	0.94550	46.60690
24	0.00324	0.94412	0.94259	45.74099
25	0.00360	0.94106	0.93937	44.88813
26	0.00406	0.93768	0.93577	44.04850
27	0.00446	0.93387	0.93179	43.22596
28	0.00480	0.92971	0.92748	42.41730
29	0.00510	0.92524	0.92289	41.61947
30	0.00533	0.92053	0.91807	40.83025
31	0.00553	0.91562	0.91308	40.04650
32	0.00573	0.91055	0.90794	39.26654
33	0.00593	0.90533	0.90264	38.49009
34	0.00606	0.89996	0.89723	37.71684
35	0.00610	0.89451	0.89178	36.94369
36	0.00619	0.88905	0.88630	36.16736
37	0.00620	0.88354	0.88081	35.38957
38	0.00629	0.87807	0.87530	34.60724
39	0.00636	0.87254	0.86977	33.82319
40	0.00653	0.86699	0.86416	33.03642

Life Table- Coloureds cont'd

Age	q_x	l_x	L_x	e_x
41	0.01121	0.81143	0.80688	25.91383
42	0.01174	0.80234	0.79763	25.20191
43	0.01233	0.79292	0.78803	24.49539
44	0.01296	0.78314	0.77806	23.79503
45	0.01370	0.77299	0.76769	23.10086
46	0.01450	0.76240	0.75687	22.41479
47	0.01530	0.75134	0.74560	21.73723
48	0.01614	0.73985	0.73388	21.06721
49	0.01707	0.72791	0.72169	20.40465
50	0.01810	0.71548	0.70901	19.75025
51	0.01920	0.70253	0.69579	19.10510
52	0.02044	0.68904	0.68200	18.46931
53	0.02184	0.67496	0.66759	17.84430
54	0.02340	0.66022	0.65249	17.23159
55	0.02504	0.64477	0.63669	16.63249
56	0.02684	0.62862	0.62018	16.04685
57	0.02880	0.61175	0.60294	15.47566
58	0.03084	0.59413	0.58497	14.91975
59	0.03306	0.57581	0.56629	14.37864
60	0.03541	0.55677	0.54691	13.85313
61	0.03797	0.53706	0.52686	13.34329
62	0.04067	0.51667	0.50616	12.85016
63	0.04351	0.49565	0.48487	12.37368
64	0.04651	0.47409	0.46307	11.91379
65	0.04970	0.45204	0.44081	11.47052
66	0.05297	0.42957	0.41820	11.04427
67	0.05641	0.40682	0.39535	10.63400
68	0.06000	0.38387	0.37236	10.23981
69	0.06367	0.36084	0.34935	9.86150
70	0.06751	0.33787	0.32646	9.49805
71	0.07150	0.31506	0.30379	9.14946
72	0.07557	0.29253	0.28148	8.81552
73	0.07977	0.27043	0.25964	8.49527
74	0.08410	0.24885	0.23839	8.18830
75	0.08850	0.22793	0.21784	7.89426
76	0.09296	0.20775	0.19810	7.61219
77	0.09750	0.18844	0.17926	7.34108
78	0.10207	0.17007	0.16139	7.08015
79	0.10667	0.15271	0.14457	6.82810
80	0.11134	0.13642	0.12883	6.58370
81	0.11600	0.12123	0.11420	6.34594
82	0.12066	0.10717	0.10070	6.11305
83	0.12536	0.09424	0.08833	5.88324
84	0.13009	0.08242	0.07706	5.65480
85	0.13481	0.07170	0.06687	5.42568
86	0.13974	0.06204	0.05770	5.19316
87	0.14536	0.05337	0.04949	4.95553
88	0.15222	0.04561	0.04214	4.71336
89	0.16065	0.03867	0.03556	4.46990
90	0.16976	0.03245	0.02970	4.22973

Age	q_x	l_x	L_x	e_x
41	0.00674	0.86133	0.85843	32.25039
42	0.00704	0.85552	0.85251	31.46590
43	0.00740	0.84950	0.84636	30.68550
44	0.00771	0.84321	0.83996	29.91053
45	0.00809	0.83671	0.83333	29.13900
46	0.00840	0.82994	0.82646	28.37263
47	0.00880	0.82297	0.81935	27.60874
48	0.00920	0.81573	0.81198	26.84942
49	0.00967	0.80822	0.80432	26.09408
50	0.01027	0.80041	0.79630	25.34391
51	0.01094	0.79219	0.78786	24.60162
52	0.01169	0.78353	0.77894	23.86825
53	0.01246	0.77436	0.76954	23.14469
54	0.01331	0.76472	0.75963	22.43037
55	0.01421	0.75454	0.74918	21.72616
56	0.01517	0.74382	0.73818	21.03209
57	0.01620	0.73254	0.72660	20.34829
58	0.01730	0.72067	0.71444	19.67513
59	0.01850	0.70820	0.70165	19.01270
60	0.01983	0.69510	0.68821	18.36164
61	0.02124	0.68132	0.67408	17.72306
62	0.02280	0.66684	0.65924	17.09685
63	0.02441	0.65164	0.64369	16.48409
64	0.02614	0.63573	0.62742	15.88399
65	0.02796	0.61911	0.61046	15.29695
66	0.02990	0.60181	0.59281	14.72255
67	0.03190	0.58381	0.57450	14.16091
68	0.03404	0.56519	0.55557	13.61105
69	0.03640	0.54595	0.53601	13.07310
70	0.03894	0.52608	0.51583	12.54805
71	0.04175	0.50559	0.49503	12.03624
72	0.04487	0.48448	0.47361	11.53886
73	0.04824	0.46274	0.45158	11.05740
74	0.05184	0.44042	0.42900	10.59252
75	0.05570	0.41759	0.40596	10.14434
76	0.05977	0.39433	0.38254	9.71322
77	0.06404	0.37076	0.35889	9.29886
78	0.06844	0.34702	0.33514	8.90091
79	0.07303	0.32327	0.31146	8.51813
80	0.07770	0.29966	0.28802	8.14986
81	0.08254	0.27637	0.26497	7.79433
82	0.08747	0.25356	0.24247	7.45058
83	0.09254	0.23138	0.22068	7.11680
84	0.09770	0.20997	0.19971	6.79157
85	0.10291	0.18946	0.17971	6.47281
86	0.10829	0.16996	0.16076	6.15796
87	0.11425	0.15156	0.14290	5.84506
88	0.12120	0.13424	0.12610	5.53452
89	0.12940	0.11797	0.11034	5.22888
90	0.13820	0.10270	0.09561	4.93177

A10 Life table – Indian/Asians

Male

Age	q_x	l_x	L_x	e_x
0	0.02241	1.00000	0.99388	63.52893
1	0.00674	0.97759	0.97429	63.96860
2	0.00285	0.97100	0.96961	63.39946
3	0.00165	0.96823	0.96743	62.57923
4	0.00094	0.96663	0.96618	61.68183
5	0.00066	0.96572	0.96541	60.73950
6	0.00060	0.96509	0.96480	59.77918
7	0.00060	0.96451	0.96422	58.81477
8	0.00060	0.96393	0.96364	57.84978
9	0.00060	0.96335	0.96306	56.88421
10	0.00060	0.96277	0.96248	55.91806
11	0.00064	0.96220	0.96189	54.95134
12	0.00070	0.96158	0.96124	53.98630
13	0.00076	0.96090	0.96054	53.02376
14	0.00094	0.96018	0.95972	52.06362
15	0.00121	0.95927	0.95869	51.11223
16	0.00151	0.95811	0.95739	50.17346
17	0.00180	0.95667	0.95581	49.24849
18	0.00206	0.95495	0.95396	48.33640
19	0.00227	0.95298	0.95190	47.43507
20	0.00241	0.95082	0.94968	46.54169
21	0.00250	0.94853	0.94734	45.65284
22	0.00260	0.94616	0.94493	44.76601
23	0.00277	0.94370	0.94239	43.88140
24	0.00297	0.94109	0.93969	43.00176
25	0.00317	0.93830	0.93681	42.12822
26	0.00337	0.93532	0.93375	41.26046
27	0.00364	0.93218	0.93048	40.39815
28	0.00387	0.92878	0.92699	39.54398
29	0.00414	0.92519	0.92327	38.69553
30	0.00450	0.92136	0.91928	37.85439
31	0.00490	0.91721	0.91496	37.02324
32	0.00524	0.91272	0.91033	36.20309
33	0.00553	0.90793	0.90542	35.39122
34	0.00576	0.90291	0.90031	34.58535
35	0.00600	0.89771	0.89502	33.78277
36	0.00629	0.89232	0.88952	32.98367
37	0.00656	0.88671	0.88380	32.18934
38	0.00683	0.88089	0.87788	31.39854
39	0.00706	0.87487	0.87179	30.61113
40	0.00734	0.86870	0.86551	29.82518

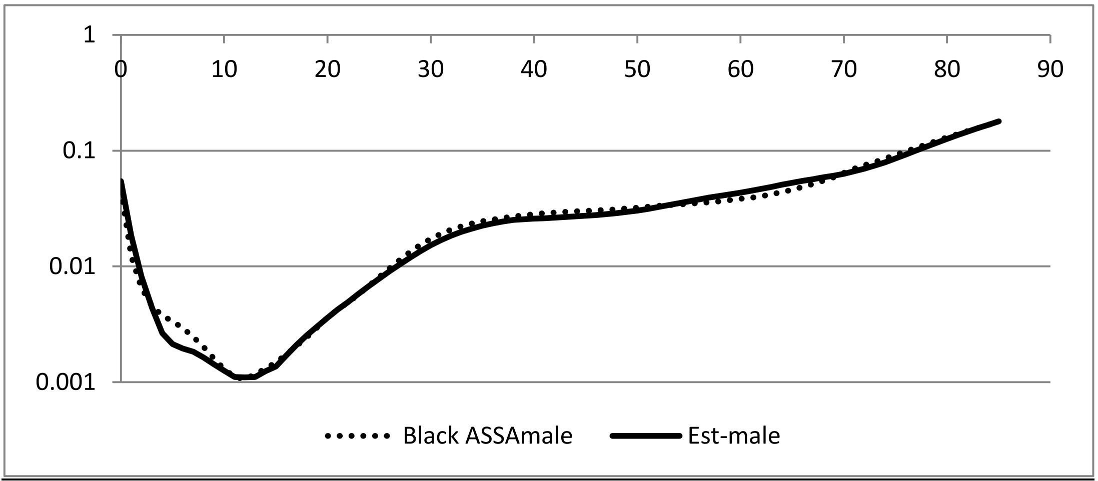
Female

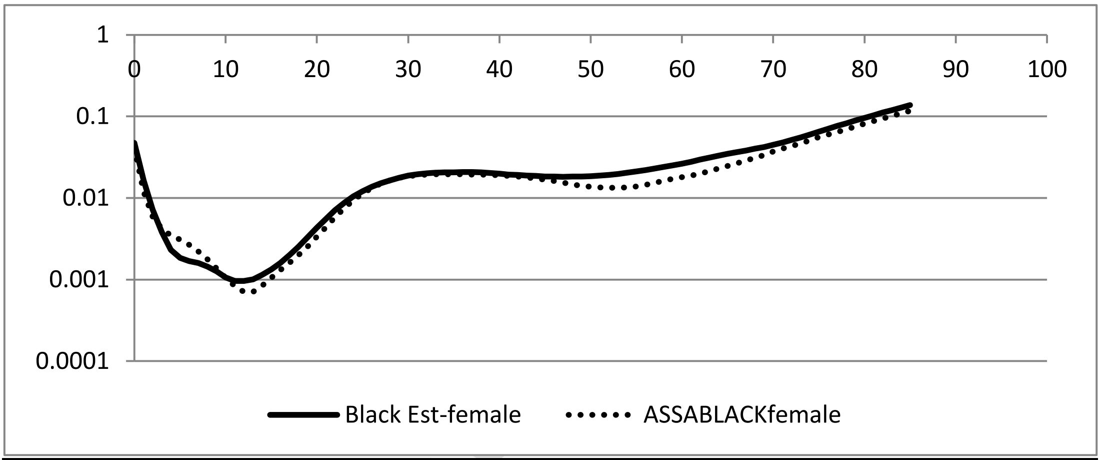
Age	q_x	l_x	L_x	e_x
0	0.01615	1.00000	0.99541	71.07419
1	0.00626	0.98385	0.98077	71.22924
2	0.00160	0.97769	0.97690	70.67508
3	0.00124	0.97612	0.97552	69.78754
4	0.00096	0.97491	0.97444	68.87368
5	0.00076	0.97398	0.97361	67.93927
6	0.00061	0.97324	0.97294	66.99045
7	0.00051	0.97264	0.97240	66.03092
8	0.00044	0.97215	0.97194	65.06425
9	0.00040	0.97172	0.97153	64.09278
10	0.00040	0.97133	0.97114	63.11822
11	0.00040	0.97094	0.97075	62.14328
12	0.00046	0.97055	0.97033	61.16795
13	0.00051	0.97011	0.96986	60.19577
14	0.00060	0.96962	0.96933	59.22613
15	0.00069	0.96904	0.96870	58.26139
16	0.00074	0.96836	0.96801	57.30136
17	0.00081	0.96765	0.96726	56.34352
18	0.00091	0.96686	0.96643	55.38870
19	0.00104	0.96599	0.96548	54.43860
20	0.00116	0.96498	0.96442	53.49485
21	0.00130	0.96386	0.96324	52.55630
22	0.00146	0.96261	0.96191	51.62407
23	0.00163	0.96121	0.96042	50.69873
24	0.00183	0.95964	0.95876	49.78086
25	0.00203	0.95788	0.95690	48.87137
26	0.00223	0.95593	0.95486	47.96993
27	0.00243	0.95379	0.95263	47.07618
28	0.00263	0.95147	0.95022	46.18979
29	0.00283	0.94897	0.94762	45.31042
30	0.00297	0.94628	0.94487	44.43775
31	0.00314	0.94347	0.94199	43.56848
32	0.00324	0.94051	0.93898	42.70422
33	0.00331	0.93746	0.93591	41.84147
34	0.00340	0.93436	0.93277	40.97870
35	0.00344	0.93118	0.92958	40.11680
36	0.00350	0.92798	0.92635	39.25362
37	0.00356	0.92473	0.92308	38.38973
38	0.00361	0.92144	0.91977	37.52504
39	0.00370	0.91811	0.91641	36.65912
40	0.00374	0.91471	0.91300	35.79340

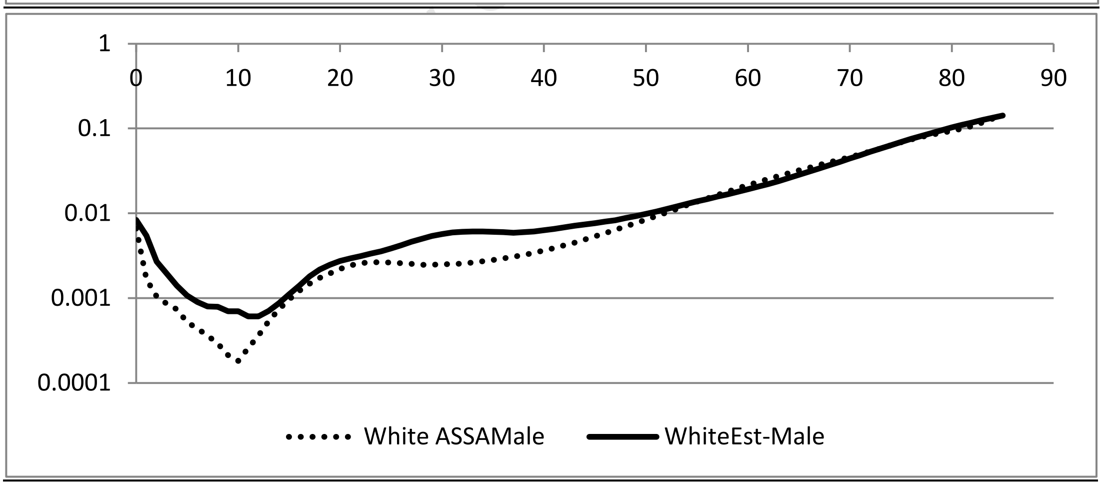
Life Table- Indian/Asians cont'd

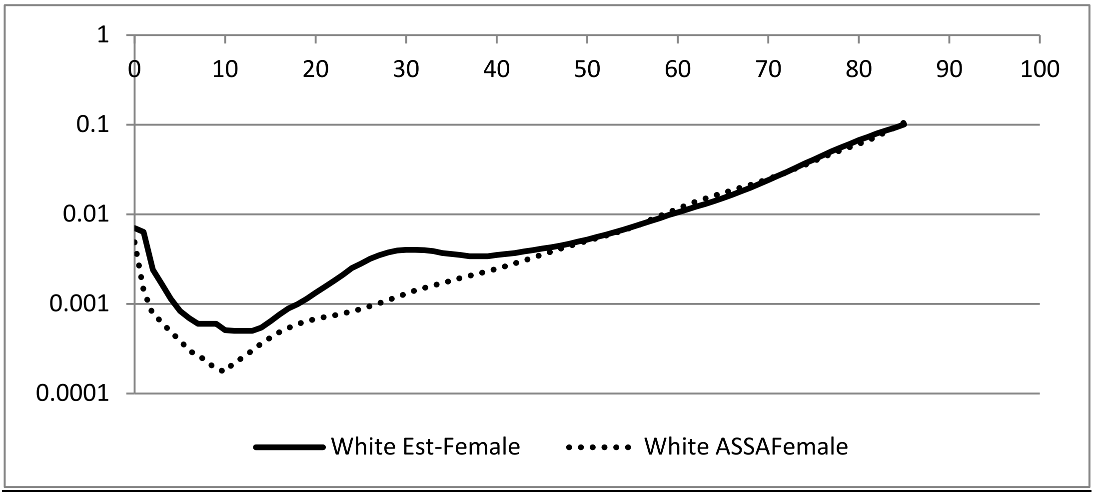
Age	q_x	l_x	L_x	e_x		Age	q_x	l_x	L_x	e_x
41	0.00770	0.86232	0.85900	29.04207		41	0.00381	0.91129	0.90956	34.92596
42	0.00814	0.85568	0.85220	28.26355		42	0.00390	0.90782	0.90605	34.05756
43	0.00864	0.84872	0.84505	27.49144		43	0.00399	0.90428	0.90248	33.18895
44	0.00920	0.84138	0.83751	26.72673		44	0.00413	0.90067	0.89881	32.31996
45	0.00974	0.83364	0.82958	25.97025		45	0.00427	0.89695	0.89504	31.45202
46	0.01026	0.82552	0.82129	25.22082		46	0.00447	0.89312	0.89113	30.58465
47	0.01076	0.81705	0.81266	24.47704		47	0.00474	0.88913	0.88702	29.71963
48	0.01130	0.80826	0.80369	23.73780		48	0.00501	0.88492	0.88270	28.85884
49	0.01190	0.79913	0.79437	23.00339		49	0.00531	0.88048	0.87815	28.00159
50	0.01260	0.78962	0.78464	22.27440		50	0.00570	0.87581	0.87331	27.14836
51	0.01340	0.77967	0.77445	21.55226		51	0.00614	0.87082	0.86814	26.30112
52	0.01427	0.76922	0.76373	20.83819		52	0.00666	0.86547	0.86259	25.46056
53	0.01531	0.75825	0.75244	20.13255		53	0.00726	0.85971	0.85659	24.62787
54	0.01650	0.74664	0.74048	19.43776		54	0.00796	0.85347	0.85007	23.80428
55	0.01777	0.73432	0.72780	18.75548		55	0.00874	0.84668	0.84298	22.99123
56	0.01921	0.72127	0.71435	18.08568		56	0.00964	0.83927	0.83523	22.18958
57	0.02080	0.70742	0.70006	17.43009		57	0.01066	0.83118	0.82675	21.40074
58	0.02247	0.69270	0.68492	16.78972		58	0.01180	0.82232	0.81747	20.62590
59	0.02434	0.67714	0.66890	16.16410		59	0.01304	0.81262	0.80732	19.86622
60	0.02636	0.66066	0.65195	15.55491		60	0.01444	0.80202	0.79623	19.12213
61	0.02854	0.64325	0.63407	14.96247		61	0.01600	0.79044	0.78412	18.39501
62	0.03090	0.62489	0.61523	14.38738		62	0.01767	0.77779	0.77092	17.68598
63	0.03337	0.60558	0.59547	13.83019		63	0.01954	0.76405	0.75659	16.99506
64	0.03601	0.58537	0.57483	13.29032		64	0.02160	0.74912	0.74103	16.32383
65	0.03881	0.56429	0.55334	12.76808		65	0.02384	0.73294	0.72420	15.67317
66	0.04177	0.54239	0.53107	12.26341		66	0.02626	0.71547	0.70607	15.04376
67	0.04491	0.51974	0.50807	11.77614		67	0.02889	0.69668	0.68661	14.43595
68	0.04825	0.49640	0.48442	11.30635		68	0.03169	0.67655	0.66583	13.85056
69	0.05180	0.47245	0.46021	10.85419		69	0.03464	0.65511	0.64376	13.28751
70	0.05547	0.44797	0.43555	10.41984		70	0.03780	0.63241	0.62046	12.74639
71	0.05940	0.42313	0.41056	10.00237		71	0.04113	0.60851	0.59599	12.22749
72	0.06347	0.39799	0.38536	9.60245		72	0.04460	0.58348	0.57047	11.73057
73	0.06771	0.37273	0.36012	9.21930		73	0.04821	0.55746	0.54402	11.25484
74	0.07211	0.34750	0.33497	8.85255		74	0.05197	0.53058	0.51680	10.79957
75	0.07665	0.32244	0.31008	8.50164		75	0.05581	0.50301	0.48897	10.36414
76	0.08131	0.29772	0.28562	8.16588		76	0.05980	0.47494	0.46074	9.94718
77	0.08610	0.27352	0.26174	7.84435		77	0.06381	0.44654	0.43229	9.54806
78	0.09100	0.24997	0.23859	7.53627		78	0.06794	0.41804	0.40384	9.16475
79	0.09596	0.22722	0.21632	7.24067		79	0.07213	0.38964	0.37559	8.79636
80	0.10096	0.20542	0.19505	6.95615		80	0.07636	0.36153	0.34773	8.44132
81	0.10600	0.18468	0.17489	6.68115		81	0.08064	0.33393	0.32046	8.09784
82	0.11104	0.16510	0.15594	6.41404		82	0.08491	0.30700	0.29397	7.76429
83	0.11614	0.14677	0.13825	6.15277		83	0.08924	0.28093	0.26840	7.43832
84	0.12120	0.12972	0.12186	5.89556		84	0.09360	0.25586	0.24389	7.11817
85	0.12627	0.11400	0.10680	5.63969		85	0.09792	0.23191	0.22056	6.80160
86	0.13156	0.09961	0.09305	5.38246		86	0.10246	0.20921	0.19849	6.48562
87	0.13752	0.08650	0.08055	5.12210		87	0.10750	0.18777	0.17768	6.16894
88	0.14473	0.07461	0.06921	4.85910		88	0.11359	0.16758	0.15807	5.85178
89	0.15348	0.06381	0.05891	4.59672		89	0.12089	0.14855	0.13957	5.53758
90	0.16291	0.05402	0.04962	4.33948		90	0.12874	0.13059	0.12218	5.23031

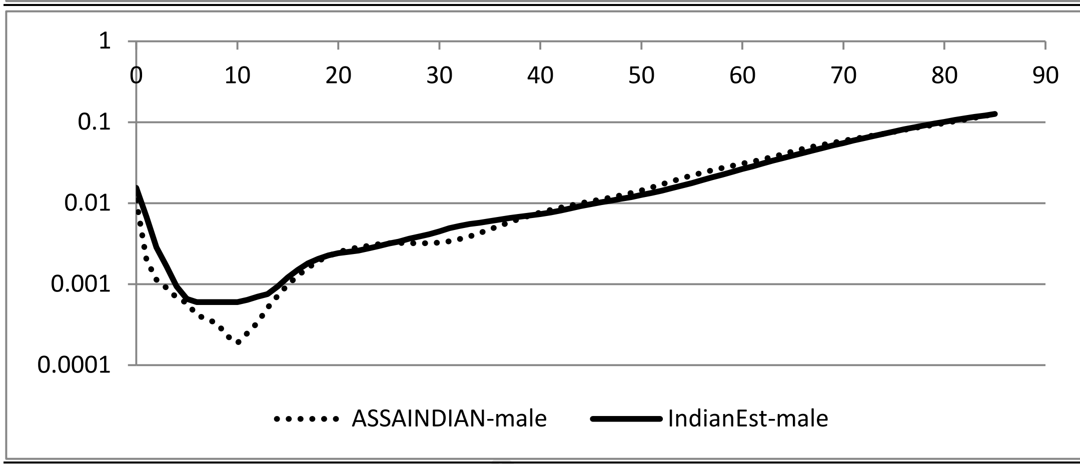
A11 Comparison between ASSA2008 life tables and life tables from this study

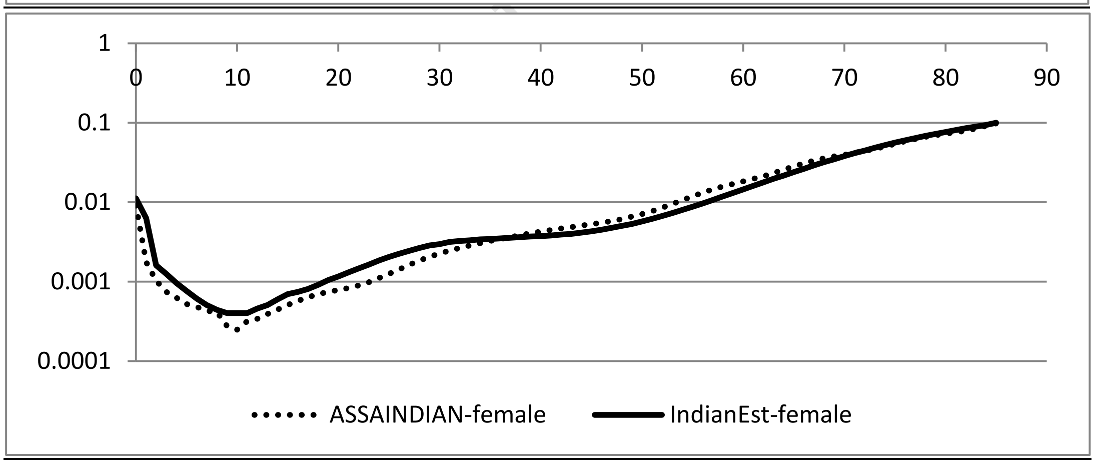


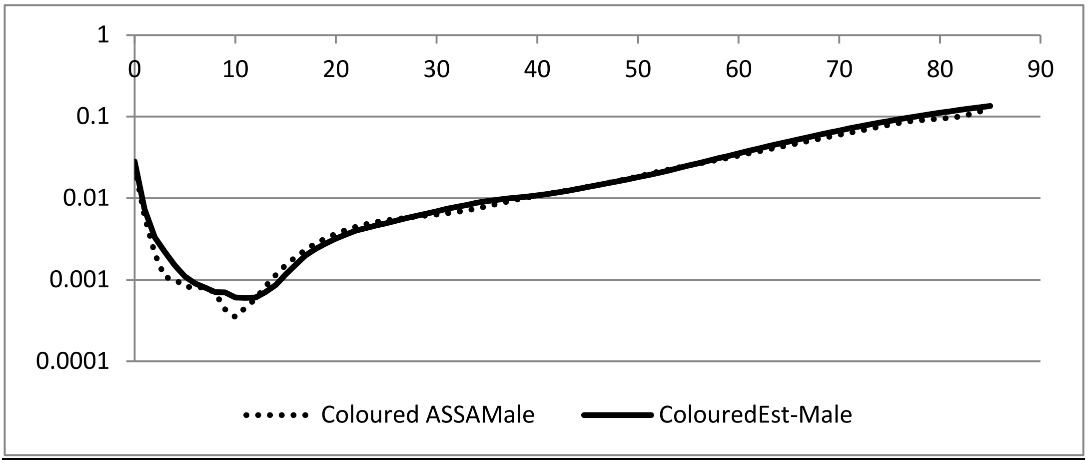


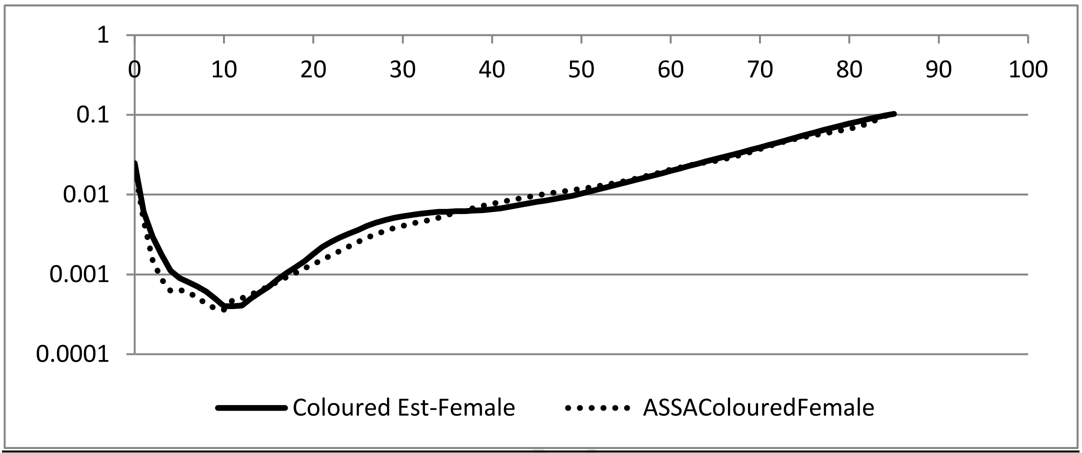



1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
White Est-Female
White ASSAFemale

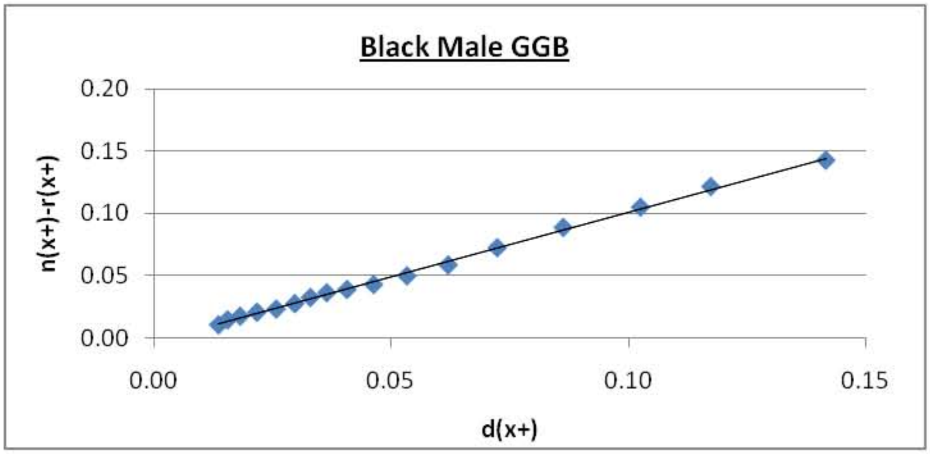

1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
ASSAINDIAN-male
IndianEst-male

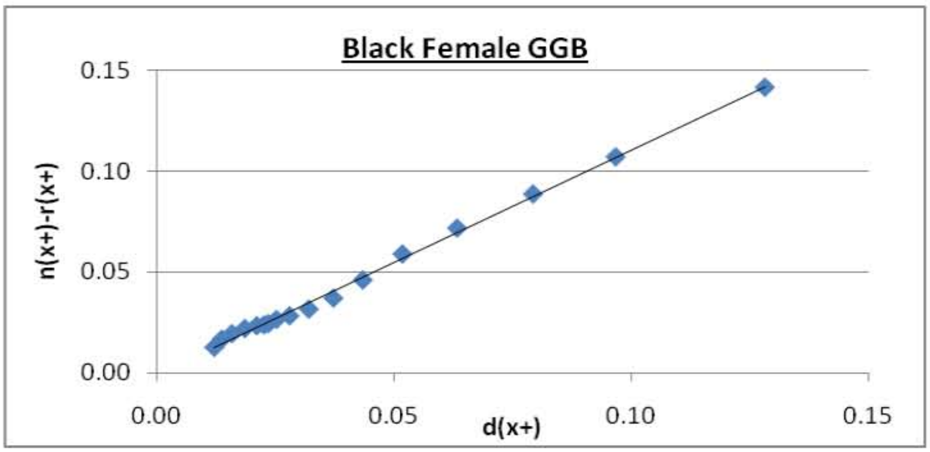

1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
ASSAINDIAN-female
IndianEst-female

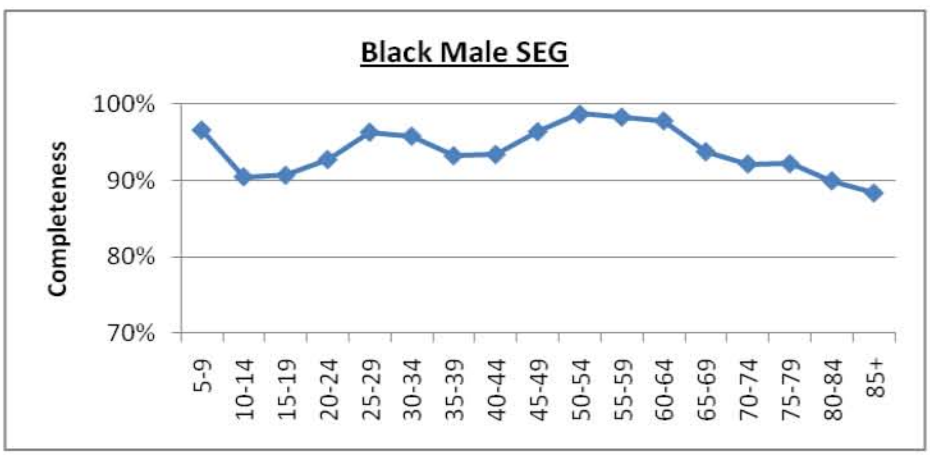

1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90
Coloured ASSAMale
ColouredEst-Male

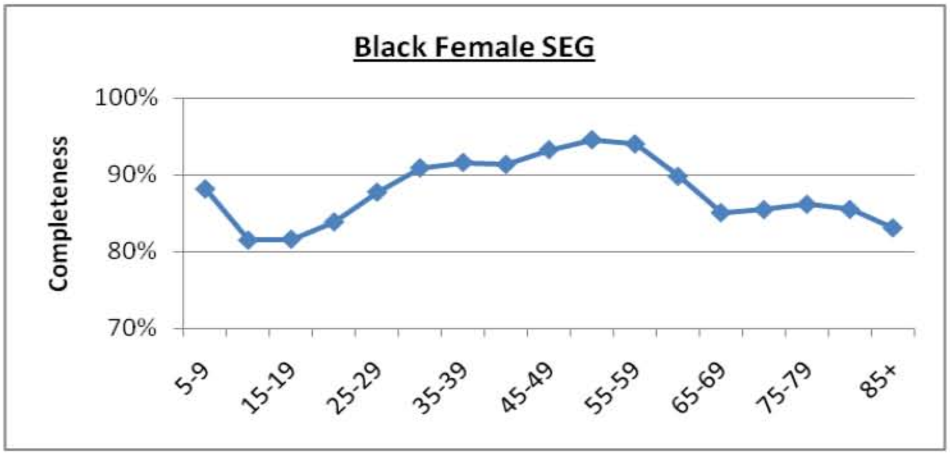

1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90 100
Coloured Est-Female
ASSAColouredFemale

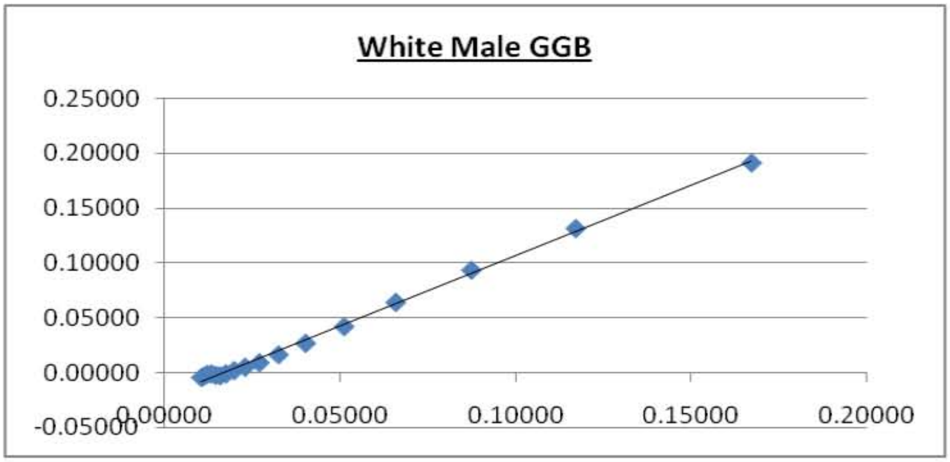
A12 Diagnostic plots for the population groups

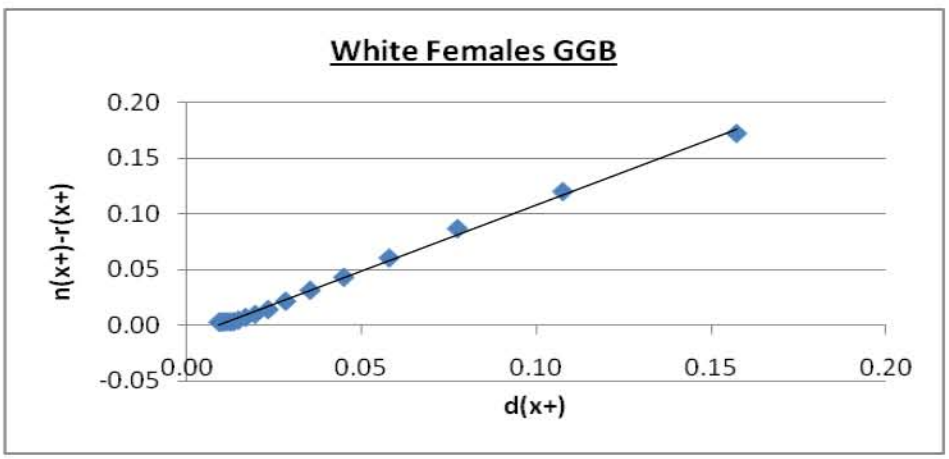


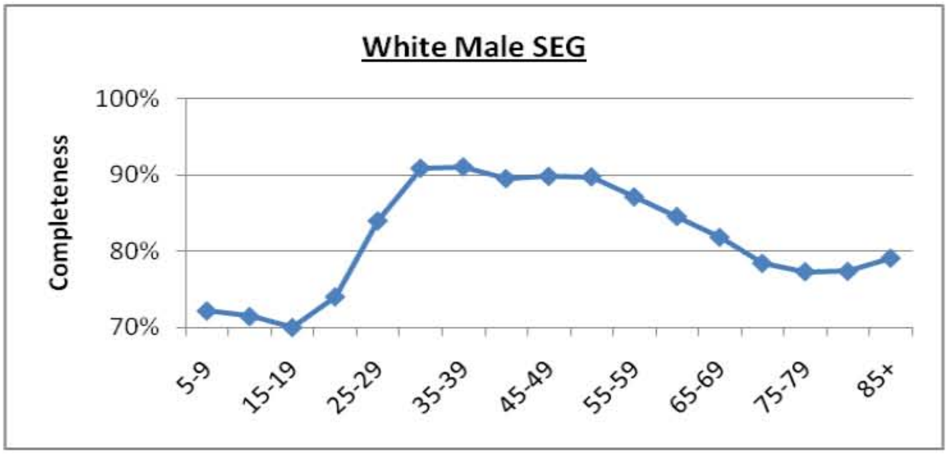


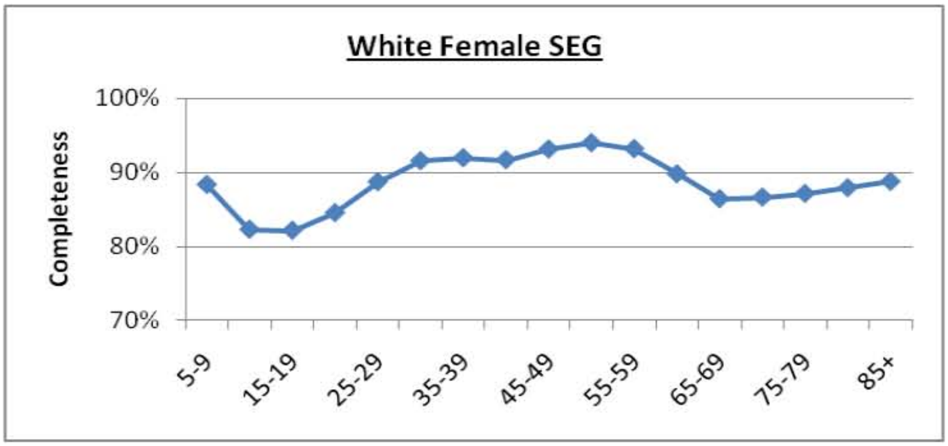


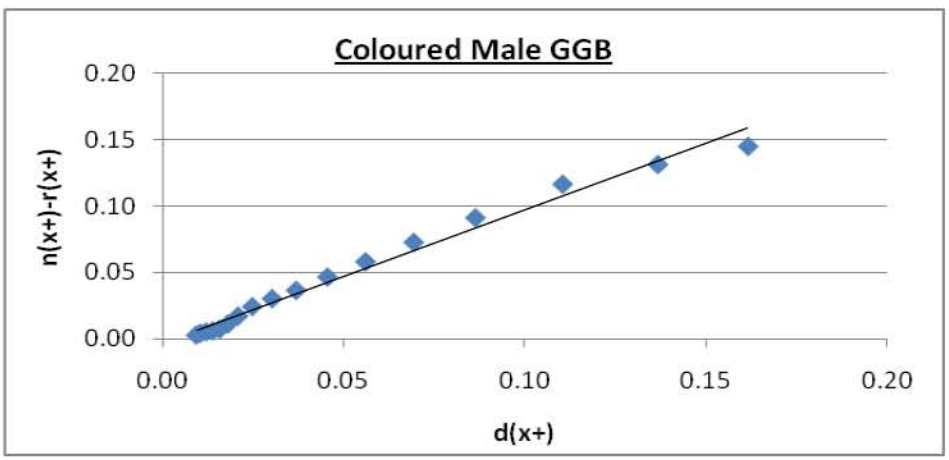

Black Female SEG
Completeness
100%
90%
80%
70%
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85+

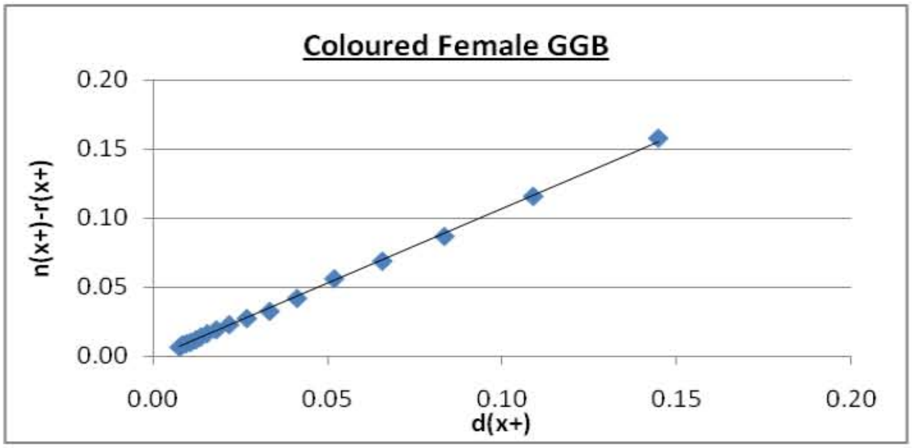

White Male GGB
0.25000
0.20000
0.15000
0.10000
0.05000
0.00000
-0.05000
0.00000
0.05000
0.10000
0.15000
0.20000

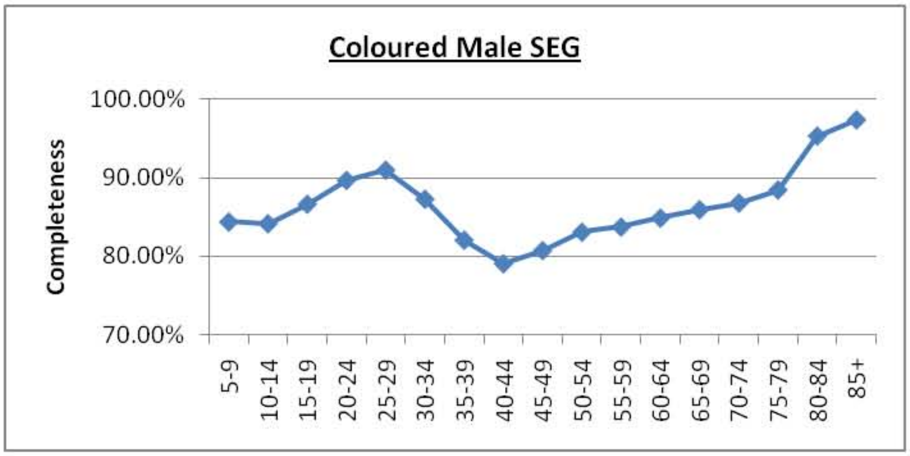

White Females GGB
n(x+)-r(x+)
0.20
0.15
0.10
0.05
0.00
-0.05
0.00
0.05
0.10
0.15
0.20
d(x+)

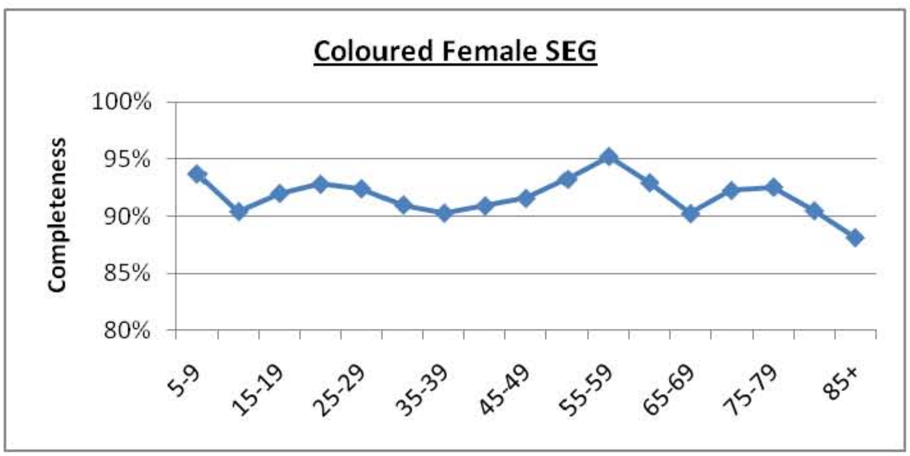

White Male SEG
Completeness
100%
90%
80%
70%
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85+

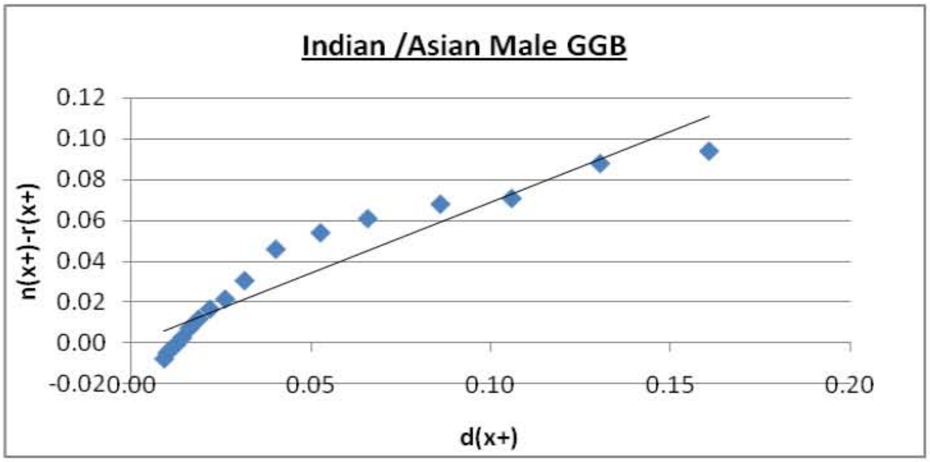

White Female SEG
Completeness
100%
90%
80%
70%
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85+

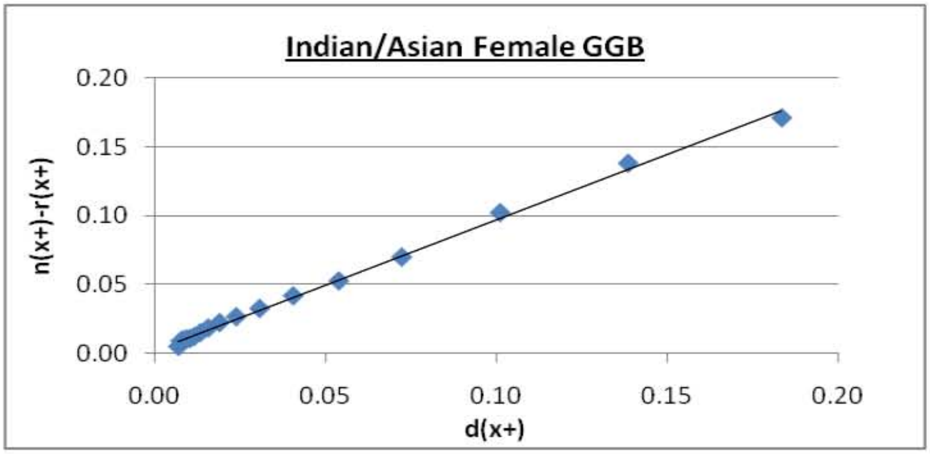

Coloured Male GGB
n(x+)-r(x+)
0.20
0.15
0.10
0.05
0.00
0.00
0.05
0.10
0.15
0.20
d(x+)

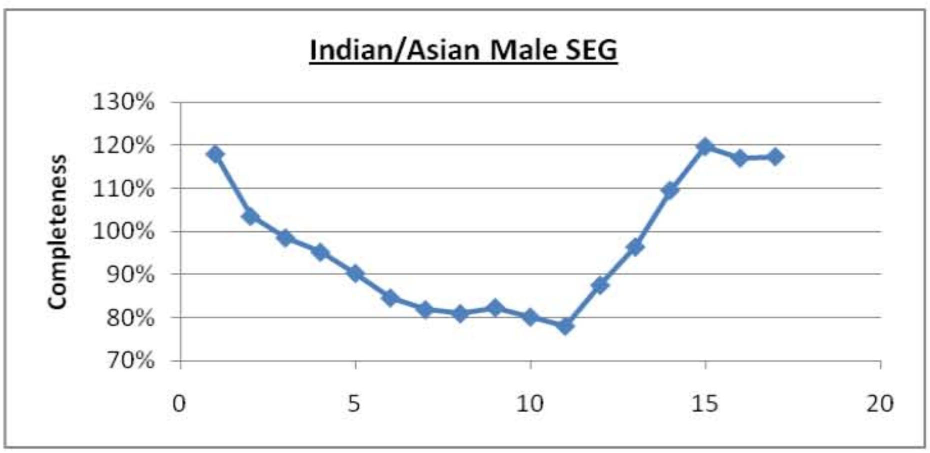

Coloured Female GGB
0.20
0.15
0.10
0.05
0.00
n(x+)-r(x+)
0.00
0.05
0.10
0.15
0.20
d(x+)

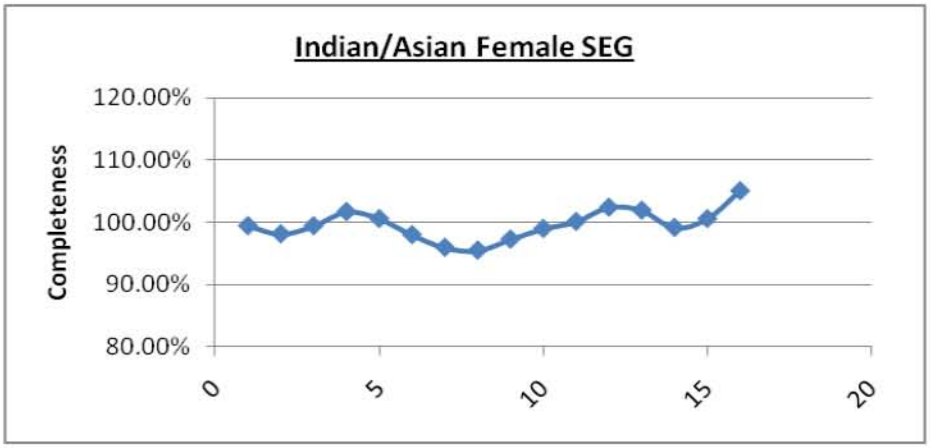

Coloured Male SEG
100.00%
90.00%
80.00%
70.00%
Completeness
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85+


Coloured Female SEG
100%
95%
90%
85%
80%
Completeness
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85+


Indian /Asian Male GGB
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.020.00
0.05
0.10
0.15
0.20
n(x+)-r(x+)
d(x+)


Indian/Asian Female GGB
0.20
0.15
0.10
0.05
0.00
0.00
0.05
0.10
0.15
0.20
n(x+)-r(x+)
d(x+)


Indian/Asian Male SEG
130%
120%
110%
100%
90%
80%
70%
0
5
10
15
20
Completeness


Indian/Asian Female SEG
Completeness
120.00%
110.00%
100.00%
90.00%
80.00%
0
5
10
15
20