Sensitivity Analysis with Simulated Data Errors: Synthetic Extinct Generations Method

By

Martin Ruzvidzo

A dissertation submitted to the faculty of Commerce of the University of Cape Town in partial fulfillment of the requirements for the Degree of Master of Philosophy in Demography
August 2009
Center for Actuarial Research (CARe)

, Thesis Presented for the Degree of
Master of Philosophy in Demography
University of Cape Town
August 2009

Abstract

This study develops the key components of the Synthetic Extinct Generations (SEG+delta) method in trying to answer research questions raised by Dorrington and Timeaus (2008) and in the process investigates the different sets of combinations of the key components of the SEG+delta method when applied to the 23 error scenarios used by Hill and Choi (2004). In addition, the study determines the pattern of estimates of ${}_{45}q_{15}$ per set of combinations, the combination that results in the best estimate of ${}_{45}q_{15}$ per scenario and per combination of scenarios and the best combination that result in best estimate of ${}_{45}q_{15}$ across all 23 error scenarios.

The current study assesses the errors in age reported in censuses by comparing the weighted average of the ages of seven countries in the sub-Saharan African region to the age error pattern used by Hill and Choi (2004). These findings suggest that there is no significant difference (except of the zig-zag pattern in age errors at older ages in the Hill and Choi scenario) in general pattern of age errors of the sub-Saharan African region and the pattern used by Hill and Choi (2004).

The age distribution of the migrants is determined by subtracting the projected population without migration from the given population of the UN Population. The resultant migration patterns from the seven countries were compared to the migration patterns applied by Hill and Choi (2004). In countries where the migration pattern was determined, the findings suggest that the general pattern of migration of the sub-Saharan African region is not significantly different from the pattern applied by Hill and Choi, except for a great deal more random fluctuation in the rates by age and low rates of migration at the older ages found in the sub-Saharan African region.

The SEG+delta method is applied using the set of combinations of key components, and the estimate of ${}_{45}q_{15}$ is determined. Performance of the SEG+delta method is assessed using the per cent difference of ${}_{45}q_{15}$ from the true value. The pattern of the per cent difference of ${}_{45}q_{15}$ from the true value was seen to differ by type of scenario (migration, age misreporting, and census coverage). It was also seen to differ by the combination of key components of the SEG+delta method. In addition it is established that each scenario had a different combination which results in the best estimate of ${}_{45}q_{15}$.

Patterns observed for these combinations of key components of the SEG+delta method suggest that the best estimates arose in the no error scenario, and the scenarios with errors the method was designed to deal with, 6, 8 and 11. The worst estimates are found in scenarios involving migration, with estimates from emigration scenarios (17 and 21) better than estimates from immigration scenarios (20 and 22). Although the SEG+delta method performed differently for different combinations of key components, a combination which results in best estimate of ${}_{45}q_{15}$ across all scenarios was determined. The combination which result in best estimate of ${}_{45}q_{15}$ across all scenarios was a combination of N_x calculated using the method suggested by Feeney (United Nations, 2002), using e_x estimated by the iterative method, using c_{x+} over the age range 15-69 to determine completeness for an open age interval of 85+.

Conclusions from this study are that some combinations work better than others in specific error scenarios. The combination which results in best estimate of mortality differs by error scenario but is the same for similar error scenarios. Most combinations have similar estimates of ${}_{45}q_{15}$ per scenario. The performance of the combinations of the key components of the SEG+delta method was seen to be the same for the results of the research done by Hill and Choi (2004), and by Dorrington and Timaeus (2008).

Acknowledgements

First and foremost I would like to thank my supervisor Professor Rob Dorrington, this thesis would have been difficult to read than it is now with my language skills being what they are and tested to the limit, thanks for all the ideas and insightfulness. I would like to thank all my friends who helped in all their wonderful ways, special thanks to T and R. I also pay tribute to all the people at Data First; I would like to thank them, for providing the census data for sub-Saharan African countries.

I would like to acknowledge the tireless efforts of my supervisor, Professor Rob Dorrington, for his invaluable contribution to this study. I would like to thank him for the time and effort invested in me, the patience and understanding. I would also like to thank him for the extra hours put in to ensure this project was completed.

I wish to thank the Andrew Mellon Foundation and the University of Cape Town, Postgraduate Funding Office for making a financial contribution to my studies.

TABLE OF CONTENTS

List of Figures

It is essential to approximate the life expectancy at the age of the open interval when applying the SEG method. The method used by Hill and Choi to determine the life expectancy at the age of the open interval is based on the Princeton Life Tables which are far from representative of the less developed countries, especially African populations with HIV deaths. While Hill and Choi are concerned about the performance of the methods in different error scenarios, Dorrington and Timaeus (2008) are even more concerned about the correct application of the methods and in trying to test the methods in less developed countries with deficient data and with HIV mortality. In this regard, Dorrington and Timaeus (2008) pointed out that the version of the SEG method applied by Hill and Choi (2004)) did not include an adaptation suggested by Bennett and Horiuchi (1981) to allow for differential coverage of one census relative to the other.

Furthermore, the corrected version of the SEG method was shown to give better estimates of ${}_{45}q_{15}$ compared to the one applied by Hill and Choi (2004). It was shown that the conclusions reached by Hill and Choi (2004) which supported the use of a combination of the GGB and SEG methods (GGB+SEG) differ from the conclusions reached by Dorrington and Timaeus (2008) which supported the adaptation of the SEG method.

Arising from the sensitivity analysis of Dorrington and Timaeus (2008) and Hill and Choi (2004) the following areas were identified as priorities for this study in order to improve the performance of the adjusted SEG method:

- What is the best method of estimating the number in the population aged x (N_x)?
- What is the best method of estimating the life expectancy at the age of the open age interval?
- What is the best age to use for the open interval?
- What is best method of determining completeness of reporting of deaths (c) from age specific measures of completeness (c_x)?
- What is the best combination of the components of the SEG+delta method that gives best estimates of ${}_{45}q_{15}$ per scenario and overall?

1.2 Statement of the Problem

The purpose of this study is to apply the SEG+delta method for estimating adult mortality to the error scenarios simulated by Hill and Choi (2004) and to investigate the extent to which the conclusions of Hill and Choi (2004), and Dorrington and Timaeus (2008) are the result of the particular combinations of components of the method SEG method they used. In addition the research considers what changes may be required to adapt the Hill and Choi (2004) data set of scenarios to be more appropriate to sub-Saharan African setting.

1.3 Justification of the study

The study will help in investigating the performance of different combinations of key components of the SEG+delta method when applied to the 23 scenarios applied by Hill and Choi (2004); thus determining which combinations works better for each scenario and the best combination that give better estimates of ${}_{45}q_{15}$ across all scenarios. In addition, this study examines the errors (migration and age exaggeration) found in less developed countries (sub-Saharan African region) and determine whether these differ to the errors used by Hill and Choi (2004).

1.4 Structure of the thesis

The thesis is structured as follows. Chapter 2 reviews past studies on the development of death distribution methods and the sensitivity analyses that have been carried out on death distribution methods. Chapter 3 describes the development of the key components of the death distribution method under study (Synthetic Extinct Generations methods) as applied by other researchers and the analytic framework to be used to investigate the performance of the methods under study. Results of the performance of the SEG+delta method are determined for the 23 different simulated error scenarios and presented in Chapter 4. Chapter 5 presents the results of the African error scenarios. Chapter 6 evaluates the results and then wraps up with a discussion of the results and conclusion.

Chapter 2 Literature Review

This chapter reviews past studies of death distribution methods that have been conducted in different populations

2.1 Death Distribution Methods

Death distribution methods can be reviewed across two intersecting dimensions: the development of the Generalized Growth Balance (GGB) and Synthetic Extinct Generations (SEG) methods; and the sensitivity analysis of these methods to data likely to be available in less developed countries.

Research has shown that most developing countries have poor or non-existent vital registration systems and have varying census coverage, which in turn may lead to difficulties in estimating mortality rates (Hill, 1987). The above issues led to the development of methods for assessing the completeness of the census coverage and death reporting (Hill, 1987). The GGB and SEG methods were seen as methods which could assist in the assessment of the completeness of census coverage and the completeness of death reporting.

The GGB method was developed from Growth Balance (GB) method proposed by Brass (1975) who used the balance equation for any open ended age group *a*+ of a population closed to migration which states that the entry rate is equal to the growth rate of the population plus the departure rate of the of the open ended age group, $b(a+) = r(a+) + d(a+)$. In order to estimate the completeness of the reporting of deaths the method assumes further that the population is stable and hence that $r(a+)=r$ for all ages *a*. The rationale of this approach is described more fully when the GGB method is described.

The method was developed to estimate the completeness of death reporting relative to coverage of an estimate of the population. In addition, this method assumed that the completeness of death reporting is constant at all adult ages.

Preston, Coale *et al.* (1980) argued that the GB method was vulnerable to age exaggeration. It has also been shown to be sensitive to the effects of destabilization resulting from a rapid mortality decline (Martin, 1980) and to significant unknown migration (Preston, Coale *et al.*, 1980). This led to Preston and Coale *et al* (1980) developing an alternative method in which the number of people in the population at each age is estimated from the number of deaths by age, and compared to that from the census to determine the completeness of reporting of deaths.

2.1.1 Preston and Coale method

The method by Preston and Coale derives its relationship from the stable population theory that relates the population of age x at time t to the deaths over age x at time t (Preston, Coale *et al.*, 1980). In addition, this method is based on the idea that the number of persons at a particular age at a point in time will be equivalent to the total number of deaths arising from this population from that time until the last survivor of the age cohort has died. Given the deaths by age $D(x)$ and the population by age $N(x)$ at the mid point of the period, estimates the size of the population from the registered deaths aged x at the last birthday, $\hat{N}(x)$ and compares these to the population $N(x)$, with the ratios $\frac{\hat{N}(x)}{N(x)}$ indicating the relative completeness of death registration.

Preston and Coale method assumes a stable and closed population and that the degree of completeness of death registration is more or less the same at all adult ages. However, the method by Preston and Coale was seen by Martin (1980) to be fairly robust to departures from stability especially to recent declines in fertility and a gradual change in mortality. It was evident that the method was more sensitive to age-misreporting and differential under-reporting of deaths by age (Martin, 1980). Thus, Martin (1980) relaxed the assumption of stability, allowing the growth rate to be age specific, namely $r(a+)$, which could be estimated from two successive censuses, but did not allow for differential completeness of enumeration between two censuses.

2.1.2 Preston and Hill method

Preston and Hill (1980) and Brass (1979) suggested methods of estimating the relative coverage of two census enumeration and the completeness of death reporting relative to one census or the other. These methods involve comparison of the number of survivors to the number at the previous census by cohort and intercensal cohort deaths to estimate the relative completeness of enumeration of the one census to the other (Luther, 1983).

The stable population assumption is often inappropriate in many contexts because of changing fertility and mortality levels and non-negligible levels of migration. As a result, Preston and Hill (1980) estimated completeness of death registration without assuming that the population was stable. Their method is based on the relationship of the size of the cohort in the second census to the size of the same cohort in the first census and to the intercensal deaths occurring to the members of the cohort. The assumption underlying this strategy was that the completeness of death registration relative to the two censuses will be constant over age (Hill, 2001). Importantly, this method estimates both the completeness of coverage of deaths relative to population enumerations and the potential change in coverage between two census enumerations.

However, the method by Preston and Hill (1980) is very sensitive to age exaggeration which tends to distort the estimated completeness of death reporting (Preston, Heuveline and Guillot, 2001). The method by Preston and Hill (1980) is less sensitive to age misreporting, but is more complicated to apply than the GGB method.

2.2 Generalized Growth Balance

Hill (1987) generalised the GB method by relaxing the stability assumption. Applying the balance equation to the population aged *x* and over one gets

$D(x+) = P_1(x+) + N(x) - P_2(x+)$, where $P_1(x+)$ and $P_2(x+)$ denote the true number of persons aged x and over in the population at the beginning and ending of some time period, respectively, $D(x+)$ denotes the true number of deaths during the period to persons aged x and over, and $N(x)$ denotes the true number of persons reaching exact age x during the period. For *x* above zero, $N(x)$ may be approximated by interpolating between the numbers in each census as follows:

$N(x) = t0.2[P_1(x-5,5)P_2(x,5)]^{0.5}$,

where *t* denotes the length of the intercensal period, $P_1(x-5,5)$ is the number of persons aged *x*-5 to *x* at the first census, i.e. those who could reach exact age *x* during the five years following the first census, and, $P_2(x,5)$ is the number of people aged *x* to *x+5* at the time of the second census, i.e. the number of survivors who reached exact age *x* during the five years preceding the second census (Hill, 1987).

The method requires estimates of the population at two time points and the reported number of deaths by age over the period in between these time points, with age specific growth rates estimated from the two censuses.

The geometric mean of $P_1(x-5,5)$ and $P_2(x,5)$, therefore estimates the average number of persons reaching exact age x during a five-year interval within the intercensal period. Multiplying this geometric mean by 0.2 gives an average number of persons reaching exact age x during any one year of the intercensal period. Multiplying this by the length of the period gives an estimate of the number of people reaching age x in the period.

The GGB method was developed by rewriting the balance equation as $N(x)-[P_2(x+)-P_1(x+)]=D(x+)$, and dividing both sides by the number of person-years lived during the intercensal period by persons aged x and over, this gives $n(x)-r(x+)=d(x+)$.

The above expressions are for true values of the parameters rather than observed quantities. To obtain an equation containing observed quantities, let k_1 and k_2 denote the coverage of enumeration at the first and second censuses, respectively, and let c denote the completeness of reporting of deaths (each assumed to be constant with respect to age). In other words $P_1^*(x+)=k_1P_1(x+)$, $P_2^*(x+)=k_2P_2(x+)$, and $D^*(x+)=cD(x+)$, where $P_1^*(x+)$, is the observed value of $P_1(x+)$, $P_2^*(x+)$ is the observed value of $P_2(x+)$ and $D^*(x+)$ is the observed value of $D(x+)$.

Thus one gets, by substitution, the observed entry rate

$n^*(x)=\frac{N^*(x)}{PYL^*(x+)}=n(x)$, where $PYL^*(x+)$ denotes an estimate of the observed number of person-years lived by the population aged x and above during the intercensal period, *calculated from the observed age distributions.*

For the growth rate *r(x+),* substitution and simplification, gives $r(x+)=r^*(x+)+\frac{1}{t}ln\left(\frac{k_1}{k_2}\right)$, where $r^*(x+)$ is the observed growth rate of the population aged x and over calculated from the observed age distributions. Also $d(x+)=d^*(x+)\left[\frac{(k_1k_2)^{0.5}}{c}\right]$, where $d^*(x+)$ is the observed death rate for the population aged x and over as calculated from the observed numbers of persons and deaths.

Substituting the expressions for $n(x)$, $r(x+)$ and $d(x+)$ into the balance equation gives;

$n^*(x)-r^*(x+)=a+bd^*(x+)$

where $a = \frac{1}{t} ln\left(\frac{k_1}{k_2}\right)$ and $b = \frac{(k_1 k_2)^{0.5}}{c}$.

This equation contains only the observable quantities $n^*(x), r^*(x+)$ and $d^*(x+)$ and the parameters c, k_1 and k_2. To estimate values for c, k_1 and k_2 a straight line is fitted to the points, $(n^*(x) - r^*(x+), d^*(x+))$.

Migration can be accommodated in the GGB method as follows (Bhat, 2002; Hill and Queiroz, 2004); $r(x+) = b(x+) - d(x+) + nm(x+)$; where $nm(x+)$ is the net migration rate for the population aged x and over. Bhat (2002) suggested the use of a standard pattern of migration since accurate data on intercensal migration are rarely available (Rogers and Castro, 1981).

2.3 Bennett and Horiuchi

Bennett and Horiuchi (1981) proposed an alternative method of using the same data as the GGB method, namely, two censuses and the number of deaths by age. The method is a generalisation of the method proposed by Preston and Coale *et al* (1980).

The formulation involves estimating the population aged *a*, at a given point in time by accumulating the estimated deaths to that cohort until the cohort is extinct. The assumption underlying this estimation of deaths is that if mortality remains constant over the period the number of deaths can be estimated from the recorded deaths in an interval, by noting that deaths at any particular age will grow at the population growth rate at that age. Based on this, the population aged *a* can be estimated from the deaths at each age above that age, by applying exponential summed age-specific growth rates from *a* to *x*, to allow for the demographic history of the population (Bennett and Horiuchi, 1981, 1984).

Preston and Coale *et al* (1980) employed the following relationship assuming the population to be stable;

$N(a) = \int_a^\infty D(x) exp[r * (x - a)]dx$; where $D(x)$ is the true number of deaths of persons aged x exactly in the current population. Note that $D(x) exp[r * (x - a)]$ is an estimate of the number of people currently aged *a* who will die at age *x* in (*x*-*a*) years time. This follows from the fact that, in a stable population, the number of deaths to people aged *a* in a given year is related to that number in the previous year by a factor of exp(*r*).

This is the period analogue (for a stable population) of the method of extinct generations set forth by Vincent (1951), by which the number of persons aged *a* at a certain time in the past can be estimated by cumulating all deaths to persons aged *a* and above which are experienced by that cohort.

If the completeness of death registration is constant at age *a* and above, then

$$\overset{*}{D}(x) = kD(x),\ \forall x \geq a$$

$$\hat{N}(a) = \int_a^{\infty} \overset{*}{D}(x) exp[r(x-a)]dx$$

The ratios $\frac{\hat{N}(a)}{N(a)}$, where $N(a)$ is estimated from the census population, estimates completeness of death reporting, when the number of registered deaths by age, the number in the population by age, and the growth rate of the population are provided.

Accumulation of $\hat{N}(a)$ and $N(a)$ from oldest to younger ages tends to smooth out some of the random distortion resulting from age misreporting and differential registration and enumeration by age. The formula which is applied for computing the estimated age distribution is

$\hat{N}(a-5) = \hat{N}(a-5)exp[5r] + {}_5\overset{*}{D}_{a-5}\, exp[2.5r]$, where ${}_5\overset{*}{D}_{a-5}$ is the number of deaths occurring within the age group *a*-5 and *a* (Bennett and Horiuchi, 1984).

When a population deviates from stability, *r* is no longer constant but rather varies with age. In such cases, the total population growth rate, *r* is often a poor approximation of $r(a)$, the growth rate of the population aged *a*. For all age-specific growth rates in a population the equation is generalised as follows;

$\hat{N}(a-5) = \hat{N}(a-5)exp(5r) + {}_5\overset{*}{D}_{a-5}\, exp(2.5\, {}_5r_{a-5})$ .... (1);

where ${}_5r_{a-5}$ is the growth rate experienced by those in the age group *a*-5 to *a*. The last two terms in equation 1 are to account for growth in the population and the number of deaths over time. Equation 1 is derived from the following relationship that holds true for any closed population:

$$N(a)=\int_{a}^{\infty}D(x)exp\left[\int_{a}^{x}r(u)du\right]dx .$$

Thus, Bennett and Horiuchi removed the assumption of stability of the entire population. Instead, they relied on the less limiting assumption that the observed number of persons in each five-year age interval is approximately equal to the corresponding number in a stable population inferred from the numbers of persons at ages *a* and *a*+5, and the observed age-specific growth rates. After all values of $\hat{N}(a)$ are calculated, we can compute the values of ${}_{5}\hat{N}_{a}$, the estimated number of persons in the age group *a* to *a*+5, by using the following expression:

$${}_{5}\hat{N}_{a}=2.5(\hat{N}(a)+\hat{N}(a+5)).$$

In the older age groups, there is often a notable amount of curvature of number of persons within each five-year age group. For computation of the ${}_{5}\hat{N}_{a}$ above the age of 60, it was suggested to fit a curve over the five-year span and then determine the area under this curve. Estimates of completeness may be derived from the median or the mean of a series of age-specific estimates such as $\frac{{}_{10}\hat{N}_{a-5}}{{}_{10}N_{a-5}}$ (the ratios of the estimated number of people between age *a*-5 and *a*+5 and the corresponding figures in the observed population).

From the relationship found to hold in populations that are stable above age *a*, the following formulation was suggested by Bennett and Horiuchi (1984),

$$\hat{N}(a)=D^{*}(a+)(exp[r(a+)e(a)]-([r(a+)e(a)]^{2}/6)),$$

with the values of the rate of growth in the open interval ($r(a+)$), and, the expectation of life at the beginning of the open interval ($e(a)$), other $\hat{N}(a)$s can be determined. The value of $r(a+)$ is derived from the data themselves, while $e(a)$ must be obtained.

The above derived SEG method was further improved when Bennett and Horiuchi (1981) suggested an adaptation, that provided the coverage of one census relative to the other was constant for all (relevant) ages, one could accounts for this by the simple addition of a constant (equivalent to …, where k1 and k2 represent the coverage of the first and second censuses respectively, and completely analogous to the allowance for differential coverage in the GGB method) to the age specific growth rates. This adaptation is termed SEG+delta (where delta $=\frac{1}{t}ln\left(\frac{k_1}{k_2}\right)$

2.4 Analysis of the sensitivity of the SEG and GGB methods to various data errors

It is clear that while there are challenges associated with the application of these methods; these are not sufficient reasons not to attempt to apply the methods. Sensitivity analysis of the GGB and SEG methods has received much attention from some researchers, with the most recent assessment done by Hill (2001), Hill (2003), Hill and Choi (2004), and Dorrington and Timaeus (2008). The substance and outputs of their empirical work have been documented in this thesis.

In addition to applying the age transfers estimated by Bhat (1990), Hill (2001) simulated three errors. In his research, the age transfers were the same for both populations and deaths and a two per cent decline in population coverage for the second observation.

Based on the general results of the research by Hill (2001), it was clear that the GGB and SEG methods performed well in perfect data. Notwithstanding this, it was evident that the combined approach (GGB to adjust the populations before applying the SEG method) performed well (Hill, 2001, 2003; Hill, Choi and Timaeus, 2005).

2.4.1 Hill and Choi, and Dorrington and Timaeus

The section that follows records and discusses the nature and primary findings of the sensitivity analysis of the GGB and SEG methods, done by Hill and Choi (2004), and Dorrington and Timaeus (2008). Generally, it has been difficult to validate the performance of any method that estimates adult mortality because there is no "gold standard" against which to measure perfqrmance (Hill, Choi and Timaeus, 2005). In this regard, these studies were undertaken to determine the performance of these methods under different error scenarios.

To examine these methods, a series of empirical research efforts on the performance of the GGB and SEG methods were reviewed. Hill and Choi (2004) sought to assess the performance of the GGB and SEG methods in 23 different error scenarios. Dorrington and Timaeus (2008) assessed the performance of the GGB and SEG methods in 23 different error scenarios but in addition to that, they also applied the original version of the adjusted SEG method, SEG+delta (Bennett and Horiuchi, 1984) and compared their conclusions to those reached by Hill and Choi (2004). Dorrington and Timaeus (2008) extended their investigation to the situation where mortality included significant HIV/AIDS deaths.

2.4.2 Hill and Choi

Hill and Choi (2004) applied the GGB and SEG methods to a known population with known number of deaths which had been subjected to a variety of simulated errors, combined in various ways, to provide guidance in answering the questions raised above. However, the 23 error scenarios are not necessarily representative of errors in less developed countries, so they could not tell the likely ranges of error in such populations.

Analysis of the effects of age misreporting revealed a mixed picture. Not surprisingly, the effects of age misreporting (as modeled by Hill and Choi), whether in the populations or in the deaths, did not have significant effects. Hill and Choi concluded that the GGB method is less affected in terms of the final mortality estimate with respect to age misstatement, indicating that the estimated census coverage and the coverage of death reporting are distorted in compensating ways.

This finding was further substantiated by the later analysis by Hill and Choi, which revealed that the combined approach gives almost accurate results. Thus, the combination (GGB+SEG) of these techniques came to be seen as having produced accurate estimates of ${}_{45}q_{15}$.

Hill and Choi examined the diagnostic plots for six scenarios and their combinations; this formed the basis of their study. The key component in reading the diagnostic plots is the interpretation of the GGB method with regards to the change of intercept and the slope. The GGB method and SEG method performed as expected in the no error scenario. For the GGB method, points at higher age groups, particularly those 70 plus categories, tended to slightly deviate from the fitted line.

The GGB and SEG methods performed as expected giving estimates of completeness of death reporting of 80 per cent in the scenario where the omission of deaths was 20 per cent. The study by Hill and Choi (2004) revealed that the GGB method and the combined method gave expected results in situations where the census coverage decline by two per cent at all ages.

It was found that the irregularities from age exaggeration in both population and deaths scenario undoubtedly played an important role in the overestimation of coverage of death reporting of the GGB and SEG methods. Looking specifically at age varying census coverage, the results revealed that the GGB method and the combined method performed poorly, with the SEG method overestimating coverage of death reporting to a lesser extent than the other two methods.

Finally, in terms of the effects of emigration on the three approaches, the GGB method and the combined method produced an estimate closer to the true value; Hill and Choi concluded that this was a result of misinterpretation of the GGB method of emigration as a decline on census coverage. Conversely, with respect to the SEG method, the diagnostic plots revealed the negative impact of emigration on its performance to countries which experience high net emigration and consequently, showing under-reporting of deaths resulting in overestimation of mortality.

2.4.3 Dorrington and Timaeus 2008

Dorrington and Timaeus (2008) compared the original results by Hill and Choi (2004) to the results using the method adapted to correct for differential census coverage (the SEG+delta method), and then went on to do the same comparison after applying these methods to the African data set (with HIV).

They found an improvement in estimates to those using the Hill and Choi approach, using the SEG+delta method. An analysis of the methods according to the distribution of mean and median indicated that the SEG+delta method provides estimates closer to the true value than the GGB method. Not surprisingly, this was supported by lower root mean square errors than the other two methods.

Further analysis of estimates of ${}_{45}q_{15}$ from these methods (SEG+delta, GGB+SEG, and GGB), revealed that the highest percentage of most accurate estimates of mortality rates (${}_{45}q_{15}$) was obtained from the SEG+delta method (65 per cent), while the worst mortality estimates were recorded from the GGB method (13 per cent). The cases where the SEG+delta method was not the best (three scenarios), the difference between them and the most accurate estimates of mortality were more than five percent. The scenarios where the SEG+delta method was undoubtedly the worst involve age misreporting in censuses, while two per cent decline in census coverage and immigration were involved in two cases where GGB was the best. In addition to the above, assessment of the results showed that the SEG+delta method was the best method, the GGB +SEG combination was only best in the remaining five scenarios (22 per cent).

When viewed in terms of the GGB +SEG combination, it became apparent that an overestimate of mortality by the SEG method was cancelled by an underestimate by the GGB method. Hence this performance of the GGB+SEG combination in these five scenarios was a consequence of either age misreporting or violation of the assumption of population closed to migration and in situations where the SEG+delta method performed worst. All the methods violated the assumption of constant census coverage and constant completeness of death reporting for all appropriate ages.

Ensuring that the SEG+delta method takes into account the pattern of HIV mortality, Dorrington and Timaeus (2008) used an alternative method to determine the life expectancy of the open age interval (a method of approximating the life expectancy as the average of those from West life Tables with the same ${}_5m_{60}$, ${}_5m_{65}$ and ${}_5m_{70}$ as the observed rates corrected for incompleteness (solved iteratively)). The alternative method improved the applicability of the SEG+delta method to populations with HIV mortality.

Importantly, the performance of the SEG+delta method generally corresponded with that obtained previously, particularly after applying the method of estimating life expectancy by Dorrington and Timaeus (2008). That is, the SEG+delta method outperformed in 15 scenarios producing mean and median of all scenarios only one per mille lower than was previously obtained, with roughly equivalent variation. In addition, this alternative method was applied to the African data set.

The distribution of the errors from the SEG+delta method applied to the African data set closely mirrored those established from the Hill and Choi data set, although the errors were bigger. In general the higher mortality rates due to HIV appears to improve the estimates from the GGB method while making the estimates from the other two methods more biased, and increasing the variation of all methods. Not surprisingly, it was noted that there was a direct correlation between the level of error in the SEG+delta method and the combination, GGB+SEG, and completeness of vital registration.

since there is no systematic criterion to be used to determine the best combination that result in best estimates of ${}_{45}q_{15}$ per scenario and across all 23 error scenarios.

Chapter 3 Methodology

This chapter outlines the methodology of undertaking the primary research to address the issues highlighted in Chapter 2. The methodology employed in this study is a combination of practices as drawn from similar studies on the specific requirements of the SEG+delta method. It gives a description of the development of the key components of the SEG+delta method.

3.1 Key Components of Synthetic Extinct Generations Method (SEG)

Based on the literature review in Chapter 2, it is apparent that the assessment of the SEG+delta method is based in the investigation of the key components of the SEG+delta method as highlighted in Chapter 1 (Dorrington and Timaeus, 2008).

The first stage in this research is therefore to identify and investigate the key components of the SEG+delta method. The second stage is to determine the differential performances of the different combinations of key components of the SEG+delta method when applied to the 23 error scenarios. After that, the data set that is typically African is constructed by replacing the components of the assumptions used by Hill and Choi (2004) with assumptions more typically African where these differ.

Finally, this study investigates all possible combinations of the key components of the SEG+delta method and validate the performance of the SEG+delta method per each combination, thus determining the combinations that result in most accurate estimates of ${}_{45}q_{15}$ per scenario and the method that gives the most accurate estimates of ${}_{45}q_{15}$ across all the scenarios. Additionally, the combination is determined for scenarios which gives better estimates of ${}_{45}q_{15}$ as compared to the combination applied from previous studies(Dorrington and Timaeus, 2008; Hill and Choi, 2004). The validation of the key components of the SEG+delta method is done using Microsoft Excel. The analysis is done for all the 23 error scenarios as developed by Hill and Choi (2004).

But to start with, the challenge in investigating adult mortality in less developed countries based on vital registration and censuses is because of data quality. Data deficiencies found in less developed countries include, incomplete vital registration, inaccurate censuses, and misreporting of age at death or age of the living (United Nations, 1983; Bhat, 1990). The next sections describe the sources of data used in the development of the key components of the SEG+delta method.

3.2 Data Sources

The investigation of the key components of the SEG+delta method was done using the scenarios created by Hill and Choi (2004). They simulated a non-stable population with known adult mortality (West female model life table, level 15, with $_{45}q_{15}$ of 0.309). Various error scenarios were simulated. A total of 23 scenarios (see Appendix 1) were created as combinations of major potential errors. Change in census coverage of the population was represented by a two per cent decline from the first to the second census. Omission of deaths was initially represented by a 20 per cent uniform omission across age groups, with two cases of age-varying omission (one increasing with age, the other declining). Age misreporting in population and/or in deaths were derived from a matrix of transfers between 5-year age groups estimated for India (Bhat, 1990). Age-varying census coverage was based on net-undercount of the male black population from the 1980 United States Census. Finally, emigration and immigration were based on a pattern of age-specific in-migration to the U.S. of Mexican males 1980-1990.

3.3 Method used to estimate completeness of death reporting

Two methods of estimating completeness of death reporting (c) are considered, one based on the estimate of N_x (Hill and Choi, 2004), and the other based on the estimate of $_5N_x$ (United Nations, 2002) were examined. A comparison of the performance following two methods was done.

The methods of estimating $_5N_x$ and N_x were:

- Method used in this study, suggested by Feeney (United Nations, 2002); $_5N_x = \dfrac{(P_2(x) - P_1(x))}{(ln(P_2(x)) - ln(P_1(x)))}$ where $P_2(x)$ are the number of survivors who reached exact age x during the five years preceding the second census and $P_1(x)$ is the number of persons aged x at the first census ; this will be used to determine N_x
- Method used by Hill and Choi (2004), $N_x = \sqrt{P_2(x) * P_1(x-5)}$

The methods are compared by applying the SEG+delta method to the 23 error scenarios from (Hill and Choi, 2004) with a combination of other key components of the SEG+delta method described in the next sections. The per cent differences of estimates of $_{45}q_{15}$ from the true value are compared. During the investigation process, the open interval is increased from 60+, 75+ to 85+, but keeping the age range used to determine the completeness fixed at 15-59. This is done to determine the effect of the open interval on the two methods of estimating completeness. In addition, the age range used to determine completeness is also increased from 15-59 to 15-64, 15-69, 15-74, 15-79, and to 15-84, subject to the maximum for a given open interval.

3.4 Determining completeness (c) from the c_xs

Dorrington and Timaeus (2008) call for research into how best c can be estimated. To achieve this, the following methods are examined: determining c from the median of c_xs ($c_x = \frac{\hat{N}_x}{N_x}$), determining c from median of ${}_5c_x$s (${}_5c_x = \frac{{}_5\hat{N}_x}{{}_5N_x}$) and determining c from the median of c_{x+}s ($c_{x+} = \frac{{}_5\hat{N}_{x+}}{{}_5N_{x+}}$). A c and its respective ${}_{45}q_{15}$ is obtained for each of the 23 error scenarios applied by Hill and Choi (2004) for each approach. These are combined with other key components of the SEG+delta method.

The results of estimating completeness of death reporting using the three proposed methods (estimating c as the median of the c_xs, estimating c as the median of ${}_5c_x s$ and estimating c as the median of the c_{x+}s) are compared. The distribution of the standard error of c is calculated in assessing the three methods mentioned above, and then the mean deviation of the estimated ${}_{45}q_{15}$ from the 23 error scenarios is employed as a criterion for deciding on the best method to determine c.

Following the determination of the best method for determining c from c_xs for the SEG+delta method, the criteria for minimising c_xs is developed and examined. The key measures used to determine the best c_x are the median deviation and the mean deviation. The estimates of completeness, c, for each of the 23 error scenarios is obtained after minimising the deviations from the mean as well as minimizing the deviations from the median. However these methods are evaluated after fixing the age range used to determine completeness (15-59) for the open age interval (85+). Thus these methods are not investigated for the 216 combinations of key components of SEG+delta method. In fact in this study c is determined by setting c to the median of the c_xs after minimising the deviations from the mean.

3.5 The best way of estimating e_x

In order to apply the SEG+delta method one needs (as described by Bennett and Horiuchi (1984) and Dorrington and Timaeus (2008)) an estimate of the life expectancy of the open age interval. Hill and Choi derived the life expectancy of the open interval from the ratio of ${}_{30}d_{10}$ and ${}_{20}d_{40}$ and a look up table based on the regression from the Princeton West Life Table (Coale, Demeny and Vaughan, 1983) produced by Bennett and Horiuchi (1984) .

Unfortunately this relationship and many others involving the use of Princeton Life Tables are far from representative of the less developed countries where mortality includes significant HIV/AIDS deaths. Thus an alternative method of approximating life expectancy is investigated. Following Dorrington and Timaeus (2008), the method which is investigated involves approximating the life expectancy as the average of those from West life Tables with the same ${}_5m_{60}$, ${}_5m_{65}$ and ${}_5m_{70}$ as the observed rates corrected for incompleteness (solved iteratively). The e_x is estimated from the data with simulated errors, using ${}_5m_{60}$, ${}_5m_{65}$ and ${}_5m_{70}$ from the West life table, the resultant adjusted deaths are used to repeat the computation. This process is repeated until the estimates of e_x no longer changes.

The analysis of the two methods of estimating life expectancy is done according to the percentage difference of the estimated ${}_{45}q_{15}$ from the true value (results of ${}_{45}q_{15}$ from the SEG+delta method after using the true value e.g. for $e(60)$, namely, 15 years from the Princeton West Life tables). The mean deviations of the ${}_{45}q_{15}$ estimates from the three methods are determined and are compared. Finally, the age of the open interval is increased from 60+ to 75+ and to 85+, and the comparison is repeated. This is done in order to determine the performance of the methods as the age of the open interval increases. The comparison is done on all combinations with each method used to estimate of e_x.

3.6 The best open age interval to use

Three different open age intervals are investigated using the SEG+delta method (60+, 75+ and 85+). This performance of these different open intervals as well as the standard deviation of the mortality estimates (${}_{45}q_{15}$) for all the 23 error scenarios is compared.

Following the completion of the investigation of the key components of the SEG+delta method, the SEG+delta method is applied to the 23 error scenarios. After the analysis of the results, they are compared to the results from the studies done by Hill and Choi (2004) and Dorrington and Timaeus (2008).

Chapter 4 Results

4.1 Introduction

This section presents the results of investigations into possible combinations of key components of the SEG+delta method.

To begin with, the key aspects of the SEG+delta method as highlighted by Dorrington and Timaeus (2008) are investigated. As indicated in Chapter 3, the SEG+delta method is applied to the Hill and Choi (2004) error scenarios to decide which combination of key components of the SEG+delta method gives the most accurate estimate of ${}_{45}q_{15}$ for each scenario and across all scenarios. This was done to allow comparability between the methods employed in previous studies and the methods to be tested by this study. The results of each scenario were investigated and the combinations of the key components of the SEG+delta method which gives the best estimate of ${}_{45}q_{15}$ are determined (for additional results see Appendix 3, Appendix 4 and Appendix 5).

For each scenario, the pattern of the estimates of ${}_{45}q_{15}$ is described per set of combinations. Thus in each scenario the combination which results in the most accurate estimate of ${}_{45}q_{15}$ is identified. In order to determine this best estimate of ${}_{45}q_{15}$ the following key components in different combinations were investigated:

- three open age intervals (60+, 75+ and 85+)
- two methods used to calculate N_x (the used by Hill and Choi (2004) and the method suggested by Feeney (United Nations, 2002)
- three methods used to determine completeness of death reporting (${}_5c_x$, c_{x+} and c_x)
- three methods used of estimating the life expectancy at the open age interval (known e_x from the Princeton West Life Table (*K*)), iterative method (*I*) proposed by Dorrington and Timaeus (2008), the method using the ratio of ${}_{30}d_{10}$ and ${}_{20}d_{40}$ and a look up table based on the regression from the Princeton West Life Table (R) used by Hill and Choi (2004))
- six age intervals used to determine completeness (15-59 last birthday, 15-64 last birthday, 15-69 last birthday, 15-74 last birthday, 15-79 last birthday and 15-84 last birthday)

4.2 Determining completeness (*c*) from the c_xs (Mean vs Median and ensuring a level set of c_xs)

In this study *c* is determined by setting it to the median of c_xs after minimising the deviations from the mean. However, in fact, two methods (mean vs median) of determining *c* from the c_{x+}s were investigated for a limited number of combinations with key components of the SEG+delta method. Setting *c* to the mean or median of c_{x+}s over the age range 15-59 gave very similar results. The median of c_{x+}s gave a standard error of the estimate from the true value of 1.4 per cent while that of the mean is 1.5 per cent. However when the age range used to estimate *c* is increased from 15-59 to 15-84, it is observed that the mean is affected by outliers (extreme values at ages above 65). However, the differences in estimates of $_{45}q_{15}$ were small.

The question of whether minimizing deviations from the mean or from the median produces more accurate estimates was investigated. A comparison was done using the age range 15-59 to determine completeness for an open age interval of 60+. The results showed that minimizing the deviation from the mean yield estimates of $_{45}q_{15}$ closer to the true value, outperforming the method of minimizing the deviation from the median in 11 scenarios. Given this, it was concluded that the better method of determining *c* is by minimizing the deviation from the mean of the c_xs and hence this method is applied in this study. Thus giving rise to the question of how well minimising deviation from the median will perform in all the combinations tested in this study, especially in scenarios where the SEG+delta method performs poorly (Scenarios 1, 3, 7, 20 and 22).

4.3 Performance of the SEG+delta method

Best estimates were obtained from the no error scenario, and also for scenarios where the errors are the ones for which the SEG+delta method was designed to correct, scenarios where there is either uniform differential coverage by censuses or uniform omission of deaths by age (estimates from scenarios 6, 8 and 11 are close to the true value).

The estimates of $_{45}q_{15}$ become worse when uniform differential coverage by censuses or uniform omission of deaths by age are combined with other errors (e.g. when combined with age misreporting in the censuses or age misreporting in the reported deaths), in scenarios 12-14. It is interesting to note that the estimates of $_{45}q_{15}$ in scenarios 12-14 are close to and have the same pattern to the estimates from scenario 5.

Estimates of $_{45}q_{15}$ are not close to the true value in scenarios 1, 3 and 7 (scenarios where census coverage varies with age), but have a similar pattern of $_{45}q_{15}$ across all combinations of key components of the SEG+delta method. The worst estimates of $_{45}q_{15}$ are from scenarios where there is a combination with immigration and age misreporting in census (scenarios 20 and 22). The estimates are similar across all combinations of key components of the SEG+delta method. However, in scenarios where the migration is strongly emigration (scenarios 17 and 21), the estimates are better.

Overall, the results revealed that, for most of the scenarios, the higher open age interval (85+) provided better estimates of $_{45}q_{15}$ compared to lower open age intervals (60+ and 75+). However the estimates of $_{45}q_{15}$ from an open age interval of 85+ have, by and large, similar patterns to the estimates from an open age interval of 75+. Generally, results within each open age interval (60+,75+ and 85+) differ by method used to determine completeness of death reporting ($_5c_x$, $c_{x+,}$ c_x), which also differ by the age range used to determine completeness (15-59, 15-64, 15-69, 15-74, 15-79 and 15-84). Finally, in most scenarios estimates also differ by method of estimating e_x with the worst estimates being produced when e_x is estimated using the method used by Hill and Choi (2004). The following sections describe in more detail the performance of the various combinations of the key components for each scenario, illustrated by the results using 75+ as the open interval, for convenience.

4.3.1 Scenarios without error or with constant under-reporting by age (Scenarios 0, 6, 8 and 11)

As can be seen from Figure 1, all combinations produce estimates of $_{45}q_{15}$ close to the true value with the exception of the combination of an open age interval of 60+ with e_x estimated from the method used by Hill and Choi (2004), where errors are greater than two per cent (but less than five per cent). The result for Scenario 6, 8 and 11 are virtually the same as shown for Scenario 0 in Figure 1.

Overall, the combination which results in the best estimate of $_{45}q_{15}$ was a combination of N_x calculated from the method used by Feeney (United Nations, 2002), e_x estimated by the method used by Hill and Choi (2004), determining completeness from c_x over the age range 15-74 with an open age interval 75+.

Estimates of ${}_{45}q_{15}$ are not close to the true value in scenarios 1, 3 and 7 (scenarios where census coverage varies with age), but have a similar pattern of ${}_{45}q_{15}$ across all combinations of key components of the SEG+delta method. The worst estimates of ${}_{45}q_{15}$ are from scenarios where there is a combination with immigration and age misreporting in census (scenarios 20 and 22). The estimates are similar across all combinations of key components of the SEG+delta method. However, in scenarios where the migration is strongly emigration (scenarios 17 and 21), the estimates are better.

Overall, the results revealed that, for most of the scenarios, the higher open age interval (85+) provided better estimates of ${}_{45}q_{15}$ compared to lower open age intervals (60+ and 75+). However the estimates of ${}_{45}q_{15}$ from an open age interval of 85+ have, by and large, similar patterns to the estimates from an open age interval of 75+. Generally, results within each open age interval (60+,75+ and 85+) differ by method used to determine completeness of death reporting (${}_5c_x$, c_{x+}, c_x), which also differ by the age range used to determine completeness (15-59, 15-64, 15-69, 15-74, 15-79 and 15-84). Finally, in most scenarios estimates also differ by method of estimating e_x with the worst estimates being produced when e_x is estimated using the method used by Hill and Choi (2004). The following sections describe in more detail the performance of the various combinations of the key components for each scenario, illustrated by the results using 75+ as the open interval, for convenience.

4.3.1 Scenarios without error or with constant under-reporting by age (Scenarios 0, 6, 8 and 11)

As can be seen from Figure 1, all combinations produce estimates of ${}_{45}q_{15}$ close to the true value with the exception of the combination of an open age interval of 60+ with e_x estimated from the method used by Hill and Choi (2004), where errors are greater than two per cent (but less than five per cent). The result for Scenario 6, 8 and 11 are virtually the same as shown for Scenario 0 in Figure 1.

Overall, the combination which results in the best estimate of ${}_{45}q_{15}$ was a combination of N_x calculated from the method used by Feeney (United Nations, 2002), e_x estimated by the method used by Hill and Choi (2004), determining completeness from c_x over the age range 15-74 with an open age interval 75+.

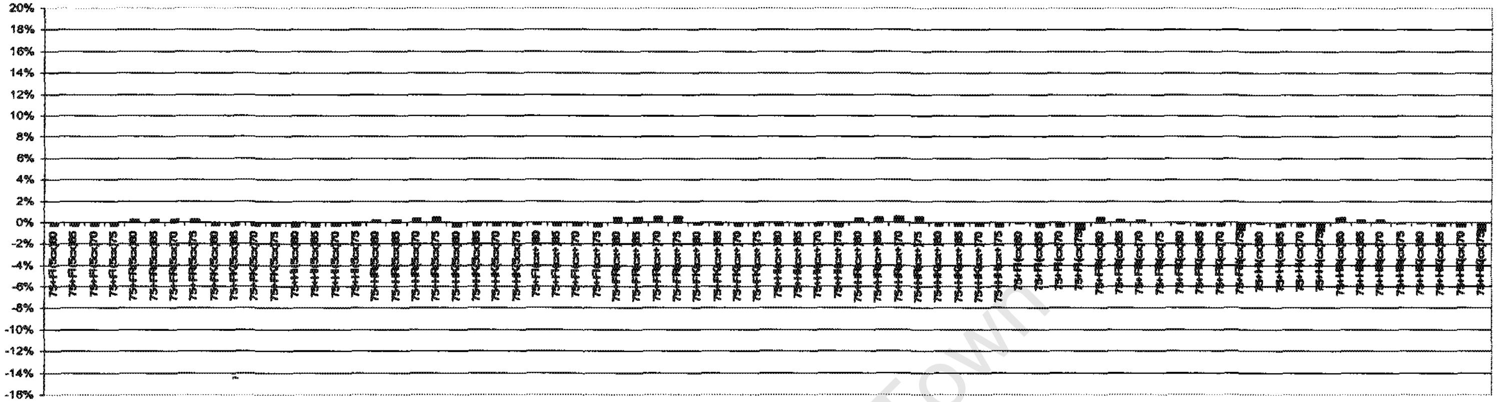

Figure 1 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 0

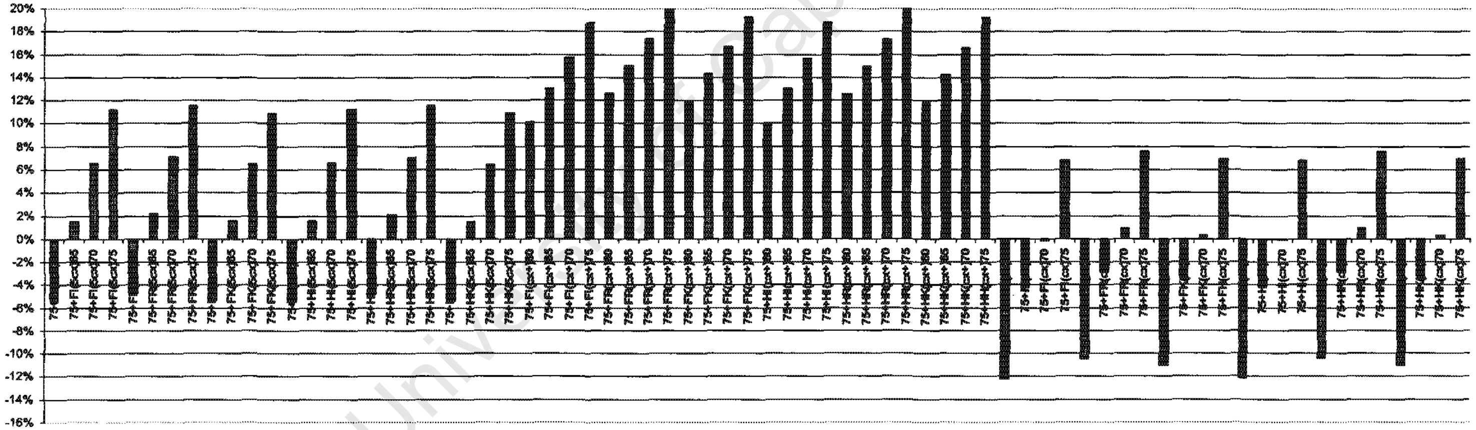

Figure 2 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 1

4.3.2 Age varying census coverage (Scenarios 1 and 7)

Figure 2 shows that estimates of ${}_{45}q_{15}$ have similar patterns within each method of estimating completeness, with estimates from the same age interval used to determine completeness having similar estimates of ${}_{45}q_{15}$ besides the different methods used to calculate N_x and different methods used to estimate e_x. In addition, results from these scenarios reveal that the most accurate estimates of ${}_{45}q_{15}$ are from estimating completeness using c_x, over the age range 15-69 for an open age interval of 75+ or greater with the exception of combinations with e_x estimated using the method used by Hill and Choi (2004) for the open age interval of 75+. The estimates of ${}_{45}q_{15}$ using c_x over the age range 15-59 to estimate completeness for an open age interval of 60+ are decidedly worse. Thus the estimates of ${}_{45}q_{15}$ are particularly affected by the method used to determine completeness of death reporting (${}_{5}c_x$, c_{x+}, c_x) and the age range used to determine this completeness.

The general pattern of performance of the SEG+delta method by different combinations of key components of the SEG+delta method described above did not differ by the open age interval applied (60+, 75+ and 85+). However the estimates of ${}_{45}q_{15}$ improved when the open age interval used to estimate completeness is lowered from 85+ to 60+, but the difference in magnitude of estimates is insignificant.

Overall, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ is a combination of N_x calculated using the method used by Hill and Choi, with e_x estimated by the iterative method, completeness determined using c_x over the age range 15-69 for an open age interval of 75+ (see Appendix 3).

4.3.3 Age misreporting in censuses (Scenario 2)

The most accurate estimates of $_{45}q_{15}$ are as a result of estimating completeness using $_5c_x$ over the age range 15-69 (similar estimates using 15-79), with the least accurate estimates obtained using the $_5c_x$ for the age interval of 15-74. Highest variation of estimates across six age intervals used to determine completeness is obtained from combinations with completeness estimated using $_5c_x$, with less variation of estimates of $_{45}q_{15}$ as a result of using c_x (see Figure 3).

The pattern described above is apparent for open age intervals 75+ and 85+. When the open age interval is lowered to 60+, the estimates of $_{45}q_{15}$ improved especially using c_{x+} to determine completeness of death reporting. Overall, the estimates did not differ by the method used to estimate e_x or by method used to calculate N_x but they are affected by the method used to determine completeness ($_5c_x$, $c_{x+,}$ and c_x).

As is evident from further analysis, a reasonable fit is produced using an open interval of 60+, estimating e_x using the method used by Hill and Choi (2004) and using c_{x+} to estimate completeness. However, the combination which results in the best estimate of $_{45}q_{15}$ is a combination with N_x calculated using the method suggested by Feeney (United Nations, 2002), e_x calculated using the method used by Hill and Choi, using c_{x+} over the age range 15-59 to determine completeness with an open age interval of 60+.

4.3.4 Age misreporting in censuses + age varying census coverage (Scenario 3) and 20 per cent VR omissions + increasing completeness with age (Scenario 9)

The estimates of $_{45}q_{15}$ have similar patterns for the same age interval used to determine completeness, but differ by the method used to determine completeness ($_5c_x$, c_x and c_{x+}).

It is interesting to, note from Figure 4 that there is an increase in the percentage difference in Scenarios 3 and 9 with an increase in the age interval used to determine completeness (with similar estimates for $_{45}q_{15}$ for the same age range used to determine completeness). Not surprisingly, the same pattern is apparent for the average deviation from the estimated $_{45}q_{15}$, but increasing with an increase in open age interval from 60+ to 85+ (2.4 per cent, 9.7 per cent, and 11.6 per cent respectively). The results reveal that all errors are greater than two per cent. However, the best fits are from using $_5c_x$ over the age range 15-64 to determine completeness for open age intervals 75+ and 85+, and estimating e_x using the method used by Hill and Choi (2004), using $c_{x\cdot}$ over the age range 15-59 to determine completeness with an open age interval of 60+ (estimates from ${}_5c_x$ and c_x have less variation across the six age intervals used to determine completeness). The worst estimates are as a result of estimating completeness using c_{x+} over the more extensive age ranges (15-79 and 15-84).

From Figure 4, the results also reveal that the estimates do not differ by method used to estimate e_x, or method used to calculate N_x but the estimates improve when the open age interval is increased from 60+ to 85+.

The pattern described above is apparent for the open age intervals 75+ and 85+. However, the combination which results in the best estimate of ${}_{45}q_{15}$ is the combination of N_x calculated by the method used by Hill and Choi, known e_x, using ${}_5c_x$ to determine completeness over the age range 15-64 for an open age interval 75+.

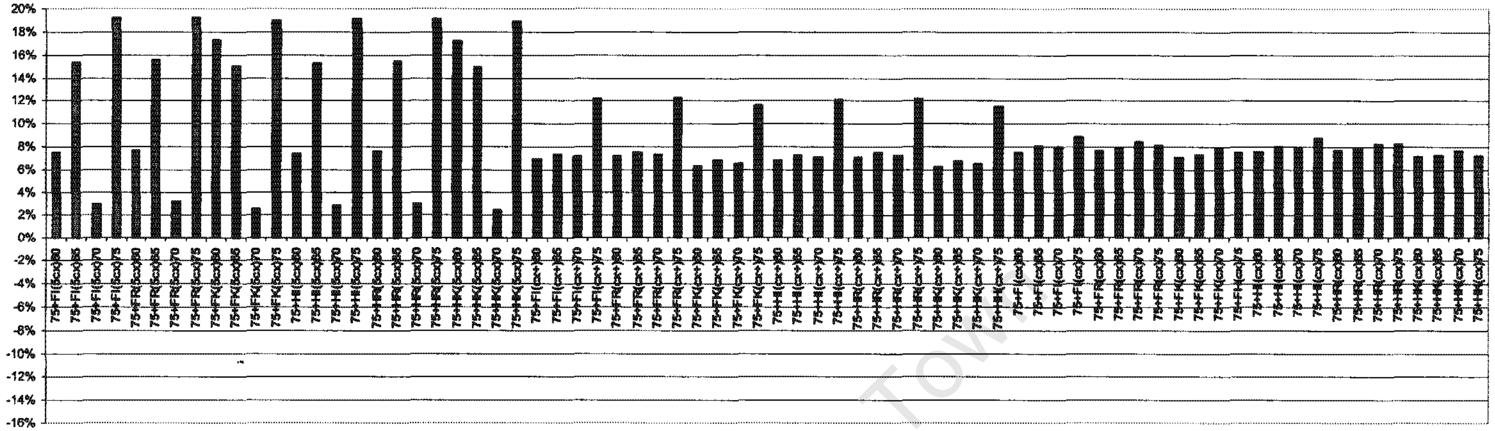

Figure 3 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 2

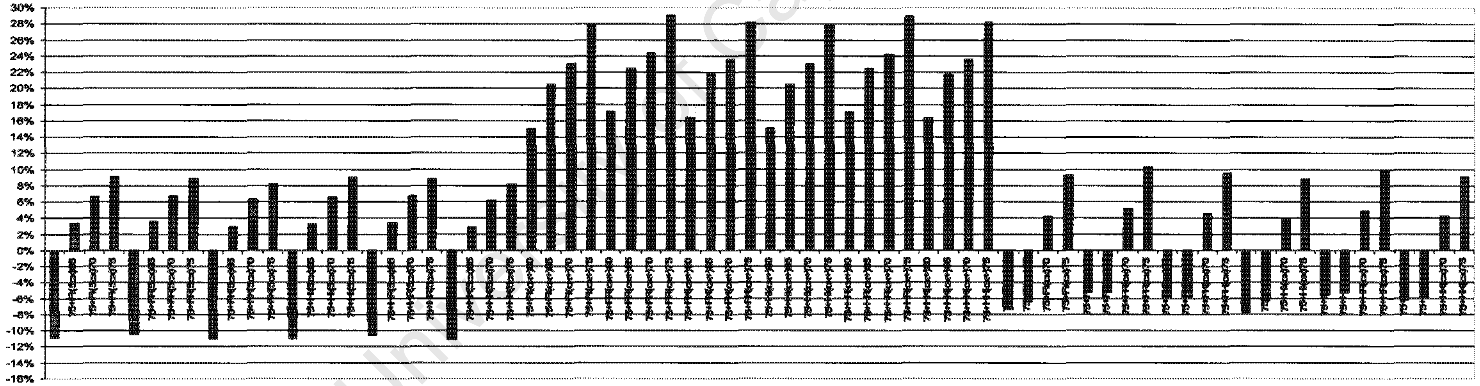

Figure 4 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 3

4.3.5 Age misreporting in VR (Scenario 4)

From Figure 5, we see that the estimates of $_{45}q_{15}$ are similar for the same age interval used to determine completeness using a given method to determine the completeness ($_5c_x$, c_{x+}, c_x). The most accurate estimates of $_{45}q_{15}$ result from estimating completeness using $_5c_x$ and c_x, with e_x calculated using the method used by Hill and Choi. This is supported by the low average deviation (two per cent) of the estimates of $_{45}q_{15}$ using $_5c_x$ and e_x calculated by the method used by Hill and Choi. However, there is little difference in the variation of estimates of $_{45}q_{15}$ across the three methods used to determine completeness of death reporting ($_5c_x$, c_x and c_{x+}).

The pattern of estimates of $_{45}q_{15}$ described occurs for both age intervals 75+ and 85+. Consistent with the pattern shown by both the percentage difference and the average deviation, short age ranges (15-59, 15-64 and 15-69) and lower open age intervals (60+ and 75+) have better estimates compared to longer age ranges (the least accurate estimate came from using the 15-84 age range) and higher open age interval (85+) across all methods used to determine completeness of death reporting. There is no significant difference which N_x is used to determine completeness, however calculating N_x using the method used by Hill and Choi had the lowest average deviation (2.7 per cent) of the estimates of $_{45}q_{15}$.

Examination of the results reveals that the best estimates are from a combination of e_x calculated using the method used by Hill and Choi (2004) for an open interval of 60+. However, the combination which results in the best estimate of $_{45}q_{15}$ is a combination of N_x calculated using the method used by Hill and Choi, e_x estimated using the method used by Hill and Choi, estimating completeness using $_5c_x$ over the age range 15-59 for an open age interval 60+.

4.3.6 Age misreporting in censuses + age misreporting in VR (Scenario 5) with constant under-reporting by age (Scenarios 12, 13 and 14)

As Figure 6 shows, using $_5c_x$ to estimate completeness produces a different pattern of estimates of $_{45}q_{15}$ for each age interval to the pattern from that produced using c_{x+} and c_x (although the estimates using c_x have the least variation of 1.4 per cent). The most accurate estimates of $_{45}q_{15}$ are from the combinations using c_{x+} over the age ranges 15-64 or 15-69 to determine completeness, while the least accurate estimates are from combinations using ${}_5c_x$ over the age range 15-74 to determine completeness.

The pattern described above is found for open intervals 75+ and 85+, but differs for the open age interval of 60+. However, the estimates of ${}_{45}q_{15}$ became worse when the open age interval was lowered from 85+ to 75+. In addition, closer inspection reveals that, with respect to average deviation of percentage difference, there is an increase from 4.0 to 4.3 per cent.

Overall, the best estimates of ${}_{45}q_{15}$ are from the open age interval of 85+; thus estimates of ${}_{45}q_{15}$ differ by open age interval. The results are seen to be better when key components are combined with known e_x or with e_x estimated using the iterative method, with the worst estimates as a result of combinations with e_x estimated using the same method as Hill and Choi. However there was no significant difference in the average deviation of the estimated ${}_{45}q_{15}$ from all the three methods used to estimate e_x at the age of the open interval (4.4, 4.3 and 4.3 per cent respectively).

As is evident from this analysis the most accurate estimates are obtained from a combination with c_{x+} over the age range 15-64 to determine completeness and also from combinations using c_x over age ranges 15-74, 15-79 and 15-84 to determine completeness for an open age interval of 85+ and also from a combination of using ${}_5c_x$ over that age range 15-69 to determine completeness with an open age interval of 75+.

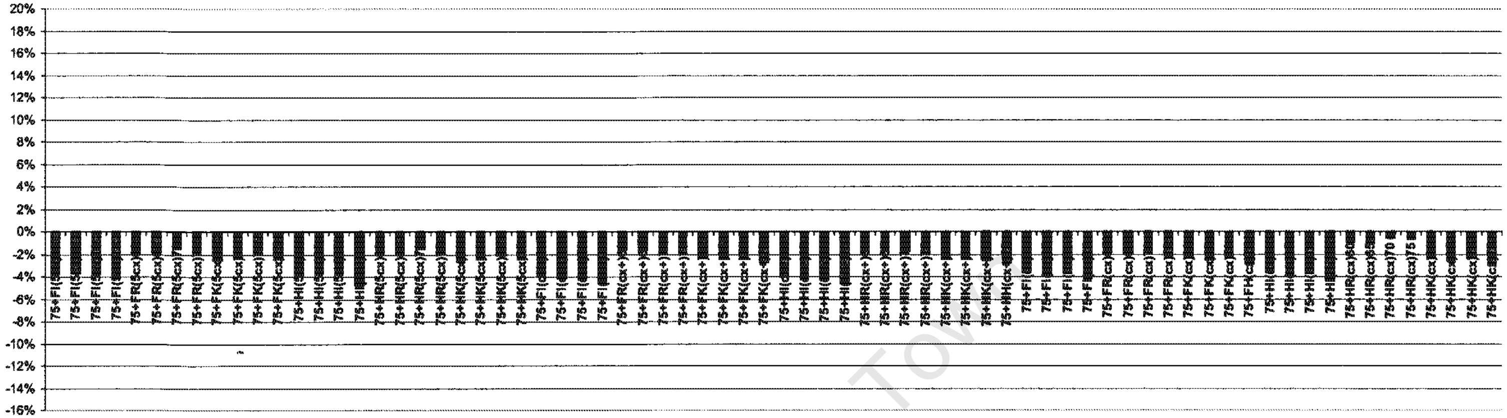

Figure 5 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 4

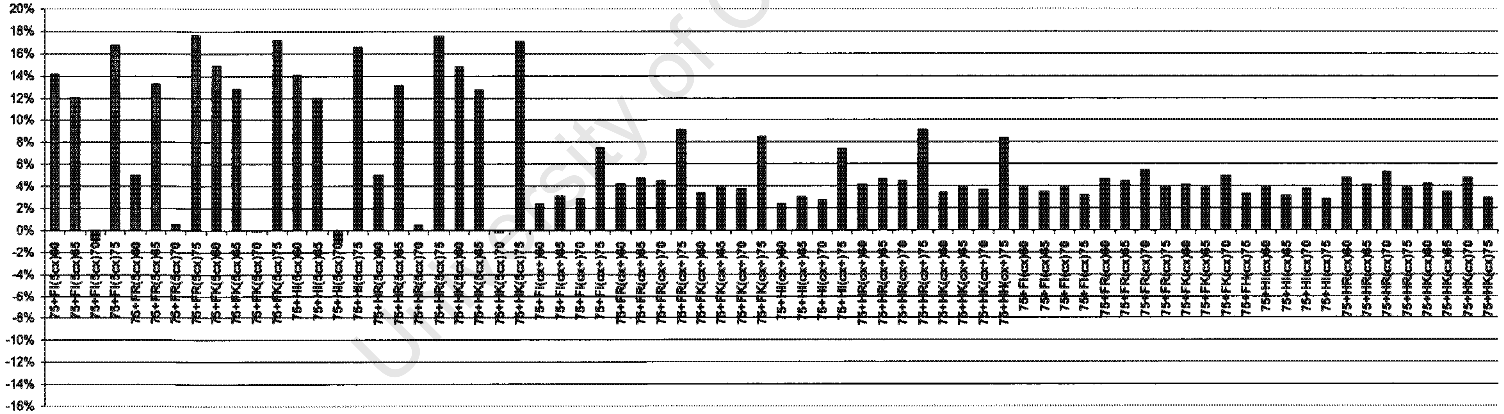

Figure 6 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 5

In conclusion, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ for these scenarios was a combination of N_x calculated using the method used by Hill and Choi, e_x estimated by a method used by Hill and Choi, with ${}_5c_x$ over the age range 15-64 used to determine completeness for an open age interval 85+.

4.3.7 VR 20% omission + decreasing completeness with age (Scenario 10)

As can be seen from Figure 7, all estimates of ${}_{45}q_{15}$ from all combinations are poor, mostly similar (errors between six and eight per cent) except the worst (errors greater than nine per cent) from estimating e_x using the method used by Hill and Choi (2004) for open age interval 60+. Estimates from the application of the known e_x outperform the other two methods of estimating e_x. The pattern above is observed for all open age intervals but the higher the open age interval the better the estimates of ${}_{45}q_{15}$. However the difference in estimates of ${}_{45}q_{15}$ is small.

However, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ for scenario 10 is a combination of N_x calculated from the method suggested by Feeney (United Nations, 2002), known e_x, using ${}_5c_x$ over the age range 15-64 to determine completeness for an open age interval of 75+.

4.3.8 Emigration (Scenarios 15 and 16)

Close examination of the results in Figure 8 shows clearly that the most accurate estimates were from the combination of e_x using the method used by Hill and Choi and using c_x to determine completeness for an open age interval of 75+. Most of estimates from the open age intervals 85+ and 75+ are good except for the combinations using c_x over the age range 15-59 or 15-64 to determine completeness for an open age interval of 85+; a combination of e_x based on the method used by Hill and Choi and using c_x or c_{x+} over the age ranges 15-64, 15-69 or 15-74 to ,determine completeness for the open age interval of 75+. Further analysis reveals that the open age interval of 85+ produces estimates of ${}_{45}q_{15}$ which are closer to the true ${}_{45}q_{15}$, with less variation compared to the open age intervals of 60+ and 75+. However, estimates of ${}_{45}q_{15}$ have similar pattern for open intervals 75+ and 85+.

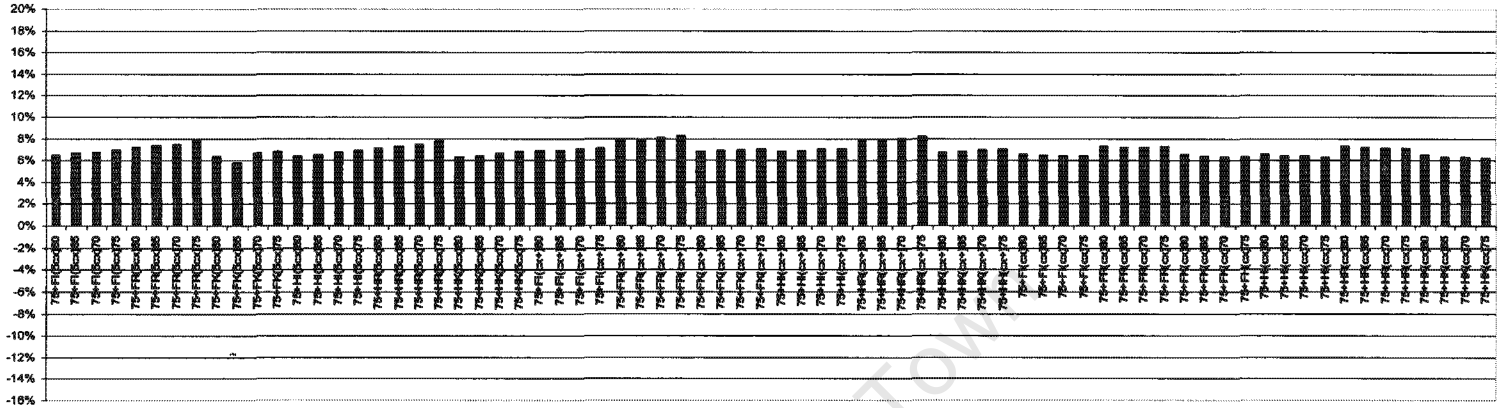

Figure 7 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 10

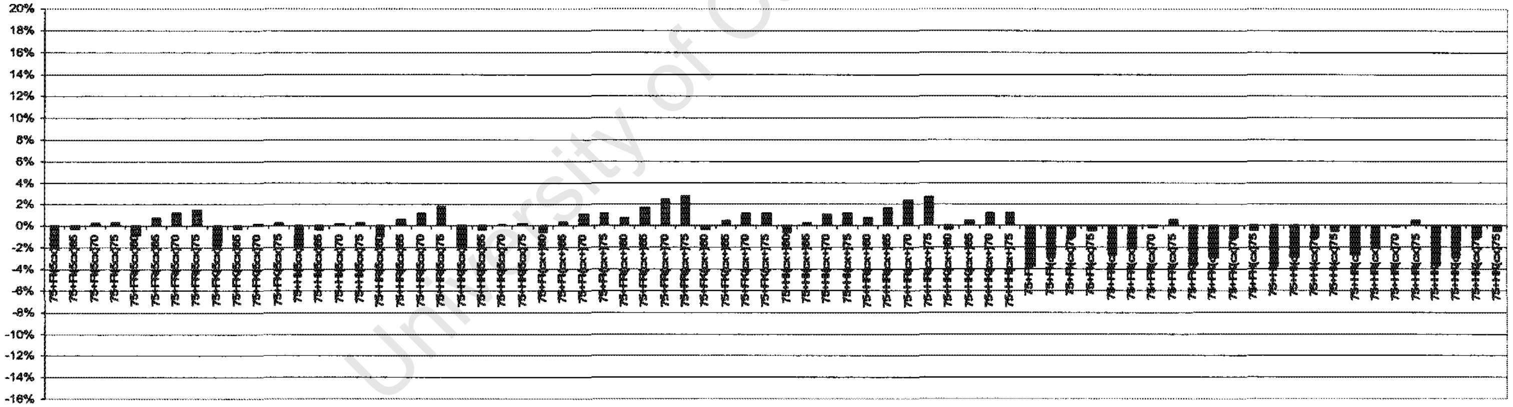

Figure 8 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 15

However, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ was a combination of N_x calculated using the method suggested by Feeney (United Nations, 2002), e_x estimated by the iterative method, using c_{x+} over the age range 15-84 to determine completeness for an open age interval of 85+.

4.3.9 Emigration + census coverage 2% decline + age misreporting in censuses (Scenario 17)

Figure 9 shows similar patterns of estimates of ${}_{45}q_{15}$ for the same age range used to determine completeness but different patterns by the method used to determine completeness of death reporting (${}_5c_x$, c_x, c_{x+}) with surprisingly different magnitudes of error.

Overall, the results suggest that using ${}_5c_x$ to determine completeness gives the largest variation of estimates compared to using c_{x+} and c_x (3.9, 3.0 and 2.3 per cent respectively). This pattern was noted for both the 75+ and 85+ open age intervals. However the estimates improved with the increase in open age interval from 60+ to 85+, although the difference is small.

For all the these results, it was concluded that the most accurate estimates are from using the open age interval of 60+ but not combined with e_x estimated using the method used by Hill and Choi and using ${}_5c_x$ to determine completeness, with iterative and known e_x, using c_{x+} to determine completeness, with e_x estimated using the method used by Hill and Choi or, known e_x combined with estimates of completeness using c_x. However, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ for scenario 17 is the combination of N_x calculated using the method used by Hill and Choi (2004), known e_x, using ${}_5c_x$ over the age range 15-59 to determine completeness for an open age interval of 60+.

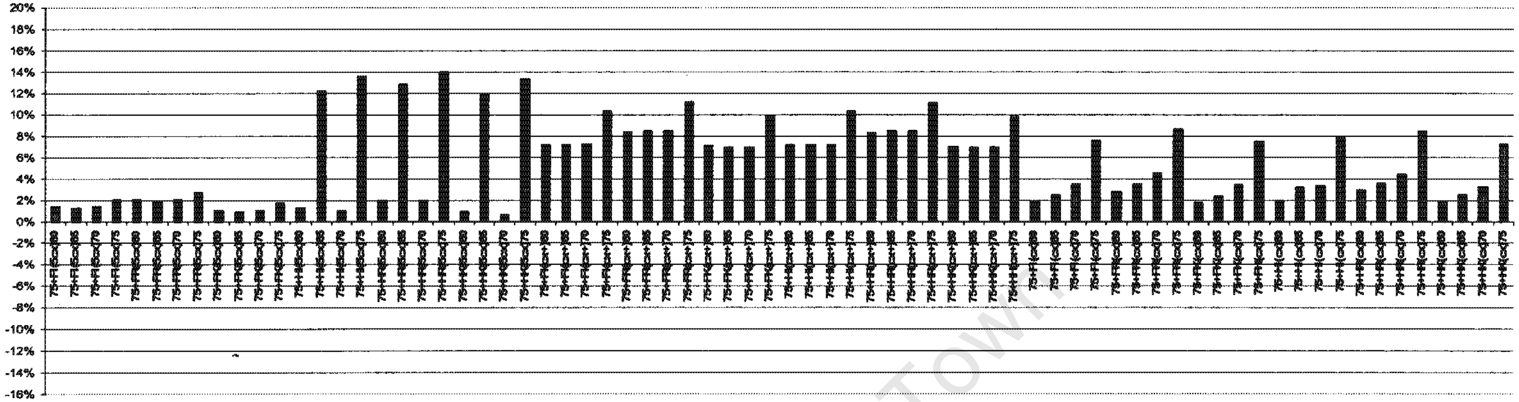

Figure 9 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 17

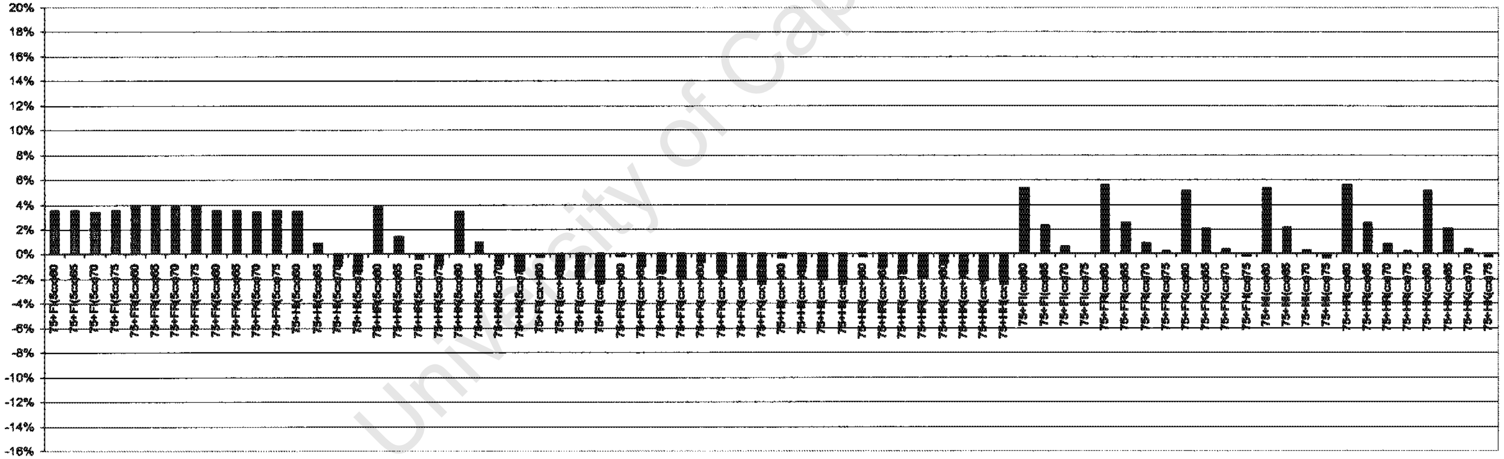

Figure 10 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 18

4.3.10 Immigration (Scenarios 18 and 19)

As is evident from Figure 10, the pattern of estimates of ${}_{45}q_{15}$ differ by method used to calculate N_x, or by method used to determine completeness (${}_5c_x$, c_{x+}, c_x), or by age range used to determine completeness. There is little or no variation between estimates of ${}_{45}q_{15}$ for the method used to determine completeness (almost two per cent). In addition, the pattern of estimates of ${}_{45}q_{15}$ did not differ by the open age interval used, with the exception of the open age interval of 60+. The estimates of ${}_{45}q_{15}$ from combinations with N_x calculated using the method used by Hill and Choi have less variation (2.1 per cent) as compared to the estimates with N_x calculated using the method suggested by Feeney (2.6 per cent). The estimates of ${}_{45}q_{15}$ improved (closer to the true ${}_{45}q_{15}$) when the open interval is lowered to 75+.

As a result of this, the best fit came from using c_x over age ranges 15-69 and 15-74 to determine completeness for the open age intervals 75+ and 85+ or, using c_{x+} over the age range 15-59 to determine completeness for an open age interval of 75+. Estimates of ${}_{45}q_{15}$ are slightly worse using e_x estimated using the method used by Hill and Choi and using ${}_5c_x$ over age ranges 15-64 and 15-69 to determine completeness for an open age interval of 85+ or, using known e_x, using c_{x+} to determine completeness for an open age interval of 60+.

However, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ for scenario 18 and 19 was a combination with N_x calculated using the method suggested by Feeney, e_x estimated using the iterative method, using c_x over the age range 15-74 to determine completeness for an open age interval of 75++(see Appendix 3 and Appendix 4).

4.3.11 Immigration + census coverage 2% decline + age misreporting in censuses (Scenarios 20 and 22)

Generally all models (combinations) produced poor estimates. However the most accurate estimates of ${}_{45}q_{15}$ are from a combination of N_x calculated using the method suggested by Feeney, known e_x, using ${}_{5}c_x$ over the age range 15-74 to determine completeness for an open age interval of 85+, and also from the combination of e_x estimated using the method used by Hill and Choi, using c_{x+} over the age range 15-59 to determine completeness for an open age interval 60+. Generally combinations with using c_{x+} to determine completeness for open age interval of 60+ produces most accurate estimates of ${}_{45}q_{15}$ with least accurate estimates coming from using ${}_{5}c_x$ to determine completeness. The estimates also differ by method of estimating e_x.

Further analysis reveal that the estimates of ${}_{45}q_{15}$ become worse when the open age interval is lowered to 75+, but the general patterns across all combinations remain the same. Thus the change in open interval impacted on the magnitude of estimates of ${}_{45}q_{15}$. The same pattern was noted when the open age interval was further lowered to 60+; thus the higher the open age interval the better the estimates of ${}_{45}q_{15}$.

On the whole, for scenario 20 and 22, the combination which results in the most accurate estimate of ${}_{45}q_{15}$ was a combination of method of N_x calculated using the method suggested by Feeney, e_x estimated using the method used by Hill and Choi, using c_{x+} over the age range 15-59 to determine completeness for an open age interval of 85+(see Appendix 3).

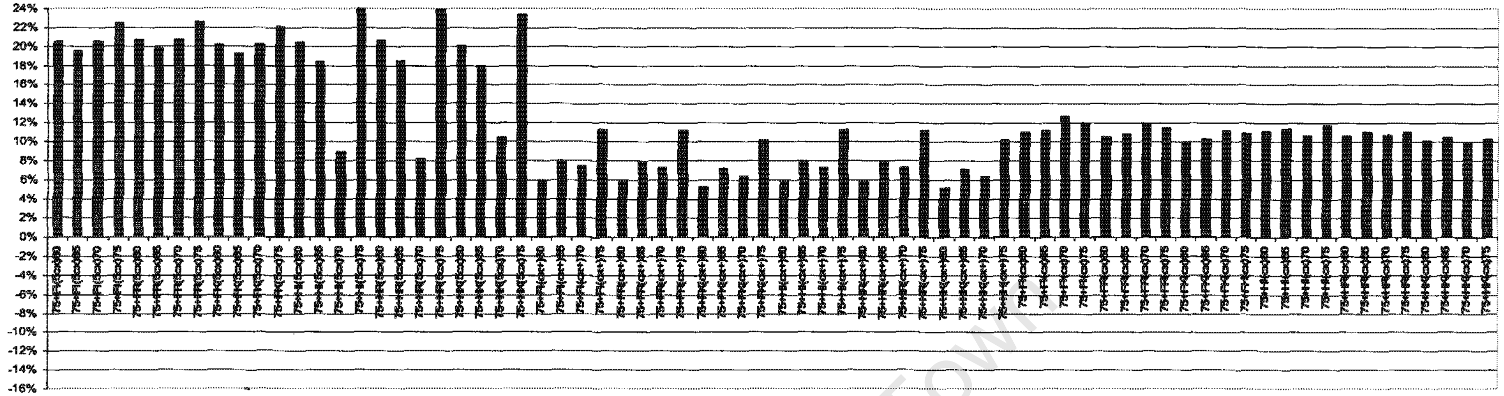

Figure 11 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 20

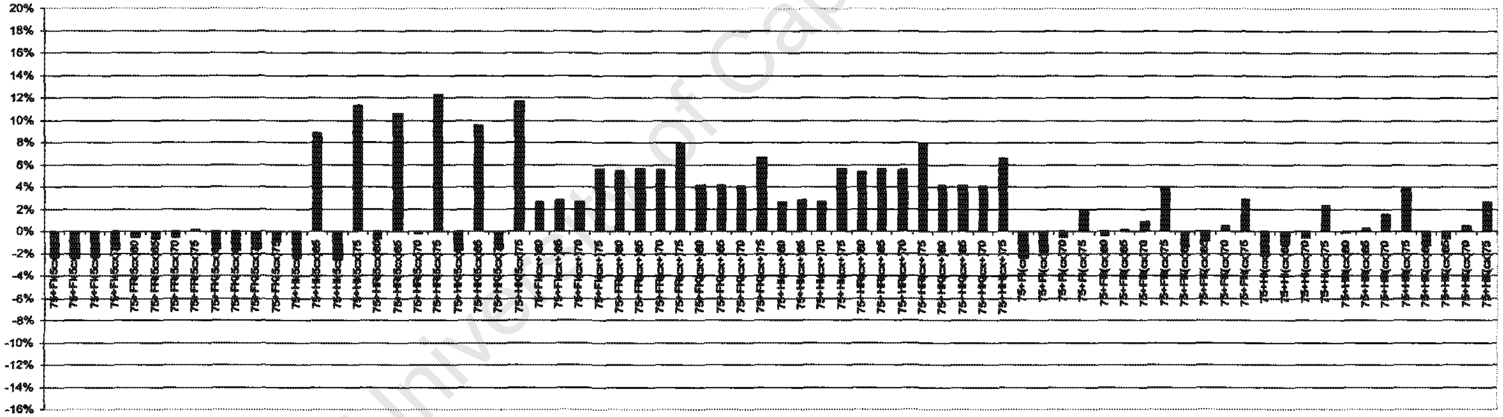

Figure 12 Percent difference between the estimated and true 45q15 by different combinations of key components of the SEG+delta method: Scenario 21

4.3.12 Emigration + age misreporting in censuses + age misreporting in VR + census coverage 2% decline + VR 20% omission (Scenario 21)

As Figure 12 shows, the pattern of estimates of $_{45}q_{15}$ differ by the method used to determine completeness of death reporting ($_5c_x$, c_{x+}, c_x) and by age range for the method used to calculate N_x. The estimates of $_{45}q_{15}$ from the combinations of key components of the SEG+delta method with N_x calculated using the method suggested by Feeney, outperformed estimates from combinations with N_x calculated using the method used by Hill and Choi.

Collectively, the most accurate estimates were from a combination of using c_{x+} over age ranges 15-59, 15-64, 15-69 and 15-79 to determine completeness ; a combination of using c_x over age range 15-74 to determine completeness with an open age interval of 85+; a combination of using c_x over age ranges 15-59, 15-64 and 15-69 to determine completeness for an open age interval of 75+ (excluding combinations of e_x estimated using the iterative method, using c_x over age ranges 15-59 and 15-64 to determine completeness). However, the combination which results in the most accurate estimate of $_{45}q_{15}$ was a combination of N_x calculated using the method suggested by Feeney, known e_x, using c_{x+} over the age range 15-79 to determine completeness for an open age interval of 85+(see Appendix 3 and Appendix 4).

4.3.13 Conclusion

Overall, the age range of 15-69 outperformed the other five age ranges used to determine completeness, with least variation in most combinations with other key components of the SEG+delta method. For combinations using e_x estimated using the iterative method over the age range 15-69, outperformed the other two methods used to estimate the life expectancy at the three open age intervals. There is an insignificant difference in estimates of $_{45}q_{15}$ for the two methods used to calculate N_x (the method suggested by Feeney (United Nations, 2002) and the method used by Hill and Choi (2004)).The open age interval of 85+ outperformed the other two open age intervals. Looking at the performance of the methods used to determine completeness, $_5c_x$ and c_{x+} performed better than c_x. However the difference in performance is small.

Looking at scenarios with migration, following the pattern assumed in the scenarios, the best age ranges to be used for estimating completeness when migration is known are 15-69 or 15-74, since these outperformed the other age ranges. They both have least variation of estimates and the most accurate estimates of ${}_{45}q_{15}$ for scenarios 15 and 18. In conclusion, given that the migration pattern is known in a population the combination which results in the most accurate estimate of ${}_{45}q_{15}$ is the combination of N_x calculated using the method either suggested by Feeney (2002) or using the method used by Hill and Choi (2004), using e_x estimated by the iterative method, using c_{x+} over the age range 15-69 to determine completeness for an open age interval of 85+.

Finally the combination which result in the most accurate estimates of ${}_{45}q_{15}$ across all the scenarios is a combination of N_x calculated using the method suggested by Feeney (United Nations, 2002), using e_x estimated by the iterative method, using c_{x+} over the age range 15-69 to determine completeness for an open age interval of 85+.

Chapter 5 African error scenarios

This chapter considers what changes need to be made to the Hill and Choi scenario data set in order to test the applicability of the method to sub-Saharan countries. Thus it considers first the suitability of the population age structure and the numbers of deaths by age to reflect the impact of an HIV/AIDS epidemic of a number of years, and then it considers the patterns of scenarios (age distortions and migration by age) in a selection of sub-Saharan African countries.

A number of authors (Bhat, 2002; Dorrington and Timaeus, 2008; Hill and Choi, 2004) expressed concerns about potential sources of bias in data from the less developed countries (age errors in deaths and censuses).

Most studies done on the death distribution methods formally or intuitively applied simulated age distortions. Ewbank (1981) proposed a method to estimate the extent of systematic errors (age misreporting) in different countries using the age data.

The distribution of the age misreporting was undertaken by assessing the comparison of recorded age distribution per observation with an estimated or stable age distribution (Ewbank, 1981). The criterion they employed was seen to have critical limitations. Limitations worth noting are the fact that identifying age reporting errors is difficult. This is complicated further by the fact that reported age distributions are also distorted by omissions in the populations in various age groups.

Ewbank (1981), Pullum (2006) analysed DHS data (1985-2003). They measured the accuracy of data by age using the Whipple's Index and the Index of Concentration. The Whipple's Index was used to assess the general degree of age heaping at ages ending in 0 or 5 in a population (United Nations, 1955). The index essentially compares the reported population at ages ending in the target digits with the proportion expected on the assumption that population is a linear function of age. For a particular age range often 23 to 62, the population with ages ending in the target digits is divided by one tenth of the total population, the result multiplied by 100 and divided by a number of different target digits. A value of 100 indicates no preference for those digits; wherever values are over 100 indicate positive preference of them.

Ewbank (1981) compared the age distribution from censuses and post-enumeration surveys. In addition, Ewbank compared ages reported in censuses and surveys with those shown on birth certificates. The age specific sex ratios of the enumerated population were model, for the years 2000 and 2005. This would reflect an African population with a prevalence of HIV in adults aged 15-49 of around 17 per cent, on which the impact of African data errors different from the original set by Hill and Choi (2004) might be measured. The data would be simulated after setting the migration in the ASSA model to zero. An example of how the scenarios could be created is given by Dorrington and Timaeus (2008) (and described in Appendix 2).

The ASSA2003 AIDS and demographic model uses a wide range of empirical evidence from different sources to provide the number of South Africans directly affected by HIV/AIDS. In addition the projections allow for the impact of major current interventions.

The second stage of the development of the African dataset is to determine whether the age errors (particularly age exaggeration at the older ages) applied by Hill and Choi differ from the age errors found in sub-Saharan African countries. To accomplish this, the censuses from seven sub-Saharan African countries and their midyear population estimates are used to determine the pattern of distortions found in practice. The census data were obtained from US Census Bureau and the midyear population estimates were obtained from UN Population Division. The censuses and their respective mid-year population estimates were scaled to the same total population and adjusted to have the same age groups and open age interval.

The age errors are determined by dividing the census by the respective mid-year population estimates for each age group and the weighted average of the age errors of the seven countries is determined. The patterns of errors by age for the seven countries were compared to the age error pattern applied by Hill and Choi (2004).

5.2 Age errors

As indicated earlier, the sample of countries to be examined in the study is chosen to represent age distortion in the most recent censuses in the sub-Saharan African region.

Looking at the female age errors of the seven countries (Figure 13), we see a slight difference in the distribution and magnitude of errors of the census from UN Population Division estimates amongst the seven countries across all age groups. However, the general pattern is similar across the selected countries. There are no consistent or significant age errors at ages 15-64, but rising errors after the age 65. As can be seen in Figure 13, there are similarities in the pattern of age errors applied by Hill and Choi (2004) and those found in sub-Saharan African countries.

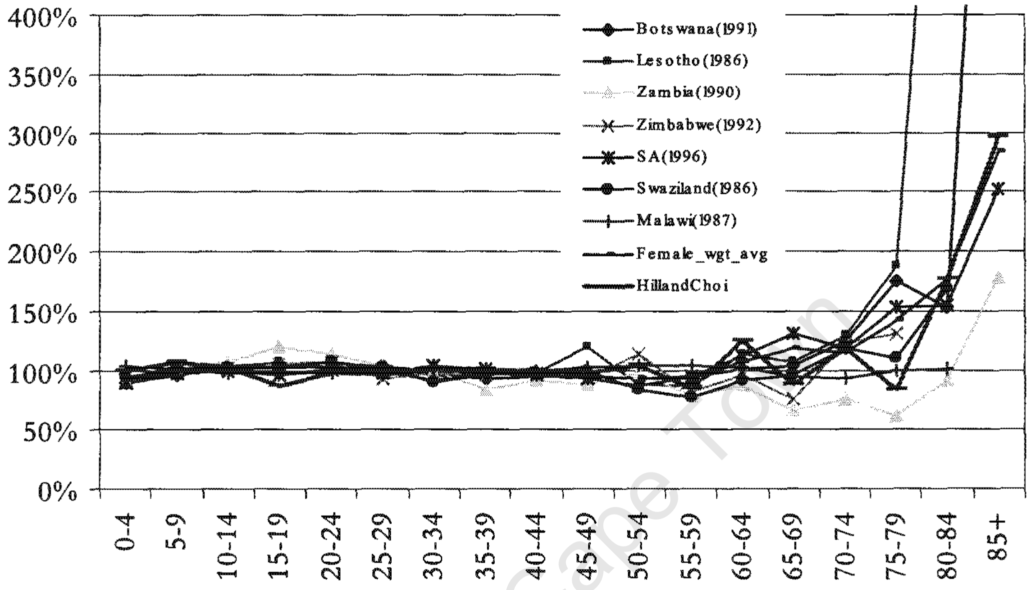

Figure 13 Sub-Saharan African Distortions for Females

An analysis of the pattern of age errors after the age 65, according to the seven countries given, reveals that the highest age errors are obtained in the open interval of each country.

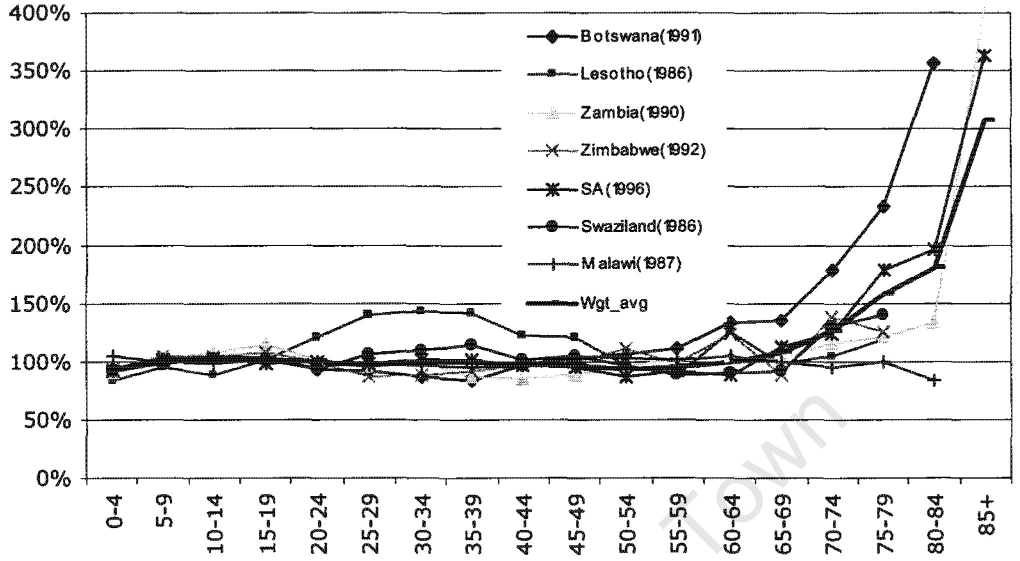

Figure 14 Sub-Saharan African Distortion for males

Examination of the pattern for males reveals almost the same pattern of age errors to that of females for the 15-64 age groups (ratios of 96 per cent to 100 per cent); however the magnitude appears to be higher for males than for females across this age range (15-64). In addition, age errors in males reveal a smooth upward pattern after age 65 as compared to the zig-zag upward pattern apparent in females.

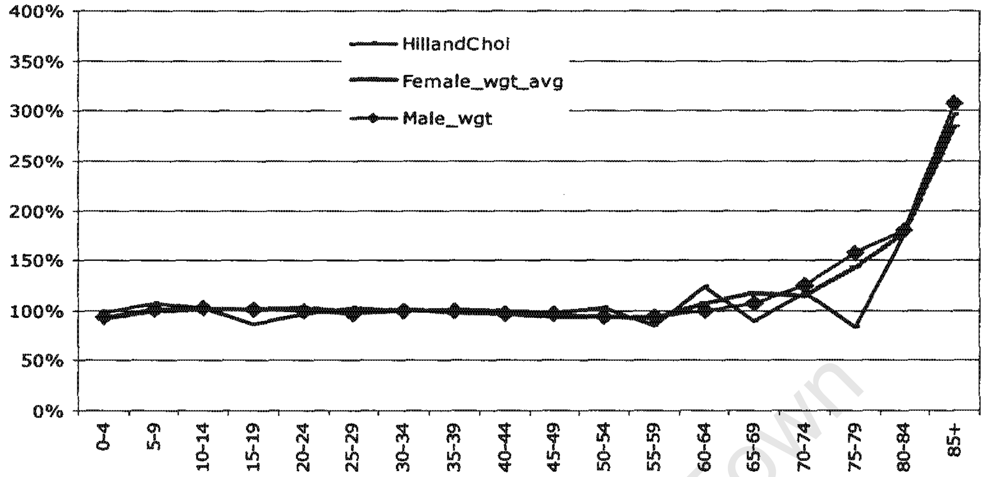

Figure 15 Hill and Choi and the Sub-Saharan African General Pattern

The pattern of age errors applied by Hill and Choi is entirely consistent with the weighted averages of the sub-Saharan African distortions apart from the zig-zag pattern which is not apparent in an averaged pattern of sub-Saharan African region.

Overall, the findings of the analysis tell us the general pattern of the African age errors (digit preference and age exaggeration, transfers across boundaries), in the sub-Saharan African region is similar to the pattern applied by Hill and Choi (2004). Taken together, these findings suggest that there is no significant difference (except of the zig-zag pattern at older ages which is found in individual countries) in general pattern of age errors of the sub-Saharan African region and the pattern applied by Hill and Choi (2004).

5.3 Migration

Following the determining of the pattern of age errors representative of African censuses, the third stage is to determine the age distribution of migration for the same seven countries from the sub-Saharan African region. The purpose of this aspect of the research is to investigate the primary difference between migration pattern applied by Hill and Choi (2004) and the migration patterns found in the sub-Saharan African region.

The absence of recent research on African migration complicates matters, thus the age distribution of migrants implied by the UN Population Division estimates of population and deaths for sub-Saharan African countries were investigated (understanding that these are not empirical estimates but rather merely assumptions from a model, designed to reproduce estimates consistent with census populations). Survival factors are calculated from the population estimates and the deaths from the 5-year period as given by the UN Population Division (United Nations, 2006). The cohort component method is used to project the population for two five-year periods, 1995 to 2000 and 2000 to 2005. For the purpose of this study, a comparison of the projected population without migration and the UN Population Division estimates gives the number of surviving migrants for the given period. The age distribution of the migrants and the migration pattern is determined. The resultant patterns from the seven countries are compared to the patterns applied by Hill and Choi (2004).

In this section, the pattern of net migration experienced by sub-Saharan African countries and the pattern postulated by Hill and Choi (2004) are compared.

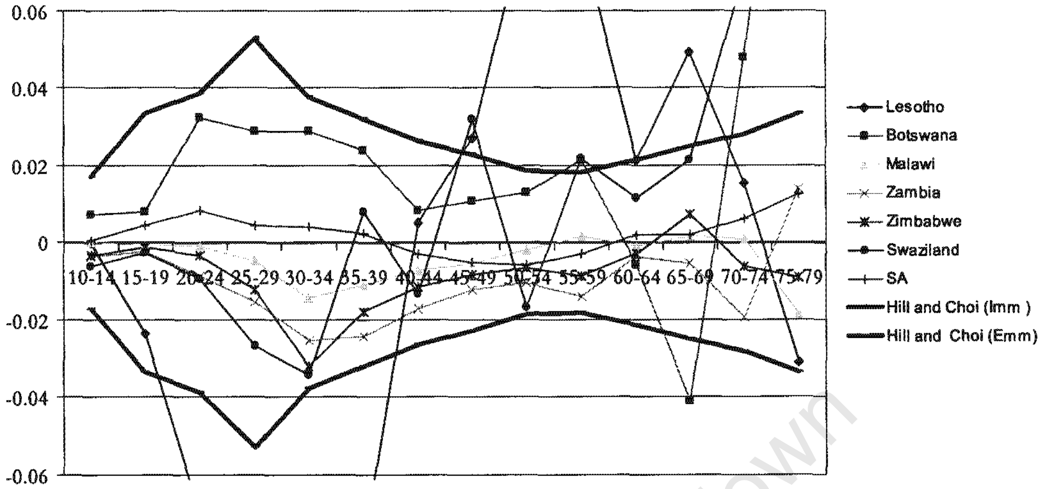

Figure 16 Net migration, females 1995 - 2000

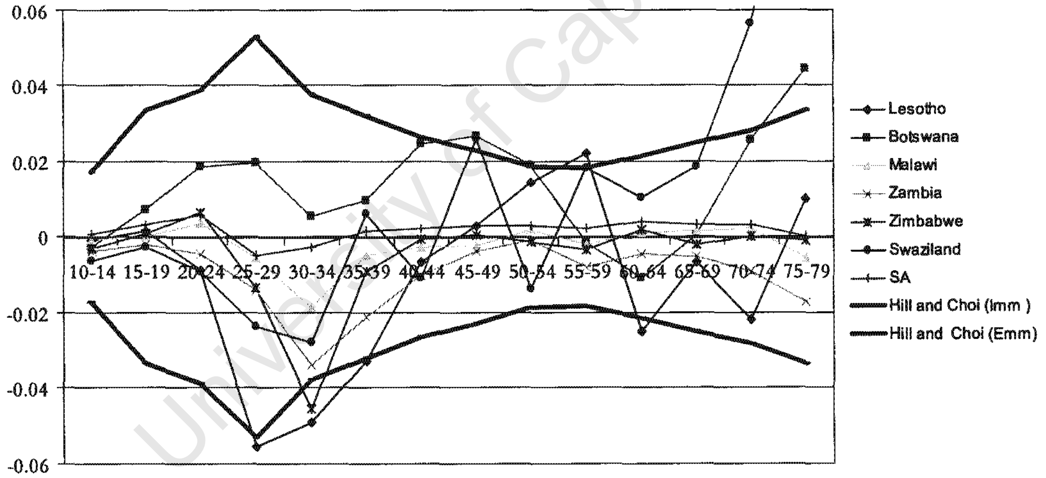

Figure 17 Net migration, females 2000 – 2005

As Figure 16 and Figure 17 show, an analysis of the seven countries reveals that there is no general pattern of net migration in the sub-Saharan African region. Some countries have net immigration patterns and some countries have net emigration patterns. In addition, net migration differs by age group for the seven countries. Taken together, these findings suggest that the general pattern of migration applied by Hill and Choi (2004) is higher than that of the sub-Saharan African region at many ages (Figure 16). Results from most of the countries reveal that at the old ages the rates of migration are low, while the migration pattern of Hill and Choi (2004) shows an upward trend after age 60.

The same results were noted for rates of migration for males for 1995-2000 and 2000-2005 periods (Figures 18 and 19).

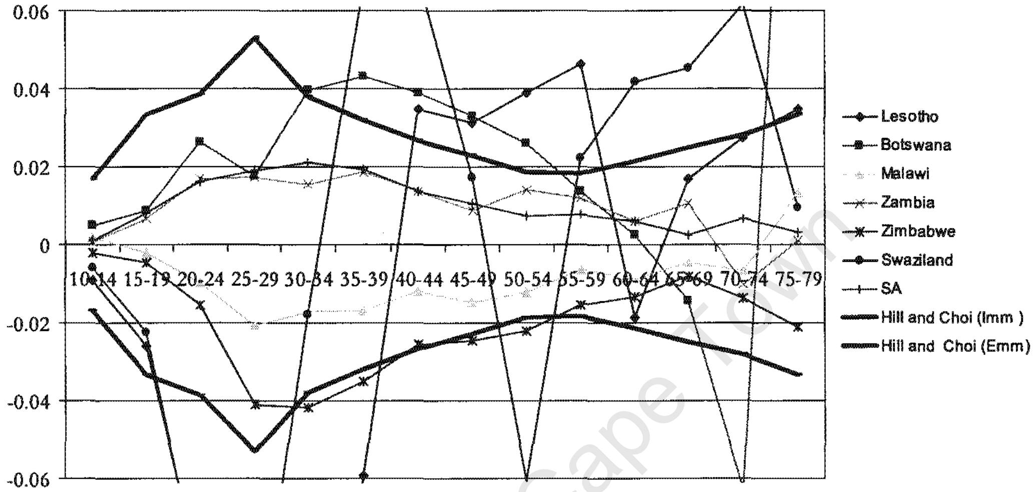

Figure 18 Net migration, males 1995 to 2000

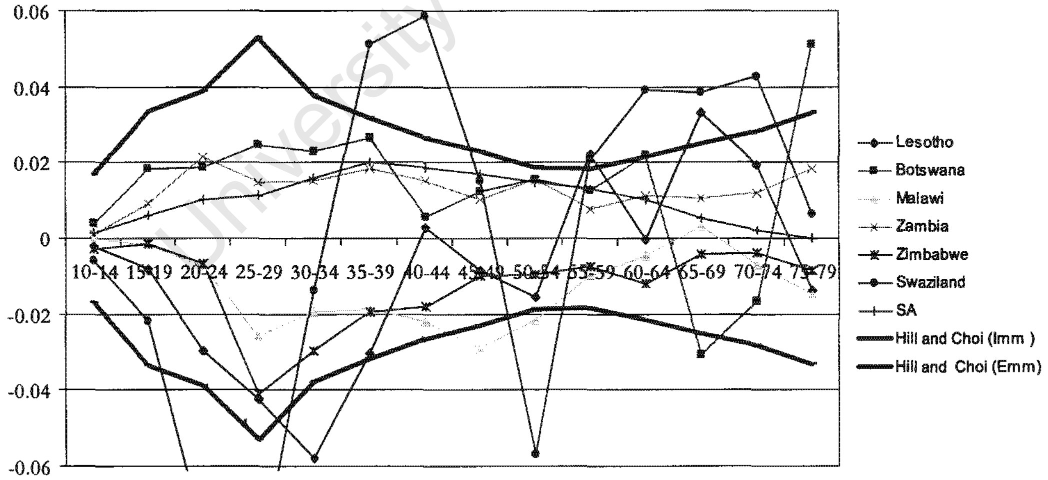

Figure 19 Net migration, males 2000 to 2005

As can be seen from Figures 18 and 19, migration rates tend to vary considerably by age and between countries. It was noted that the variation is great for small countries within South Africa (Lesotho and Swaziland) and for males in particular, these countries show much higher out migration in the working ages (20-59). However, old ages have low net migration and deaths are high (not so much with AIDS), thus adjusting the migration pattern by Hill and Choi (2004) is not expected to affect mortality estimates. From this, it would be sensible to retain the migration pattern applied to the original data set.

Chapter 6 Discussion and Conclusions

6.1 Introduction

This chapter presents the discussion of results of investigations on key components of the SEG+delta method developed in Chapter Four. The aims of this study were to determine the differences in performance of different combinations of the SEG+delta method when applied to the 23 error scenarios developed by Hill and Choi. In addition, the research also investigated the primary difference between the patterns and levels of migration and age exaggeration applied by Hill and Choi (2004) and those found in the sub-Saharan African region. The research also compared the conclusions made by Dorrington and Timaeus (2008), and Hill and Choi (2004) to the conclusions reached when different combinations of components of the SEG+delta method are applied to the 23 error scenarios.

6.2 Summary of results of the performance of SEG+delta method

The estimates of ${}_{45}q_{15}$ from the combination of different key components of the SEG+delta method revealed that even in perfect data, there is still variation in the estimates of ${}_{45}q_{15}$. Looking at the general pattern of estimates from the no error scenario, it is seen that the estimates are very close to each other and very close to the known value; the small percentage differences are seen to be as a result of the methods used to estimate the life expectancy at the age of the open interval, methods used to estimate N_x and the age intervals used to determine completeness.

Not only are the most accurate estimates obtained for the no error scenario, but also for scenarios where the errors are those for which the SEG+delta method was designed to deal with, scenarios where there is either uniform differential coverage by censuses or uniform omission of deaths by age, separately and together (estimates from scenario 6, 8 and 11 are close to the true value). The small percentage difference in the estimates of ${}_{45}q_{15}$ from scenarios 6, 8 and 11 are for the same reason for the differences in the estimates from the no error scenario. It was noted that the differences in estimates of ${}_{45}q_{15}$ across all these scenarios are small, thus most combinations give similar answers. The conclusions above for Scenarios 0, 6, 8 and 11 are the same as the conclusions reached by Dorrington and Timaeus (2008), and Hill and Choi (2004).

It is evident from Figure 1 that the most accurate estimates of ${}_{45}q_{15}$ for these scenarios all come from the same combination of key components of the SEG+delta method (namely, estimating e_x from the iterative method, N_x calculated using the method used by Feeney (United Nations, 2002), using c_{x+} over the age range 15-64 to determine completeness for an open age interval of 85+).

As has been observed before by Hill and Choi, and Dorrington and Timaeus, the estimates of ${}_{45}q_{15}$ become worse when the scenarios include other errors (e.g. when combined with age misreporting in the censuses and age misreporting in the reported deaths), in scenarios 12-14. The estimates of ${}_{45}q_{15}$ in scenarios 12-14 are close to and have the same pattern to those from scenario 5. Thus the pattern and variation of estimates from scenarios 12-14 are as a result of age misreporting in census.

The estimates of ${}_{45}q_{15}$ had similar patterns across all combinations of key components of the SEG+delta method, but the estimates are not that close (the SEG+delta method did not perform well) in Scenarios 1, 3 and 7 (scenarios where census coverage varies with age), with surprisingly different magnitudes. The most accurate estimates of ${}_{45}q_{15}$ for the Scenarios 1, 3 and 7 come from a combination of N_x calculated using the method used by Hill and Choi (2004), e_x estimated using the iterative method, using c_x over the age range of 15-69 to determine completeness for an open age interval of 75+.

The SEG+delta method continued to perform worst (as found by Dorrington and Timaeus) in Scenarios 20 and 22 where there is a combination with immigration and age misreporting in census. The pattern of estimates from the two scenarios is similar. However the combination which results in the most accurate estimate of ${}_{45}q_{15}$ is a combination of N_x calculated using the method used by Feeney, with e_x estimated by the method used by Hill and Choi, using c_{x+} over the age range 15-59 to determine completeness for an open age interval of 60+. These conclusions are the same as those reached by Dorrington and Timaeus, however the combination that gave the most accurate estimate of ${}_{45}q_{15}$ used the e_x estimated by the method used by Hill and Choi for an open age interval of 60+, while Dorrington and Timaeus used the $e_x$estimated using the iterative method over the age range of 15-69 to determine completeness for an open age interval of 85+.

In scenarios where the migration is strongly emigration, Scenarios 17 and 21, the estimates of ${}_{45}q_{15}$ are better than those for Scenarios 20 and 22, which are strongly immigration. However the combination which results in the most accurate estimate of ${}_{45}q_{15}$

is a combination of N_x calculated using the method used by Hill and Choi (2004), with e_x estimated using the method used by Hill and Choi, completeness determined using ${}_5c_x$ over the age range 15-59 for an open age interval of 60+.

Looking at migration, where there is a known migration pattern , the results reveal that the most accurate estimates of ${}_{45}q_{15}$ are as a combination of key components of the SEG+delta method over the age ranges 15-69 and 15-74, since these outperformed the other four age ranges with least variation. In conclusion, given that the migration pattern is known in a population, the best model includes the combination of N_x calculated using the method used by Feeney (United Nations, 2002), determining completeness using c_{x+} with e_x estimated by the iterative method over the age range 15-74 for an open age interval of 85+.

Overall, the SEG+delta method performed worst where there was misreporting in censuses or where there was immigration or emigration. The study came to the same conclusions reached by Hill and Choi (2004), and Dorrington and Timaeus (2008) for the performance of the SEG+delta method but the combination which results in most accurate estimates of ${}_{45}q_{15}$ per set of scenarios differ to the ones Hill and Choi, and Dorrington and Timaeus suggested. It has been noted that some combinations work better than others, this points to the need for further research into why some combinations give better estimates than others.

The differences in estimates of e85 from the iterative method were close to the known e85. Thus the effect of the life expectancy was small. However one of the problems with using a low age of the open interval is precisely the inaccuracy of estimates of e_x at this age. The differences between the estimates of e_x from the iterative method and known e_x are too small (less than two per cent) to have much impact on the estimates of ${}_{45}q_{15}$ in this research. It is recommended to use the estimates from the iterative method proposed by Dorrington and Timaeus (2008) method in this type of study since the method used by Hill and Choi is not applicable to life tables which reflects HIV/AIDS mortality.

6.3 Limitations of the study

There are very few benchmarks against which the results of $_{45}q_{15}$ from this study can be measured. The one that does exist is the sensitivity analyses study done by Hill and Choi (2004), and Dorrington and Timaeus (2008), which investigate the performance of the SEG+delta and GGB methods, for a fixed combination of key components of these methods. However, there were critical differences between the two studies that must be taken into consideration.

The age ranges used to determine completeness in this study started from the age range of 15-59. This age group is where HIV deaths and migration are most pronounced. The SEG+delta method identifies emigrants as unreported deaths (Hill and Choi 2004). Thus there is a need to investigate the performance of different combinations of the SEG+delta method when completeness is determined using age ranges starting from other ages above 60 in trying to eliminate the effects of HIV mortality and migration (Murray, Rajaratnam *et al.*, 2009). However the research by Murray, Rajaratnam *et al* (2009) was done on simulated data for low income countries and used real data for high-income countries. In addition, the age ranges tested in this study all start with age 15, which have been shown by this research to be a poor choice especially with populations with migration. For future research, it would be worth focusing on age rages starting with 5 and on age 30 as lower limits.

When estimating e_x for the age of the open interval of 85+, it will be better to use $_5m_x$ at lower ages to remove the impact of age exaggeration on the estimates of e_x. However, when using younger ages, the estimates are affected by high HIV mortality especially in less developed countries. However the other method used to estimate e_x proposed by Dorrington and Timaeus of fitting the Gompertz curve to the old age mortality (assuming that the mortality rates follow a Gompertz curve (Bongaarts, 2004)) was not tested in this study. Thus estimating e_x by fitting the Gompertz curve to the old age mortality needs further research.

In addition, methods of estimating e_x are affected by old age mortality, age exaggeration and high HIV mortality. Hill and Choi derived the life expectancy of the open interval from the ratio of $_{30}d_{10}$ and $_{20}d_{40}$ and a look up table based on the regression from the Princeton West Life Table (Coale, Demeny and Vaughan, 1983) produced by Bennett and Horiuchi (1984) which do not represent the mortality in less developed countries affected by HIV, e.g. sub-Saharan Africa. Thus if any pattern of age errors or age distribution and age distribution of deaths from sub-Saharan Africa, the methods do not perform well. Thus further research is needed in order to devise such a rule for populations with HIV. In addition, when applying the death distribution methods, including SEG+delta, to real data it is difficult to know the true value of the estimates (${}_{45}q_{15}$), thus it is very difficult to assess how the method fits the data; several data errors may prevent at the same time in ways not predicted by simulated analysis.

This research investigated the age errors that are common in less developed countries. The most common errors are age overstatement and age digit preference in both the reported deaths and the enumerated population, and under reporting of deaths. The method used to determine the pattern of age errors found in the sub-Saharan African region depend on standard population estimates from UN Population Division assumptions which depend on the mortality pattern similar to that of the population under study. In addition, the results of age errors at older ages found in this research should not be relied upon since it was determined that the UN Population Division underestimated the population for South Africa for the older ages (conclusions reached from the research by Machemedze (2009)).

The results revealed greater variation of migration pattern for small countries in South Africa (Lesotho and Swaziland) and for males in particular. These countries show higher out migration in the working ages, hence there is need to test if this is important when the SEG+delta method is applied is such a context.

However the pattern found from the sub-Saharan African region should not be relied upon since the validity of the assumptions applied by UN to project the populations is not known. In addition, there are no accurate estimates of international migration against which the results of the migration pattern from this study can be measured. The investigation of the pattern and level of migration at ages as high as 80 by five-year age interval is one of the limitations beyond which the research could not go. To this end, in terms of migration from the sub-Saharan African region, there is room for further research in how the migration pattern can be determined and its impact on the performance of the SEG+delta method.

6.4 Conclusions

The results include no major surprises. However, the results reveal significance differences between adjacent estimates with different age ranges (e.g. those with different age ranges used to determine completeness). This could be an artifact of the zig-zag pattern in the data. If so, this suggests an element of randomness in the outcome (estimate of ${}_{45}q_{15}$).

From the results above, the conclusion must be that some combinations work better than others in the 23 error scenarios. In addition, the combination that results in the most accurate estimate of ${}_{45}q_{15}$ differ per scenario but for scenarios with the same type of error the most accurate estimates come from the same combination of key components of the SEG+delta method. This is so since the most accurate estimates of ${}_{45}q_{15}$ for Scenarios 0, 6, 8 and 11 all come from the same combination of key components of the SEG+delta method (no error scenario, and scenarios which the SEG+delta method was designed to cope with).The above result is also apparent for Scenarios 1, 3 and 7 (scenarios where census coverage varies with age) and scenarios 20 and 22 (which give the worst estimates of ${}_{45}q_{15}$) where there is a combination of immigration and age misreporting in census.

The results of the study revealed that a lot of combinations have similar patterns of estimates of ${}_{45}q_{15}$. This is besides the scenarios where all combinations result in worst estimates of ${}_{45}q_{15}$. Thus, scenarios where all combinations results in worst estimates (scenarios 17, 20, 21 and 22) were not used to determine the combination of the key components of the SEG+delta method that results in the most accurate estimates of ${}_{45}q_{15}$ across all scenarios, since these are poor combinations per key components of the SEG+delta method.

This study concluded that in most scenarios the combination which results in most accurate estimate of ${}_{45}q_{15}$ per scenario is the same as the combination suggested by Dorrington and Timaeus (2008) . The results revealed that at the high open age intervals, the results do not differ by method used to estimate e_x. The higher the open age interval the better the estimates of ${}_{45}q_{15}$. However in most scenarios the difference in errors is small.

Given the uncertainty around estimates of life expectancy it is always better to use higher age open intervals, 85+. In addition, the estimates of ${}_{45}q_{15}$ from the SEG+delta method are slightly better for the longer age ranges (15-69 and 15-74) used to determine completeness. The findings also reveal that the two methods used to calculate N_x seem to give similar results of $_{45}q_{15}$. Thus the results do not differ by the method used to calculate N_x. This appears best to use the longer age range and calculate N_x using the method suggested by Feeney (United Nations, 2002).

With regards to age exaggeration scenarios, there is no significant difference (except of the zig-zag pattern at older ages) in the general pattern of age errors of the sub-Saharan African region and the pattern applied by Hill and Choi (2004). In addition, the general pattern of migration in the sub-Saharan African region is not significantly different from the pattern applied by Hill and Choi, except for random fluctuation (see Figure 16). Results from most of the countries revealed that at the old ages the rates of migration are low, while the migration pattern by Hill and Choi (2004) shows an upward pattern (after age 60). Thus, it was decided to retain the migration pattern and age error pattern applied to the original data set by Hill and Choi. However there is need for further research on how the mortality estimates are affected if the migration pattern by Hill and Choi is adjusted at older ages.

There are a number of areas where further research is needed. This study did not investigate fully the performance of the SEG+delta method when different methods (such as setting c to the mean of c_xs) of determining c from the c_xs, in fact only setting c to the median of c_xs was applied in this study.

Further studies could be undertaken to assess the impact of applying the irregular pattern of migration noted from countries in sub-Saharan Africa since the zig –zag pattern shown in Chapter 5 might reduce the impact of migration on the performance of the SEG+delta method. Since the patterns of migration from the sub-Saharan African estimates vary by age (net emigration at some ages and net immigration at some ages), the impact of this migration pattern might be different to the strong net migration patterns applied by Hill and Choi (2004).

Finally, one can apply the various combinations of the SEG+delta method to the African data with African distortions (migration pattern and age error pattern determined in Chapter 5) and compare with the results from previous studies done by Hill and Choi (2004), and Dorrington and Timaeus (2008). Additionally, there is a need to apply the SEG+delta method to the male data, since this study and many studies did before focused on the performance of the SEG+delta method on female data. However Dorrington and Timaeus have already applied the SEG+delta method to an African female dataset. It will be recommended to do the similar analysis for the African male data with HIV mortality.

References

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

Bennett, N. G. and Horiuchi, S. 1984. "Mortality Estimation from Registered Deaths in Less Developed Countries", *Demography* **21**(2):217-233.

Bhat, P. N. M. 1990. "Estimating Transition Probabilities of Age Misstatement", *Demography* **27**(1):149-163.

Bhat, P. N. M. 2002. "General Growth Balance Method: A Reformulation for Populations Open to Migration", *Population Studies* **56**(1):23-34.

Bongaarts, J. 2004. *Long-Range Trends in Adult Mortality: Models and Projection Methods.* Policy Research Division Working Paper No. 192. Population Council. Available:

Brass, W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data.* Chapel Hill North Carolina: Carolina Population Centre.

Brass, W. 1979. "A procedure for comparing mortality measures calculated from intercensal survival with the corresponding estimates from registered deaths", *Asian and Pacific Census Forum* **6**(2):5-7.

Coale, Demeny and Vaughan. 1983. *Regional Model Life Tables and Stable Populations.* New York: Academic Press.

Dorrington, R. and Timaeus, I. M. 2008. "Death Distribution Methods for Estimating Adult Mortality: Sensitivilty Analysis with Simulated Data Errors, Revisited" Paper presented at PAA Conference.

Ewbank, D. C. (ed). 1981. *Age Misreporting and Age Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis.* Committee on Population and Demography, Report No.4 Washington, D.C.: National Academic Press

Hill, K. 1987. "Estimating Census and Death Registration Completeness", *Asian and Pacific Population Forum* **1**(3):8-13, 23-24.

Hill, K. 2001. "Methods for Measuring Adult Mortality in Developing Countries: A Comparative Review," Paper presented at Paper presented to the International Population Conference,Salvador, Brazil.

Hill, K. 2003. "Adult Mortality in the Developing world; What we know and how we know it " Paper presented at Presented at United Nations Population Division Workshop on HIV/AIDS and Adult Mortality in Developing Countries. New York, 8th-13th September.

Hill, K. and Choi, Y. 2004. "Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors," Paper prepared for Adult Mortality in Developing Countries Workshop. The Marconi Center, Marin County, California, 8 -11 July.

Hill, K., Choi, Y. and Timaeus, I. M. 2005. "Unconventional approaches to mortality estimation", *Demographic Research* **13**(12):281-300.

Hill, K. and Queiroz, B. (2004). "Adjusting General Growth Balance Method for Migration. Paper presented AMDC. Berkeley.

Luther, N. Y. 1983. "Measuring changes in census coverage in Asia", *Asian and Pacific Census Forum* **9**(3):7-16.

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

Martin, L. G. 1980. "A modification for use in destabilised populations of Brass's technique for estimating completeness of death registration", *Population Studies* **34**(2):381-395.

Murray, C. J. L., Rajaratnam, J. K., Marcus, J. *et al.* 2009. "Reducing Ignorance about Adult Mortality: Improving Methods for Evaluating the Completeness of Death Registration," Paper presented at PAA Conference. 19 March.

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

Preston, S. and Hill, K. 1980. "Estimating the Completeness of Death Registration", *Population Studies* **34**(2):349-366.

Preston, S. H., Heuveline, P. and Guillot, M. 2001. *Demography:Measuring and Modelling Population Process.* Oxford: Blackwell Publishing.

Pullum, T. W. 2006. *An Assessment of Age and Date Reporting in the DHS Surveys 1985-2003* The University of Texas at Austin and Macro International Inc. Calverton, Maryland.

Rogers, A. and Castro, L. J. (eds). 1981. *Model Migration Schedules.* RR-81-030.Laxenburg, Austria: International Institute for Applied Systems Analysis.

United Nations. 1955. *Methods of Appraisal of Quality of Basic Data for Population Estimates.* Population Studies. Manual II, Series A, Population Studies No. 23. New York: United Nations:

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

United Nations. 2002. *Methods for Estimating Adult Mortality.* New York: Department of Economic and Social Affairs.

United Nations. 2006. *World Population Prospects: The 2006 Revision.* New York:Department of Economic and Social Affairs.http://esa.un.org/unpp/. Accessed: 6 October 2008

Vincent and 1951. "La Mortalité des Vieillards (The mortality of the aged)", *Population* **6**(2):181-204.

Appendix 1: List of data error scenarios

Scenario Number	Description
0	No error
1	Age varying census coverage (based on net-undercount of male black population from the 1980 United States Census)
2	Age misreporting in censuses (derived from a matrix of transfers between 5-year age groups estimated for India)
3	Age misreporting in censuses+age varying census coverage
4	Age misreporting in VR (derived from a matrix of transfers between 5-year age groups estimated for India)
5	Age misreporting in censuses+age misreporting in VR
6	Census coverage decline of 2%
7	Census coverage decline of 2%+ age varying census coverage
8	VR 20% omission
9	VR 20% omission+increasing completeness with age (linearly to 100% by age 85+)
10	VR 20% omission+decreasing completeness with age (linearly to 70% by age 85+)
11	Census coverage 2% decline + VR 20% omission
12	Age misreporting in censuses+age misreporting in VR+census coverage 2% decline
13	Age misreporting in censuses+age misreporting in VR+VR 20% omission
14	Age misreporting in censuses+age misreporting in VR+census coverage 2% decline+VR 20% omission
15	Emigration (based on a pattern of age-specific in-migration to the U.S. of Mexican males 1980-1990)
16	Emigration+census coverage 2% decline
17	Emigration+census coverage 2% decline+age misreporting in censuses
18	Immigration (based on a pattern of age-specific in-migration to the U.S. of Mexican males 1980-1990)
19	Immigration+census coverage 2% decline
20	Immigration+census coverage 2% decline+age misreporting in censuses
21	Emigration+age misreporting in censuses+age misreporting in VR+census coverage 2% decline+VR 20% omission
22	Immigration+age misreporting in censuses+age misreporting in VR+census coverage 2% decline+VR 20% omission

Source: Dorrington and Timaeus (2008)

Appendix 2: Creating of African Aids scenarios

1. **Age Varying Census coverage**

   Assumed the same pattern of differential coverage as used in the original scenario

2. **Age misreporting in the Census**

Applied the same Matrix of Transition Probabilities of Age Misstatements, India, 1971-1981 as was applied by Hill and Choi (2004). Thus, the same structure of the age misreporting of a particular age to the numbers in the other ages as found in the original scenario, the totals were rebalanced to sum to the original total by use of an adjustment factor.

3. **Age misreporting in the Vital Registration**

   A similar method as from Age misreporting in the Census stated in (2) was applied.

4. **Census coverage decline by 2% and 20% omission in Vital Registration**

   Applied change to the complete data. The actual linear and exponential model used for the 20% omission in Vital Registration was determined by using a solver function (was done in such a way that the 20% omission after applying exponential or linear model remains the same as the 20% omission of the true population)

5. Increasing completeness with age (to 100% by age 85+, and to 70% by age 85+). Applied the same adjustment process as used in the original data set (this is designed to ensure linearly changing completeness while maintaining the same overall level of completeness.

6. **Emigration**

   True population after emigration = true population without + migration rate per true population from the original data set times the true African population. True deaths allowing for emigration = true deaths without allowing for emigrants – death rate from the original times the number of emigrants in each age

7. **Immigration**

   True population after immigration = true population without - migration rate per true population from the original data set times the true African population. True deaths allowing for immigration = true deaths without allowing for emigrants + death rate from the original times the number of immigrants in each age

8. All the combinations were done using the above created scenarios for the 2000 population, 2005 population and the intercensal deaths (2000 to 2005).

   Source: Dorrington and Timaeus (2008)

Appendix 3: Best combination per scenario which resulted in the most accurate of ${}_{45}q_{15}$

Error scenario	Least % Difference	Combination	Open age interval	Method of calculating N_x	Method of estimating e_x	Method of determining completeness	Fitted age interval
No error	0.006%	75+FH(cx)75	75+	Feeney	Hill and Choi	cx	15-74
1	0.136%	75+HI(cx)70	75+	Hill and Choi	Iterative	cx	15-69
2	0.005%	60+FH(cx+)60	60+	Feeney	Hill and Choi	cx+	15-59
3	2.890%	75+HK(5cx)65	75+	Hill and Choi	Known	5cx	15-64
4	0.032%	60+HH(5cx)60	60+	Hill and Choi	Hill and Choi	5cx	15-59
5	0.023%	85+HH(5cx)65	85+	Hill and Choi	Hill and Choi	5cx	15-64
6	0.018%	75+FH (cx)75	75+	Hill and Choi	Hill and Choi	cx	15-74
7	0.152%	75+FI(cx)70	75+	Feeney	Iterative	cx	15-69
8	0.008%	75+FH(cx)75	75+	Feeney	Hill and Choi	cx	15-74
9	9.064%	75+HK(5cx)65	75+	Hill and Choi	Known	5cx	15-64
10	5.770%	75+FK(5cx)65	75+	Feeney	Known	5cx	15-64
11	0.009%	75+FH(cx)75	75+	Feeney	Hill and Choi	cx+	15-74
12	0.009%	85+HH(5cx)65	85+	Hill and Choi	Hill and Choi	5cx	15-64
13	0.023%	85+HH(5cx)65	85+	Hill and Choi	Hill and Choi	5cx	15-64
14	0.009%	85+HH(5cx)65	85+	Hill and Choi	Hill and Choi	5cx	15-64
15	0.009%	85+FI(cx+)85	85+	Feeney	Iterative	cx+	15-84
16	0.023%	85+FI(cx+)85	85+	Feeney	Iterative	cx+	15-84
17	0.302%	60+HK(5cx)60	60+	Hill and Choi	Known	5cx	15-59
18	0.046%	75+FI(cx)75	75+	Feeney	Iterative	cx	15-74
19	0.069%	75+FI(cx)75	75+	Feeney	Iterative	cx	15-74
20	0.191%	85+FH(cx+)60	85+	Feeney	Hill and Choi	cx+	15-59
21	0.054%	85+FK(cx+)80	85+	Feeney	Known	cx+	15-79
22	0.191%	85+FH(cx+)60	85+	Feeney	Hill and Choi	cx+	15-59

Appendix 4: Average Deviation of estimates of ${}_{45}q_{15}$ by age of the open interval and by the fitted age range

Scenario	Average Deviation Open Age Interval			Average Deviation by age Interval					
	85+	75+	60+	15-59	15-64	15-69	15-74	15-79	15-84
No Error	0.3%	0.3%	1.1%	0.5%	0.2%	0.2%	0.3%	0.2%	0.3%
1	6.8%	7.7%	5.4%	7.6%	6.3%	5.8%	4.2%	1.3%	1.4%
2	4.2%	3.5%	4.2%	3.2%	3.5%	2.3%	4.0%	3.3%	4.2%
3	11.6%	9.7%	2.4%	9.0%	10.0%	8.5%	9.1%	8.8%	6.0%
4	2.5%	0.9%	1.5%	1.8%	1.7%	2.0%	2.1%	2.4%	2.6%
5	4.0%	4.1%	4.3%	4.2%	4.3%	2.4%	5.9%	1.9%	3.0%
6	0.3%	0.4%	1.6%	0.8%	0.2%	0.3%	0.3%	0.2%	0.2%
7	6.8%	7.8%	5.5%	7.7%	6.3%	5.8%	4.2%	1.3%	1.4%
8	0.3%	0.3%	1.2%	0.5%	0.2%	0.2%	0.3%	0.2%	0.2%
9	0.6%	0.3%	0.5%	0.4%	0.3%	0.3%	0.3%	0.4%	0.5%
10	0.2%	0.4%	1.4%	0.7%	0.3%	0.3%	0.3%	0.2%	0.1%
11	0.3%	0.4%	1.6%	0.8%	0.2%	0.3%	0.3%	0.2%	0.2%
12	4.4%	4.1%	3.8%	4.0%	4.3%	2.4%	5.9%	1.9%	4.6%
13	3.7%	4.1%	3.9%	3.9%	4.3%	2.4%	5.9%	1.9%	1.3%
14	4.4%	4.1%	3.9%	4.0%	4.3%	2.4%	5.9%	1.9%	4.6%
15	0.8%	1.3%	1.9%	1.4%	1.3%	0.7%	0.6%	0.2%	0.3%
16	0.8%	1.3%	2.4%	1.6%	1.3%	0.7%	0.7%	0.2%	0.3%
17	4.1%	3.4%	2.8%	2.8%	3.7%	2.7%	2.9%	3.7%	4.9%
18	2.4%	2.1%	1.6%	2.1%	1.6%	1.6%	1.8%	1.6%	1.7%
19	2.4%	2.1%	1.7%	2.1%	1.6%	1.7%	1.6%	1.6%	1.7%
20	4.6%	4.7%	7.5%	5.8%	4.4%	3.3%	5.3%	4.2%	4.8%
21	2.2%	3.2%	2.6%	2.2%	3.8%	1.9%	3.6%	1.7%	2.1%
22	4.6%	4.5%	7.5%	5.9%	4.3%	3.1%	5.3%	4.2%	4.8%
Overall	2.5%	3.2%	2.4%	2.2%	3.5%	2.0%	2.9%	1.6%	1.7%

Appendix 5: Average Deviation of estimates of 45q15 by method used to estimate ex and method used to determine completeness

Scenario	Method used to estimate ex			Method used to determine completeness		
	Iterative	Hill and Choi	Known ex	$_5c_x$	C+	c_x
No Error	0.2%	0.6%	0.2%	0.3%	0.4%	0.4%
1	7.9%	7.5%	7.8%	6.1%	2.8%	7.2%
2	3.9%	3.8%	4.2%	6.2%	3.4%	0.4%
3	11.7%	11.5%	11.5%	7.4%	7.6%	8.9%
4	2.4%	3.3%	2.7%	2.0%	4.0%	1.9%
5	4.4%	4.0%	4.3%	7.0%	2.6%	1.4%
6	0.2%	0.8%	0.2%	0.4%	0.5%	0.4%
7	7.9%	7.5%	7.8%	6.1%	2.7%	7.2%
8	0.2%	0.6%	0.2%	0.3%	0.4%	0.4%
9	0.5%	0.8%	0.5%	0.3%	0.6%	0.5%
10	0.2%	0.7%	0.3%	0.4%	0.4%	0.3%
11	0.2%	0.8%	0.2%	0.4%	0.5%	0.4%
12	4.3%	4.2%	4.4%	6.5%	2.7%	1.3%
13	4.1%	4.0%	4.2%	7.2%	2.6%	1.2%
14	4.3%	4.2%	4.4%	6.5%	2.7%	1.3%
15	1.2%	1.1%	1.1%	0.8%	0.7%	1.3%
16	1.2%	1.2%	1.1%	0.9%	0.8%	1.4%
17	4.2%	3.8%	4.1%	3.9%	3.0%	2.3%
18	2.4%	2.5%	2.5%	2.0%	1.1%	2.3%
19	2.4%	2.5%	2.4%	2.0%	1.2%	2.4%
20	4.9%	4.9%	4.7%	3.7%	3.3%	0.8%
21	2.7%	2.9%	2.9%	4.0%	2.6%	1.5%
22	4.8%	4.9%	4.9%	3.5%	3.4%	0.8%