

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

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

**Estimating adult mortality in South Africa using
orphanhood and year of death data from the 2008
National Income Dynamics Study**

By

Takura Kupamupindi

A dissertation submitted to the faculty of commerce in partial
fulfilment of the requirements for the degree of Master of
Philosophy in Demography

Centre for Actuarial Research
University of Cape Town
December 2010

PLAGIARISM DECLARATION

I, Takura Kupamupindi hereby declare that the work on which this thesis is based is original, except where otherwise cited, and that neither the whole nor any part of it has been submitted for another degree at this or any other university.

I empower the University of Cape Town, to reproduce for the purpose of research either the whole or any portion of its contents.

Signature

Date

University of Cape Town

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof Rob Dorrington for his patience and tireless effort in guiding me from the beginning to the end of my thesis. I would also like to thank Associate Prof Tom Moultrie for the knowledge he gave me during my coursework and for everything he has done for me.

Special thanks goes to my wife Venah for her love and support, to my parents Mr and Mrs Kupamupindi for everything they have done for me and their understanding in my long absence, to my siblings; Patience, Nathan and Sam for all the encouragement and help, to my in-law Mr Kasayi for his help and support, to my 'demography' friends for making this journey a memorable one, and to all my relatives and friends who directly or indirectly supported me.

I am very grateful to the Andrew W Mellon and Hewlett Foundation and the University of Cape Town (UCT) Postgraduate Funding Office for their financial assistance which facilitated my studies at UCT.

Finally I would like to thank God for giving me the gift of life, for giving me this opportunity to further my studies, for giving me hope, wealth and the wisdom.

ABSTRACT

The overall objective of this research is to investigate whether using year of death data to produce estimates of time location is a better approach than the method developed by Brass and Bamgboye (1981) and whether estimates of mortality produced using year of death data are any better than those derived using the conventional orphanhood method. In this research, year of death data from the 2008 National Income Dynamics Study (NIDS) are used to estimate the time location of mortality rates as suggested by Chackiel and Orellana (1985). Estimates of mortality derived from using year of death data are then compared to estimates derived using the conventional orphanhood methods from the 1996 and 2001 censuses, the 2007 Community Survey and the 2008 NIDS survey. Estimates of mortality rates over time derived from the ASSA 2003 model are used as the benchmark to determine which estimates best represent the level and trend of mortality in South Africa. Year of death data from the NIDS survey contain a high non-response rate, so to correct for that problem, deaths with unknown year of death were distributed proportionately. Results and analysis show that distributing deaths with unknown year of death data proportionately according to those with known year of death results in time reference points being located more in the recent past than those produced by using the Brass and Bamgboye (1981) method. The analysis also shows that estimates of mortality produced from using the year of death data are not internally consistent and for some years underestimate mortality. Since the NIDS data contain data on survival of parent at ages 5 and 15 of respondent, the research uses the data to impute year of death to deaths with unknown year of death. Results show us that the estimates produced from using these data are internally consistent and stretch for a much wider span of years.

TABLE OF CONTENTS

PLAGIARISM DECLARATION	2
ACKNOWLEDGEMENTS	3
ABSTRACT.....	4
TABLE OF CONTENTS	5
LIST OF TABLES.....	7
LIST OF FIGURES.....	8
1 INTRODUCTION.....	10
1.1 Background.....	10
1.2 Objectives of the Research	11
1.3 Significance of this Research	11
1.4 Organisation of the dissertation.....	12
2 LITERATURE REVIEW.....	13
2.1 Introduction	13
2.2 The conventional orphanhood method	13
2.3 Limitations.....	16
2.4 Additional data: year of death of parent.....	21
2.5 Trends and levels of Adult Mortality in South Africa	23
3 DATA SOURCE AND METHODOLOGY	28
3.1 Introduction	28
3.2 Data source and quality	28
3.3 Theconventional orphanhood method	33
3.4 The conventional synthetic cohort method	36
3.5 Using year of death data	37
3.6 The ASSA 2003 model and how it was used.....	41
4 RESULTS	44
4.1 Introduction	44
4.2 Time location of estimates	44
4.3 Single survey estimates of maternal and paternal orphanhood	49
4.4 Synthetic Cohort Estimates of Maternal and Paternal orphanhood...54	

5	DISCUSSION AND CONCLUSION.....	57
5.1	Introduction.....	57
5.2	The problem of missing data.....	57
5.3	Estimates of time location and mortality.....	58
5.4	Synthetic cohort estimates of maternal and paternal orphanhood....	60
5.5	Limitations of the Study.....	61
5.6	Conclusions and Scope for the Future.....	62
	REFERENCE.....	63
	APPENDIX A: TABLES.....	65
	APPENDIX B: STATA CODE.....	68

University of Cape Town

LIST OF TABLES

Table 3.1	The percentages of respondents who reported year of father's death, by sex and population group.....	32
Table 3.2	The percentages of respondents who reported year of mother's death, by sex and population group.....	32
Table 3.3	Distribution of respondents who did not know father/mother year of death and whether father/mother was alive at age 5 and 15	33
Table 3.4	Antenatal HIV prevalence, South Africa	36
Table 3.5	Estimates of adult mortality in South Africa, 35q15.....	42
Table 4.1	Comparison of empirical and theoretical time reference points, paternal	45
Table 4.2	Comparison of empirical and theoretical time reference points, maternal	45
Table 4.3	Comparison of empirical and theoretical time reference points, empirical time location calculated from year of death data imputed by using data on survival of father at age 5 and 15	47
Table 4.4	Comparison of empirical and theoretical time reference points, empirical time location calculated from year of death data imputed by using data on survival of mother at age 5 and 15	47

LIST OF FIGURES

Figure 2.1	Trends in the probability of dying between ages 15 and 50, South Africa	24
Figure 2.2	Trend in mortality from the 1996 Census	26
Figure 3.1	Proportion of respondents not knowing the year of death of father/mother by age	30
Figure 3.2	Proportion of respondents who did not know father's year of death, by sex	31
Figure 3.3	Proportion of respondents who did not know mother's year of death by sex	32
Figure 3.4	Distribution of reported years of deaths	38
Figure 3.5	Distribution of reported years of death, after assuming that deaths with missing years of death occurred before 2003.25	40
Figure 3.6	Distribution of reported deaths before and after using orphanhood data at age 5 and 15 of the respondent	41
Figure 3.7	The probability of dying between ages 15 and 50, South Africa	42
Figure 4.1	Male mortality estimates derived from the 2008 NIDS survey plotted using reference points from the Brass & Bamgboye method and from year of death data (YOD)	46
Figure 4.2	Female mortality estimates derived from the 2008 NIDS survey plotted using reference points from the Brass & Bamgboye method and from year of death data (YOD)	46
Figure 4.3	Estimates of paternal mortality from the 2008 NIDS survey plotted using different reference points	48
Figure 4.4	Estimates of maternal mortality from the 2008 NIDS survey plotted using different reference points	48
Figure 4.5	Estimates of male mortality derived using the conventional orphanhood method and using year of death data, after distributing deaths with unknown YOD proportionately	50
Figure 4.6	Estimates of female mortality derived using the conventional orphanhood method and using year of death data, after distributing deaths with unknown YOD proportionately	50
Figure 4.7	Estimation of male mortality derived from orphanhood data and parents year of death, assuming all unknown years of death occurred 5 years prior to survey	52
Figure 4.8	Estimation of female mortality derived from orphanhood data and parents year of death, assuming all unknown years of death occurred 5 years prior to survey	52
Figure 4.9	Estimation of male mortality derived from orphanhood data and father's year of death, with unknown year of death distributed using data on survival of father at age 5/15 of the child	53

Figure 4.10	Estimation of female mortality derived from orphanhood data and mother's year of death, with unknown year of death distributed using data on survival of mother at age 5/15 of the child	54
Figure 4.11	Paternal synthetic cohort estimates produced using orphanhood data and year of death data	55
Figure 4.12	Maternal synthetic cohort estimates produced using orphanhood data and year of death data	56

University of Cape Town

1 INTRODUCTION

1.1 Background

The estimation of adult mortality still remains an area of great concern in Sub-Saharan Africa in the light of the HIV/AIDS pandemic. While a great amount of resources has been allocated by governments and non-governmental organisations to the estimation of levels and trends of child mortality, adult mortality still lags behind. The lack of reliable vital registration systems has made it difficult to obtain reliable measurements of adult mortality.

The lack of reliable vital registration systems has forced demographers to turn to indirect techniques to estimate adult mortality. Over the years a number of indirect methods have been developed, one such method is the orphanhood method. It is one of the oldest techniques of indirect estimation used in the measurement of adult mortality. It utilises the simple question “is your mother/father alive” to estimate levels and trends of adult mortality. The method has the advantage that the question asked is fairly simple and can be asked in a census or even in relatively small surveys. Despite the numerous advantages the method has and its success in estimating mortality in some countries it suffers from a number of limitations. This has led to a number of scholars developing and modifying the method so as to make it produce reliable estimates of mortality. Some scholars have advocated filter questions to improve the estimates. One such filter question is the question suggested by Chackiel and Orellana (1985) which asks year of death of the parent. Feeny (2001) highlights a possible lack of robustness of the dating procedure developed by Brass and Bamgboye (1981) in populations such as those in Southern Africa which have a rapidly rising level of mortality. This possible lack of robustness is attributed to the assumption made by the Brass and Bamgboye method that mortality is declining linearly yet in actual fact in Southern Africa mortality was increasing at a rapid rate in the late 1990s. Chackiel and Orellana (1985) suggest that year of death data can be used to calculate better estimates of time location than the ‘theoretical’ methods such as the one developed by Brass and Bamgboye (1981). Chackiel and Orellana(1985) also argue that year of death data produces estimates of mortality which are more consistent than those produced by the conventional orphanhood methods such as the method developed by Brass and Hill (1973).

The use of year of death data in estimating adult mortality has been successfully applied in South America, particularly in Bolivia, Costa Rica and Honduras but has failed in an application in East Africa because of poor quality data (Chackiel and Orellana 1985, Timæus 1991a). Timæus (1991a) suggests that the approach needs to be studied more in populations where knowledge of dates is not widespread, such as those in Sub-Saharan countries. The method still remains relatively untested in Sub-Saharan Africa especially in populations experiencing an HIV/AIDS epidemic. This research has set out to investigate whether using the year of death data from the National Income Dynamics Study (NIDS) survey produces better estimates of time location than the Brass and Bamgboye method and whether year of death data produces better estimates of mortality than the orphanhood method without the aid of filter questions (conventional orphanhood method).

1.2 Objectives of the Research

The first objective of this research is to investigate whether using year of death data produces better estimates of time location than the method developed by Brass and Bamgboye (1981). The second objective of this research is to investigate whether the year of death data produces better estimates of mortality in terms of level and trend than using the conventional orphanhood methods (i.e. the orphanhood method plus the synthetic cohort method).

1.3 Significance of this Research

The importance of this research stems from the need to estimate accurately levels and trends of adult mortality. Estimates of mortality are important in policy making especially in this era of HIV/AIDS. In the absence of reliable vital registration systems the onus is on demographers to develop methods that can estimate mortality levels accurately.

The question on survival of fathers and mothers is used in estimating the number of orphans. This means that the question is likely to be included in many surveys and censuses; therefore there is a need to develop and improve the orphanhood method so as to produce reliable estimates of mortality.

Chackiel and Orellana (1985) suggest that using the year of death data can enable one to obtain estimates of mortality which stretch for more than 20 years before the survey.

This could be useful in cases where we want to document the history of mortality in a country which has not had a chance of conducting many surveys because of economic or political reasons.

1.4 Organisation of the dissertation

This research is made up of five chapters. Chapter 2 reviews the relevant literature which is related to the research. Chapter 3 introduces the data source and gives an outline of methods that were used to produce results. Chapter 4 presents the results obtained after carrying out the analysis. Chapter 5 reflects on the results that were obtained and highlight possible areas of future research.

University of Cape Town

2.1 Introduction

The purpose of this chapter is to review work that has been done on the orphanhood method and the various attempts that have been made to improve the method. The first part of the literature review, section 2.2, looks at the idea behind the conventional orphanhood method (the conventional orphanhood method in this research means the orphanhood method without any additional questions to assist in improving estimates of mortality or their time location) and looks at the subsequent development of the method to its current form. The second part of this review, section 2.3, covers the limitations of the method and the various attempts that have been made to address these problems. The third part of this chapter, section 2.4, covers the work done by Chackiel and Orellana on using year of death data to produce estimates of adult mortality. The final part, section 2.5, looks at the levels and trends of adult mortality in South Africa.

2.2 The conventional orphanhood method

The use of the survival status of fathers/mothers to estimate mortality can be traced as far back as 1939 when Lotka proposed using life table functions to estimate the number of orphans (Luy 2009). A scholar named Henry, in 1960, was the first to suggest using the number of orphaned children to estimate adult mortality in cases where fertility and mortality schedules are known (Luy 2009). To estimate adult survivorship using the survival of fathers/mothers, the proportion of children with father/mother alive is needed for each age group of the children. The idea underlying the method is that since mothers must be alive at the birth of their children, the child's age at the time of the survey, x , represents the period to which the mother is exposed to the risk of mortality. If the age of the mothers at the birth of these children is y , and assuming that the mortality of mothers and children is independent, then the proportion of the mothers of these children aged x is a measure of survivorship of the mothers between ages x and $x+y$, that is l_{x+y}/l_y . The approach here is to use the proportion of children in a given age group with mother alive to calculate l_{x+y}/l_y , where y represents a suitable age near the mean age of child bearing, M , and x represents a suitable age near the midpoint of each age group. Conditional survivorship probabilities for a convenient age and duration are then calculated by using either a weighted average of the survival ratios for the

adjacent age groups or by applying regression equations using these ratios as independent variables. These probabilities are then converted into estimates of mortality by using a suitable life table as standard. To make the translation the logit relation developed by Brass in 1971 is used namely: $\lambda(l(x)) = \alpha + \beta * \lambda(l^s(x))$, where

$$\lambda(l(x)) = 0.5 \left(\frac{1-l(x)}{l(x)} \right) \text{ and } l(x) \text{ and } l^s(x) \text{ equal to the proportion surviving to age } x$$

in the fitted and standard life table, respectively, while alpha, α , the level of mortality is calculated for each of the conditional survival probabilities, given a suitable value for β , which represents the shape of mortality (Bhat 2004).

Brass and Hill (1973) developed a simple and convenient approach to estimate mortality from orphanhood data. Assuming a stable population the number of women aged y , $N(y,t) = Be^{-rt}l(y)$ where B is the number of births at time t , and given constant age specific fertility rates (ASFRs), $f(y)$, the number of births to women aged $y = Be^{-rt}l(y)f(y)$. Thus the proportion of children aged x with living mothers can be

$$\text{expressed as } S(x) = \frac{\sum_y e^{-rt} l_y f(y) \frac{l_{y+x}}{l_y}}{\sum_y e^{-rt} l_y f(y)} \text{ or in five year age groups, can be expressed as}$$

$${}_5S_a = \frac{\int_a^{a+5} e^{-rx} l(x) \int_s^w e^{-rx} l(y+x) f(y) dy dx}{\int_a^{a+5} e^{-rx} l(x) \int_s^w e^{-rx} l(y) f(y) dy dx},$$

where w and s represents the child-bearing age range and r is the rate of natural increase (Brass and Hill 1973, Timæus 1992).

Brass and Hill developed a set of weighting factors $w(n)$ to convert the proportions of children with living mothers, ${}_5S_a$, into estimates of female survival. The weighting factors were introduced to account for the differences in age patterns of fertility and mortality (United Nations 1983). This conversional process involves a weighted average which relates the probabilities of survival to the survivorship probabilities

$$l(25+n)/l(25) = w(n) {}_5S_a + (1-w(n)) {}_5S_a.$$

For estimating paternal mortality the procedure is basically the same as with maternal mortality except that there is the need to account for the fact that males are exposed to the additional risk of dying between the time of conception and the time of birth, which is roughly 9 months or 0.75 of a

year. Brass and Hill (1973) did not take this into account in deriving their weighting factors, which means that the estimates from the weighted average approach are biased since the period of exposure to the risk of dying for fathers refers is measured from the time of birth of the child instead of the time at conception (Timæus 1992).

Timæus(1992)estimated new coefficients replacing the ones suggested by Hill and Trussell (1977). Estimates produced using coefficients estimated by Timæus (1992) are similar to estimates of maternal orphanhood derived using Hill and Trussell (1977) regression coefficients but produce better estimates than those produced using the Brass and Hill approach. The idea behind the regression based approach is to introduce an intercept term, which the Brass and Hill method lacks. This makes it possible to model the relationship between life table measures of survivorship and proportions with living fathers which in some situations the Brass and Hill method is not be able to model. The equivalent expression to the above equation for ${}_5S_a$, taking into consideration that the exposure to risk for males is 0.75 years longer is given by:

$${}_5S_a = \frac{\int_a^{a+5} e^{-rx} l(x) \int_s^w e^{-rx} l(y+x) / l(y-0.75) f(y) dy dx}{\int_a^{a+5} e^{-rx} l(x) \int_s^w e^{-rx} l(y) f(y) dy dx}$$
 and the function which relates

the probabilities of survival to the proportions surviving is

$${}_n p_{35} = \beta_0(n) + \beta_1(n)M + \beta_2(n) {}_n S_{n-5} + \beta_3(n) {}_n S_5$$
 for male mortality. For female mortality the equations is given by ${}_n p_{35} = \beta_0(n) + \beta_1(n)M + \beta_2(n) {}_n S_{n-5}$ where $\beta_0, \beta_1, \beta_2$ and β_3 are the new regression coefficients.

The method is been generally confined to estimating maternal mortality. This is because subjects are assumed to know more about the survival status of their mothers than that of their fathers. In African settings, whether due to migrant labour or for other reasons, fathers were often absent from households. This has rendered paternal estimates less reliable than maternal mortality, generally leading it to overestimate the mortality of fathers. However with the development of society, science and methods of data collection more is now known about the survival status of fathers than, say, 20 years ago. Although we still face similar problems as highlighted above the method is now applied to estimate both maternal and paternal mortality.

Initial trials of the orphanhood method in three African countries (Chad, Malawi and Kenya) by Blacker (1977) revealed it to be a simple and convenient way of obtaining the level of adult mortality in countries without a reliable vital registration system. He points out that the method performed reasonably well judging by the internal consistency of the estimates and from the fact that the estimates were consistent with estimates produced by other methods. The method was also applied successfully in four countries in Latin America (Timæus and Graham 1989). Despite the initial success results, the method was criticised because of limitations which emanate from the method itself and from its assumptions.

2.3 Limitations

2.3.1 The problem of time location

If mortality is decreasing, which has been the case in many developing countries just before the era of AIDS, orphanhood estimates would refer to conditions in the distant past and not in the recent past. To account for this Brass and Bamgboye (1981) developed a formula for estimating the time location to which the survival probabilities refer. The procedure assumes that mortality has been declining linearly over time and that the shape of adult mortality can be represented by a standard life table. Using this assumption Brass and Bamgboye developed two simple equations to estimate the time reference points to which mortality estimates apply. By considering children aged N years with mother alive at time of survey and assuming that the life table function $l(x)$ is linear over age they considered two cases to which their equations apply. The first case is the situation where mortality is heavy, where they estimate $T = \frac{1}{2}N$, that is, if deaths are uniformly distributed in the age span of N years, the time reference, T years prior to the survey, is located at the midpoint of the respondents age span. The second case is the situation where mortality is lighter, here they estimate $T = \frac{1}{2}N(1 - C(N))$,

where $C(N) = \frac{\ln(S(N))}{3} + f(N + M) + 0.0037(27 - M)$, $S(N)$ is the proportion of respondents aged N with mother alive, M is the mean age of childbearing at the time the respondents were born, and $f(N+M)$ is a standard function of age, whose value can be interpolated from the values presented in Table A.1 of the appendix. The $C(N)$ is introduced to adjust for the fact that if the survivorship function $l(x)$ is non-linear at

older ages, deaths will not be uniformly distributed, they will be centred more towards the survey date than the date implied by $\frac{N}{2}$ (Brass and Bamgboye 1981).

2.3.2 Estimates out of date

An inherent limitation of the original method is that the estimates produced by this method refer to dates way before the survey date. Therefore to obtain an estimate of mortality for the recent past one is forced to make an assumption about the trend of the estimates, probably that mortality has remained constant. In an attempt to resolve this problem Zlotnik and Hill (1981) developed a method for estimating survivorship probabilities for intersurvey periods. The method requires data sets from consecutive surveys or censuses separated by five years or a multiple of five years.

Assuming that migration is not significant to the changes in proportion during the intersurvey period, the method involves “chaining” together cohorts from the first and second survey to obtain a hypothetical intersurvey cohort of respondents. Let $S(n,1)$ and $S(n,2)$ be the proportions with mother/father alive for respondents aged between n and $n+5$ for the first and second surveys respectively. Then the proportion with mother/father alive for the hypothetical cohort can be expressed as follows:

$$S(n,s) = S(n,2), \text{ for } n < T, \text{ and } S(n,s) = S(n-T,s) \frac{S(n,2)}{S(n-T,1)} \text{ for } n \geq T, \text{ where } T$$

denotes number of years between the two surveys (United Nations 1983).

Timæus (1986) developed a much more flexible approach than that developed by Zlotnik and Hill. This approach uses the growth rate of proportions with mother/father alive over the intersurvey period instead of the actual cohort survival proportions.

Although the method produces almost the same results, the approach by Timæus is more flexible in that it can be applied to situations where the data sets are not separated by a multiple of five years (Timæus 1986). He assumes that the number of persons aged x to $x+5$ in the first census/survey is $p_1(x,5)$ and the second census is $p_2(x,5)$ and estimates the intercensal growth rate as $r(x,5) = \ln[p_2(x,5)/p_1(x,5)]/t$, where t is the intercensal period in years. The average number of person-years lived per annum

between the two surveys is given by $N(x,5) = \frac{p_2(x,5) - p_1(x,5)}{tr(x,5)}$ and synthetic survival

ratios are estimated by $R(x,5) = \frac{N(x+5,5) \exp\{2.5r(x+5,5)\}}{N(x,5) \exp\{-2.5r(x,5)\}}$. Using this approach

and given proportions of children with mother/father alive in the first and second census, $S_1(x,5)$ and $S_2(x,5)$ respectively, synthetic ratios are given by

$$R(x,5) = \frac{S(x+5,5) \exp\{2.5r(x+5,5)\}}{S(x,5) \exp\{-2.5r(x,5)\}}$$
 and the proportion with surviving parents for

the hypothetical cohort by $S^*(x,5) = R(x-5,5) * S^*(x-5,5)$, where

$$S^*(5,5) = \frac{S_1(5,5) + S_2(5,5)}{2}$$
. Once $S^*(x,5)$ is obtained, the method of obtaining the

conditional survivorship is the same as that in the single survey/census enquiry.

The advantage of using this kind of variation of the orphanhood method is that it provides the analyst with up-to-date estimates of mortality than the variation developed by Brass and Hill (1973). However the major weakness with this method is the bias introduced if differential reporting exists. Timæus and Graham (1989) point out that in situations where reporting improves between enquiries the method is bound to overestimate mortality.

2.3.3 The adoption effect

One of the most serious limitation of the orphanhood method is the so called “adoption effect”. The adoption effect is a term given to the situation where a respondent answers “yes” to the question “is your mother/father alive” when, in actual fact, the biological mother/father is not alive. In African culture it is quite common for an orphan to call a guardian parent, mother or father when in actual fact they are not the biological parent. Udjo (2007) provides us with evidence of this by quoting the former president of South Africa Nelson Mandela in 1994 as follows: “In African culture we do not make the same distinctions among relatives practiced by whites. We have no-half brothers or a half sister, my mother’s sister is my mother; my uncle’s son is my brother; my brother’s child is my son, my daughter...” This misrepresentation of information has been identified as a cause of distortions of mortality estimates in African populations and is more pronounced in the responses of the younger age groups, that is, those aged 15 years and below (Blacker 1977, Timæus and Graham 1989). Palloni, Massagli and Marcotte (1984) in their study reveal that the effect of the adoption effect is that it underestimates the level of mortality and hence exaggerates the decline in mortality. The

exaggerated decline results from the fact that the estimates derived from data from the younger age groups greatly underestimate mortality relative to the other age groups, which then results in an exaggerated decline in mortality, since the mortality levels portrayed by the younger age groups would be much lower than the other groups.

Timæus (1991b) developed a method of estimating adult mortality using data on orphanhood in adulthood. His idea is to use data of those children who had living mothers at exact age 20 so as to minimise the bias introduced by the adoption effect. He combines this with the idea of Zlotnik and Hill (1981) of using two data sets to create synthetic cohorts so as to obtain up-to-date estimates of mortality. Using the same ideas as in conventional method the proportion of children aged 20 and above with a living mother or father at exact age 20 can be represented by the following equation

$$\frac{{}_5S_x}{S(20)} = \frac{\int_a^{a+5} e^{-rx}l(x)S(x)dx}{\int_a^{a+5} e^{-rx}l(x)dxS(20)}$$

. Timæus introduced new estimation equations similar to

those developed by Hill and Trussell (1977) to adjust for the fact that exposure starts 20 years after the mean age at child bearing. For mothers, survivorship was estimated from a base age of 45, namely: $\frac{l(25+n)}{l(45)} = \beta_0(n) + \beta_1(n)\bar{M} + \beta_2(n)\frac{{}_5S_{n-5}}{S(20,t)}$ and for fathers,

survivorship was based on age 55, i.e.,

$$\frac{l(35+n)}{l(55)} = \beta_0(n) + \beta_1(n)\bar{M} + \beta_2(n)\frac{{}_5S_{n-5}}{S(20,t)} + \beta_3(n)\frac{{}_5S_{n-5}}{S(20,t)}$$

. Although this new

approach produced encouraging results, Timæus points out that the method is susceptible to age reporting errors.

2.3.4 Problems introduced by HIV/AIDS

Timæus and Nunn (1997) using data from the Demographic Surveillance site in Masaka, Uganda investigated whether the orphanhood method can be successfully applied to a population experiencing an HIV/AIDS epidemic. They identified two sources of bias that might distort estimates that are produced using the orphanhood method. The first source of bias is selection bias resulting from mother to child transmission and through reduction in fertility of women infected by the virus. They developed a mathematical model linking the survival of mothers and mortality in a population experiencing an HIV/AIDS epidemic. They applied this model to the data from rural Uganda so as to

quantify the extent of the bias. The major conclusion is that selection bias resulting from mother to child transmissions and the reduction in fertility is likely to be a small. By making use of a few simplifying assumptions they developed a formula to adjust for selection bias. Given the that the proportion of women alive who would have given birth a years ago in the absence of the impact on fertility of HIV is $S(a)$ and given the proportion of respondents who report that mother is alive at the time of the survey is $S^*(a)$, and assuming that the seroprevalence P of women in antenatal clinics is representative of the population of women infected, they showed that

$$\frac{S(a)}{S^*(a)} = \frac{1 - hP}{1 + \frac{1 - F}{F}P}$$

where h represents the vertical transmission and F is the reduction

in fertility of HIV positive women. By assuming that the fertility of HIV positive mothers is 20 per cent lower than that of HIV negative mothers and that the rate of mother to child transmission is 25 per cent, they showed that the above equation produces the following approximation which corrects for selection bias:

$${}_5S'_x = \frac{1 - 0.25P}{1 + 0.25P} {}_5S_x^* \approx (1 - 0.5P) {}_5S_x^* .$$

To allow for the possibility that some

seropositive mothers survive for more than five years after the birth of their children the adjustment is halved for respondents who are in the 5-9 age group such that

$${}_5S'_x = (1 - 0.25P) {}_5S_x^* .$$

The second source of bias that Timæus and Nunn identified is the bias introduced by using regression coefficients (used in converting proportions with living mother to life table survivorship) estimated by Timæus (1992) in populations experiencing an HIV/AIDS epidemic. Regression coefficients from Timæus (1992) were estimated at a time when the prevalence of HIV was very low and in some populations the virus was non-existent. They identified this as a potential source of bias if these coefficients were to be used in a population which is experiencing an HIV/AIDS epidemic. Using the data from Masaka, Uganda they showed that regression coefficients from Timæus (1992) perform fairly well for children over the age of 15 and that HIV only affects estimates that come from children who are in the age groups of 5-9 and 10-14. From analysing the data they observed that there existed a relationship between survival of mothers and life table survivorship from age 25 to age 40. Using this relationship they estimated a revised set of regression coefficients for populations that are experiencing

an epidemic. However, as they clearly point out, this new set of coefficients is only provisional since the coefficients were derived from rural Uganda only.

Timæus and Jasseh (2004) used Demographic and Health Surveys (DHS) to estimate levels and trends of adult mortality from 23 sub-Saharan countries. To estimate these levels and trends in adult mortality they made use of the sibling and orphanhood data which are contained in most of DHS surveys. In using the orphanhood method they allowed for the impact of HIV/AIDS by reducing the estimated prevalence of women at the birth of their children by 47.5 per cent such that ${}_5S_x' \approx (1 - 0.475P) {}_5S_x^*$ and for children in the age group of 5-9 such that ${}_5S_x' \approx (1 - 0.247P) {}_5S_x^*$. Timæus and Jasseh (2004) considered the case of fathers, where they allowed for the impact of HIV by reducing the estimated prevalence by 60 per cent of the children with living father. They adjusted using 60 per cent so as to allow for concordance of infection in couples.

2.3.5 Other problems associated with the method

Age misreporting in developing countries is a serious problem, the main reason for this being the lack of education particularly in the case of the elderly and the lack of complete birth registration system which makes it difficult for the respondent to know the exact age of people they are reporting on. However, this problem of misstatement of age is more common in the elderly and not the younger ages and thus has little effect on orphanhood estimates since the orphanhood method does not make use of data from age groups above the 50-54 age group (Timæus and Graham 1989).

The survival of parents with more than one living child might be reported more than once which results in an increase in the proportion of children with living parents and similarly childless couples will not be represented in the survey, which could lead to a bias in the estimates if they experience different mortality from those with children. However, Palloni, Massagli and Marcotte (1984) showed that the errors which result are not serious because more often than not the effects tend to cancel out each other.

2.4 Additional data: year of death of parent

Chackiel and Orellana (1985) proposed the use of year of death data to improve estimates produced by orphanhood method. They proposed three ways in which the year of death data could be used to improve estimates of mortality. Their first

suggestion was with regards to the time location of estimates, they proposed using the mean year of death of the empirical data as an approximation for the average year of death for the age groups. Assuming that the data collected from a survey is of good quality they argued that it was more accurate to use the empirical data to estimate time reference points than the approach suggested by Brass and Bamgboye (1981) (referred to by them as ‘theoretical approach’). By using the survey data taken from the National Demographic Survey of Honduras 1983-1984 (EDENH II) they showed that as the age groups of the children get older the theoretical estimates of time reference points are biased more towards the recent past than time reference points estimated using year of death data. They also showed that time calculated from year of death data spans a greater number of years than theoretical reference points.

The second innovation they proposed was to use the year of death data to create synthetic surveys 5 and 10 years before the survey. They showed through simple calculations how proportions with mother alive at 5 and 10 years before the survey can be used to produce mortality estimates for much longer periods than produced by the conventional orphanhood method. Given that the time of the survey is t , the number of children in age group i with mother alive at the time of the survey is A_i , the number of children in age group i with mother dead is D_i and the total number of children in age group i at the time of the survey is N_i , the proportion with mother alive at time of survey is A_i/N_i and the proportion with mother alive 5 and 10 years before the survey is $(A_i + D_i(5))/N_i$ and $(A_i + D_i(10))/N_i$ respectively, where $D_i(5)$ and $D_i(10)$ represent the total number of mothers who died not more than 5 and 10 years before the survey (Chackiel and Orellana 1985). Now, given the survey at time t and the synthetic surveys at time $t - 5$ and $t - 10$ they argued that it is possible to use the proportions at these three points in time to calculate mortality. The advantage of this kind of approach is that it gives consistent estimates that stretch far back in time, as they illustrated with the data from EDENH II survey.

Chackiel and Orellana also showed that the proportions with mother alive at 5 and 10 years before the survey can be used in the same manner as was proposed by Zlotnik and Hill (1981), that is, using two censuses or surveys separated by 5 years to estimate

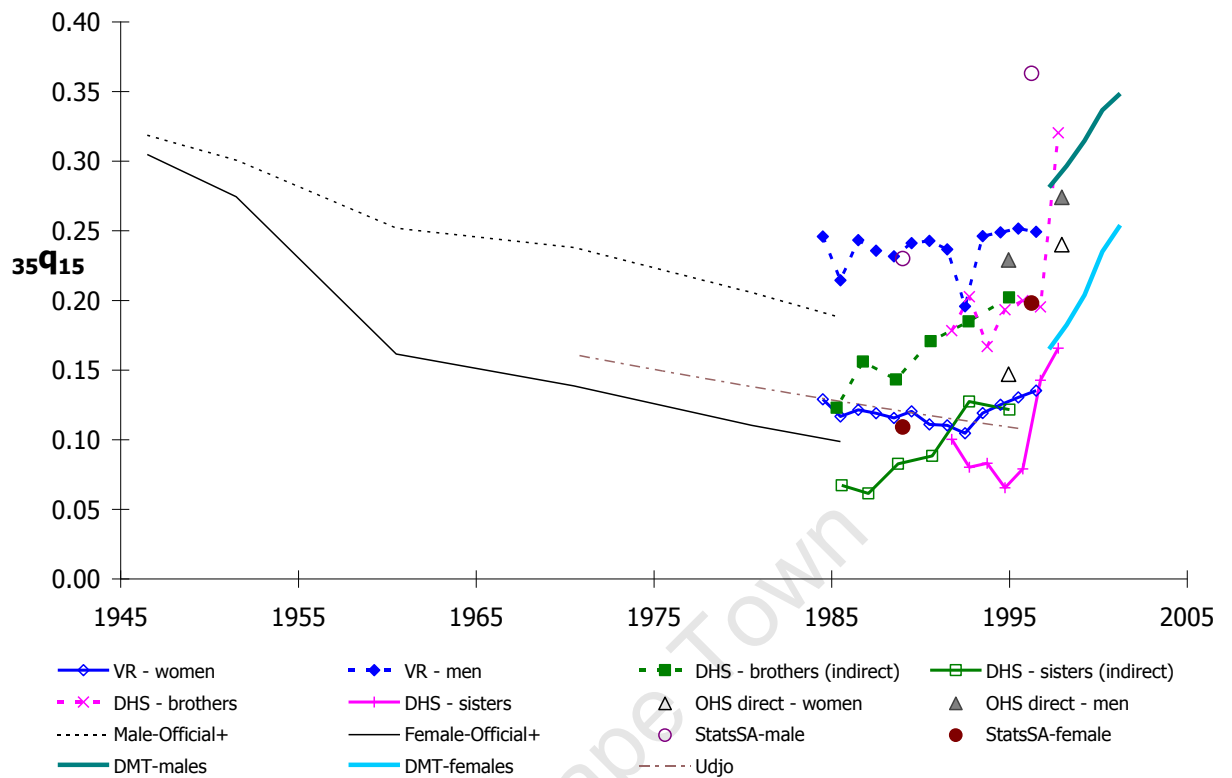
mortality for the intercensal/intersurvey period. They argue that it is advantageous to use year of death data because the Zlotnik and Hill (1981) approach suffers from differential reporting and sampling errors and also relies on the assumption that migration is not significant for the period in question (Chackiel and Orellana 1985, Timæus 1986).

2.5 Trends and levels of Adult Mortality in South Africa

Mortality estimates are a vital piece of information in effective policy making. It is particularly important in the era of HIV/AIDS for a government to know levels and trends of mortality so as to be able to plan adequately and allocate resources. Mortality has been difficult to estimate in South Africa especially before 1994. This is because of the poor vital registration system under apartheid which force demographers to rely on scanty information and data (Dorrington, Moultrie and Timæus 2004). After the collapse of the apartheid system there were efforts from the new government to acquire data through surveys and censuses. Statistics South Africa (Stats SA) launched the October Household Survey (OHS) in 1994, an annual survey which collects data on various social and economic aspects of the population, and carried out the 1996 and 2001 censuses and the 2007 Community Survey (CS). Trends in adult mortality in South Africa can be separated into three periods. The first period is the period before 1980 where mortality was falling, the second is the period where mortality remained roughly constant between 1980 and 1995, and the third is the period where mortality was increasing due to the HIV/AIDS epidemic (Dorrington, Moultrie and Timæus 2004, Bradshaw and Timæus 2006).

Dorrington, Moultrie and Timæus (2004) provide an analysis of adult mortality trends in South Africa according to various sources of mortality data and estimates obtained before and after the collapse of the apartheid system. As illustrated by Figure 2.1 the probability of a 15 year old dying by age 50, ${}_{35}q_{15}$, according to the official South African life tables and Sadie (Official +), decreased sharply for women between 1945 and 1960 from an average of 30 per cent to almost 15 per cent and then further decreased gradually to about 10 per cent around 1985.

Figure 2.1 Trends in the probability of dying between ages 15 and 50, South Africa



Source: Figure 6.2 (Dorrington, Moultrie and Timæus 2004)

The estimates of female mortality from South African Life Tables (SALTs) and Sadie are in agreement with those produced by Udjo (for female mortality) for the period 1960 to 1985 in terms of rate of decline but differ in terms of level. Although estimates from Udjo differ in terms of level from those derived from the Official life tables, they are consistent with those derived from the vital registration (VR) system after adjusting for undercount for the period 1985-95. On the other hand female estimates produced from Demographic Health Survey (DHS) data are at odds with those produced by Udjo and those derived from the vital registration system. The United Nations Population Division (UNPD) (2009) estimates ${}_{35}q_{15}$ to have been 12 per cent for females between 1985 and 1990, this gives credibility to estimates produced by Udjo, estimates derived from the vital registration system and those produced by Statistics South Africa (Stats SA) and this implies that estimates from DHS data are less reliable. Evidence from Figure 2.1 indicate that around 1995 was the period in which female mortality began to increase at a rapid rate although in some provinces in South Africa which were badly affected by the HIV/AIDS epidemic, sharp increases in deaths were recorded as early as 1993 (Tollman, Kahn, Garenne *et al.* 1999).

Estimates of female mortality produced by Dorrington, Moultrie and Timæus (2004) between 1995 and 2005 reveal that ${}_{35}q_{15}$ increased from about 17 per cent around 1997 to about 26 per cent in 2003, while the UNPD also reports this rapid upward increase in female mortality with ${}_{35}q_{15}$ estimated to be about 11 per cent for the period 1990-95, increasing to an average of 18 per cent for the period 1995-2000 and further increasing to an average of 34 per cent for the period 2000-05. The rapid increase in female mortality in the late 1990s is attributed to the HIV/AIDS epidemic. Bradshaw and Nannan (2006) for instance report that in the year 2000 alone 51 per cent of female deaths are attributable to HIV/AIDS.

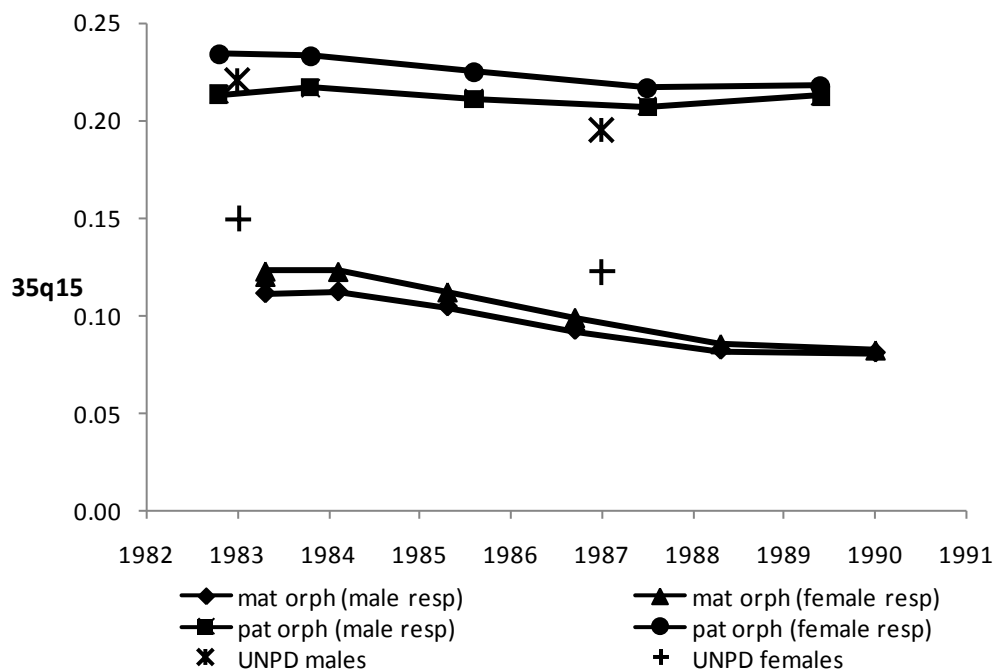
On the other hand, male mortality is estimated to have been higher than female mortality with the decline of male mortality generally following that of female mortality albeit less steeply. The probability of a 15 year old male dying by age 50 is estimated to have been around 32 per cent in 1946 decreasing steadily to about 25 per cent around 1960 and slowly decreasing to around 20 per cent in 1985. The trend in male mortality also follows the same trajectory as that of female mortality between 1985 and 2005. As is illustrated in Figure 2.1, the probability of a male aged 15 dying before age 50 was roughly constant between 1985 and 1995. Estimates derived from the vital registration system and estimates produced by Statistics South Africa indicate that ${}_{35}q_{15}$ was hovering around 25 per cent which is consistent with the UNPD (2009) which estimates ${}_{35}q_{15}$ to be on average 20 per cent for the period 1985-90 and 19 per cent for the period 1990-95. As was the case with female mortality, estimates indicate that male mortality began to increase around the mid 1990s. Dorrington, Moultrie and Timæus (2004) estimate ${}_{35}q_{15}$ to have increased from about 27 per cent around 1999 to about 35 per cent around the year 2000. This is consistent with UNPD (2009) which reports that ${}_{35}q_{15}$ increased from 18 per cent for the period 1995-2000 and to 34 per cent for 2000-05.

2.5.1 Trends and levels of mortality produced using the orphanhood method

Udjo (2005) analysed orphanhood data from the 1995, 1998 OHS and the 1996 census to estimate mortality levels and trends for adult males and females for the period 1980 - 90. He produced a set of alpha values which can be translated into estimates of ${}_{35}q_{15}$.

Figure 2.2 below illustrates mortality estimates of ${}_{35}q_{15}$ translated from the alpha values produced by Udjo (2005) using the Brass General Standard life table with beta equal to 1 for the 1996 census. As can be observed in Figure 2.2 female mortality estimates are unrealistically low especially for the more recent years. For instance, female mortality is estimated to have been on average 8 per cent in 1987 while the UNPD estimates it to be around 12 per cent for the 1985-90 period. Udjo attributes these low estimates of female mortality to the adoption effect (Udjo 2005). Estimates of male mortality on the other hand are more consistent in terms of level and trend with estimates produced by UNPD (2009) and compare well with estimates produced by Dorrington, Moultrie and Timæus(2004) and Bradshaw and Timæus (2006).

Figure 2.2 Trend in mortality from the 1996 Census



Dorrington, Moultrie and Timæus (2004) applied the synthetic cohort method of the orphanhood method to the 1996 and 2001 census data. To minimise the impact of the adoption effect they used the variation of the synthetic cohort method developed by Timæus (1991b) which only considers respondents that are aged 20 years and above. Results obtained from this method were considered to be satisfactory.

The orphanhood method was also applied by Moultrie and Dorrington (2009) to derive the level and trend of adult mortality in South Africa for the past 15 years using data from the 2007 community survey (CS) and the National Income Dynamics Study (NIDS) survey. Estimates produced from both the CS and NIDS survey are consistent

in terms of trend with those produced by Dorrington, Moultrie and Timæus (2004), Bradshaw and Timæus (2006) and those produced by the UNPD (2009). Female estimates derived from both the CS and NIDS survey are consistent in terms of trend but when it comes to level, estimates produced using the NIDS data are erratic. The probability of a 15 year old female dying by age 50, ${}_{35}q_{15}$, is estimated to be around 20 per cent before 1995 and to have increased to 30 per cent by 2004, which is consistent with estimates reported by the UNPD. The indication from male estimates produced by Moultrie and Dorrington (2009) is that mortality was generally level before 1995 and increased rapidly in the late 1990s and early 2000s before levelling off around 2002. Estimates of male mortality derived from the Community Survey are at odds in terms of level with those produced by the UNPD for the period between 1990 and 1995. Estimates from the CS indicate that ${}_{35}q_{15}$ was on average 30 per cent for the period 1990-95 while UNPD estimates it to have been on average 19 per cent for the same period. After 1995 estimates derived from both the CS and NIDS are more consistent in both level and trend with those produced by the UNPD.

3.1 Introduction

The purpose of this chapter is to introduce the data sources and to outline the various assumptions and methods used to estimate adult mortality from the data. Section 3.2 introduces the data and looks at the quality of the data while sections 3.3 and 3.4 outline how the conventional and the synthetic orphanhood method were applied and the various assumptions and adjustments that were made so as to apply these methods to the South African data sets, including adjusting for the impact of HIV/AIDS. Section 3.5 covers how the year of death data were used to estimate mortality. The final section, section 3.6, describes briefly the ASSA model and how the estimates from the model were used.

3.2 Data source and quality

In order to estimate the levels and trends of adult mortality in South Africa data on the survival of fathers and mothers from the 1996 census, 2001 census, 2007 Community Survey and 2008 National Income Dynamics Study (NIDS) were extracted.

3.2.1 The 1996 and 2001 Censuses

The 1996 and 2001 census are the first and second post apartheid censuses to be held in South Africa. The reference date for the two censuses is the midnight 9/10 October 1996 and 9/10 October 2001 respectively. Both censuses were designed to collect data on various demographic and socio-economic issues of interest. Data from the two censuses for this research are in the form of a 10 per cent sample (which is all people included from a 10 per cent systematic sample of all households plus an independent 10 per cent systematic sample of all people living in special institutions and hostels). It includes a weight variable to adjust for the undercount in the censuses and for the fact that the censuses are in the form of a 10 per cent sample. The 1996 census estimated the population to be around 40.5 million people while the 2001 census estimated the population to be around 44.7 million people, after adjusting for under count as estimated by a post enumeration survey. In both censuses respondents were asked whether their biological father and mother were alive. In the 1996 census 95 per cent answered the question on survival of father while 97 per cent answered the question on survival of mother. Unfortunately Statistics South Africa edited the 2001 census data

such that all missing data on the question on survival of father and mother were given a valid imputed response.

3.2.2 The 2007 Community Survey

In 2007 Statistics South Africa undertook a large scale survey which they called the Community Survey. The objective of carrying out the survey was for the survey to act as a bridge between successive censuses which it was decided would only be held every 10 years. The survey was run from 7-28 February 2007 but for our purposes we assume that the survey occurred on the night of 14-15 February 2007. The survey covered 246,618 households from all provinces and enumerated 949,100 persons. The Community Survey estimated the population to be around 47.9 million people which represents an increase of about 7.2 per cent from the population in the 2001 census. The question on orphanhood was asked in the survey and about 56 per cent of the respondents answered that father was alive while 43 per cent answered that father was not alive representing a non-response of around 2 per cent. For the case of mothers, about 73 per cent of the respondents answered that mother was alive at the time of the survey while 26 per cent answered that mother was not alive, leaving a 1% non-response.

3.2.3 The 2008 National Income Dynamics Study

In a bid to gain an understanding of the social and economic dynamics which govern a typical South African household, the South African government in 2008 funded a survey called the National Income Dynamics Study (NIDS). The survey was undertaken by the Southern African Labour and Development Research Unit (SALDRU) which in 2009 released the results of the first wave. The study is the first national household panel study in South Africa and it aims to extract information from randomly selected households (even if the individuals in the households migrate within the country) by administering carefully drafted questionnaires. The sample of households was extracted using a stratified two stage cluster sampling technique from the master sample used by Stats SA for its Labour Force Survey and General Household Survey.

The first wave of the survey was conducted between February and November 2008, however, about 90 per cent of the respondents were interviewed before the end of June, which is about half way through the year. For the purpose of this research we assume

that the interviews took place halfway (in proportion to the number of interviews in each month) between February 2008 and June 2008, that is, that the survey was conducted on the night of 31 March/1 April 2008. A household questionnaire was administered in each household together with individual questionnaires which were administered separately to all children and all adults in the household. All respondents aged below the age of 14 at the date of the survey were categorised as children and all respondents aged 15 and above at the date of the survey were categorised as adults. A question on the survival of mother/father was asked in the survey and for those who answered that a parent had died the year of death of the parent was asked. In addition respondents were asked about the parent's survival at age 5 and 15 of the child.

Compared to the 1996 and 2001 censuses and the 2007 Community Survey the question on survival of parents in the NIDS survey was not well answered. About 48.4 per cent of the respondents answered that their father was alive while 36.8 per cent answered that their father was not alive, representing a non-response of about 15 per cent. The question on survival of the mother had a slightly better response with 63.7 per cent saying that their mother was alive, while 22.9 per cent answered that their mother was not alive. However, of the respondents who did not have a living father/mother, 36 per cent did not know the father's year of death while 32 per cent did not know the mother's year of death.

Figure 3.1 Proportion of respondents not knowing the year of death of father/mother by age

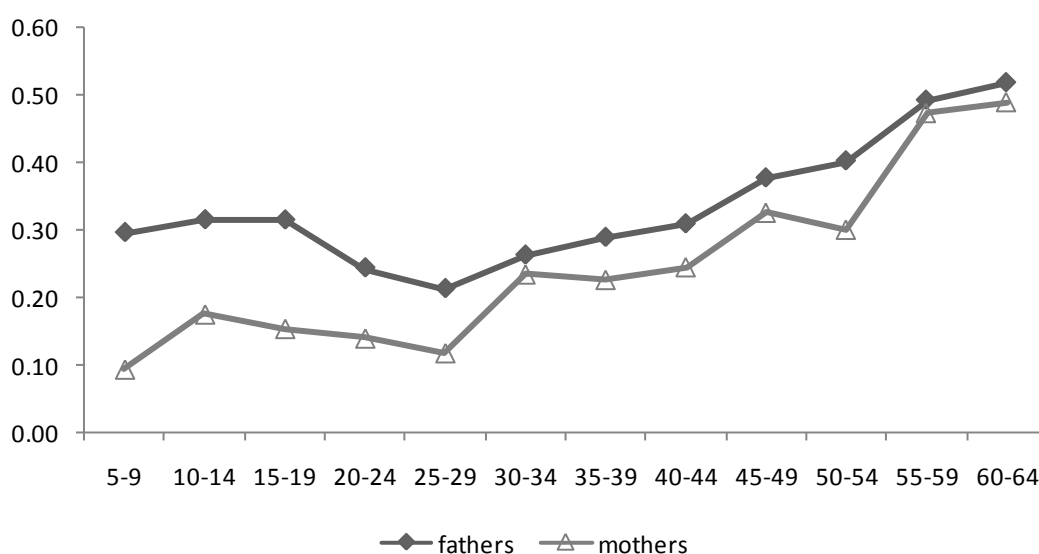


Figure 3.1 above shows the proportion of respondents who did not know the year of death of their mother/father. As expected the proportion of respondents with missing data increases with age. This is because respondents in older age groups will most likely have lost their parents in the distant past, thus they are more likely not to remember the year of death. What is also apparent from Figure 3.1 is the high proportion of respondents in the younger age groups who do not know their father's year of death. This could possibly indicate that the respondent did not know the father's whereabouts and consequently report the father as not alive, but can't give year of death.

We now look at these age distributions of the respondents by sex so as to see if there is any concentration of missing observations in either the male or female respondents. Figures 3.2 and 3.3 below show the age distribution of respondents by sex, who did not know the year of death of the father and mother respectively. As we can be observed from the two figures there is not a big difference between the distributions for the fathers and mothers.

Figure 3.2 Proportion of respondents who did not know father's year of death, by sex

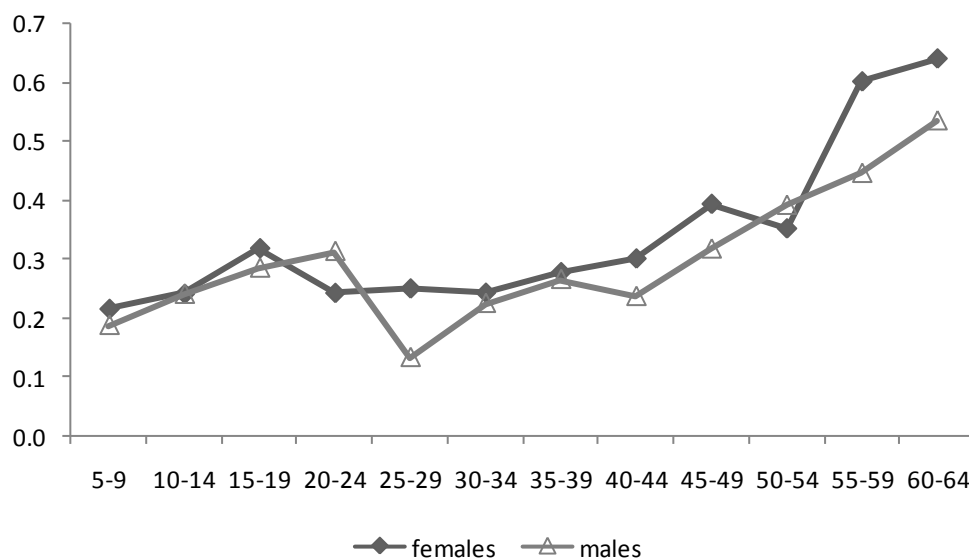
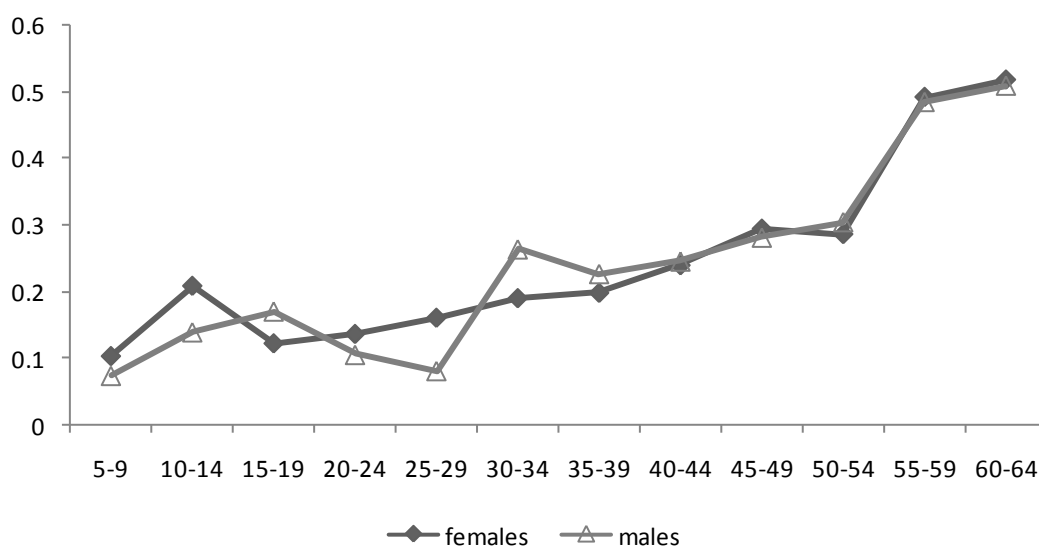


Figure 3.3 Proportion of respondents who did not know mother's year of death by sex



The demographic processes of fertility, migration and mortality vary by population group. It is therefore imperative to investigate whether the occurrence of missing observation is in anyway related to population groups. Table 3.1 and 3.2 below present the percentages of respondents who reported father's and mother's year of death by population group and sex. As illustrated in both tables the white population answered this question very well relative to the other groups, followed by the Africans. As expected all the population groups answered the question on mother's year of death better than the question on father's year of death. Although there are big differences in how the different population groups answered the question on year of death, the national percentage of those who answered is similar to that of the African population because it is the most popular of the population groups in South Africa.

Table 3.1 The percentages of respondents who reported year of father's death, by sex and population group

	<i>African</i>	<i>Asian/Indian</i>	<i>Coloured</i>	<i>White</i>	<i>National</i>
Males	68.0%	59.8%	49.8%	83.4%	68.0%
Females	60.2%	55.7%	48.6%	76.1%	60.6%

Table 3.2 The percentages of respondents who reported year of mother's death, by sex and population group

	<i>African</i>	<i>Asian/Indian</i>	<i>Coloured</i>	<i>White</i>	<i>National</i>
Males	71.2%	62.7%	52.6%	73.9%	69.7%
Females	65.4%	74.6%	45.2%	81.9%	65.7%

Having looked at the data on missing year of death by sex and population group we look at the data on orphanhood at age 5 and 15. Table 3.3 below shows the distribution

of respondents who did not know the year of death of father/mother. As can be seen, of the respondents who did not know the year of their father's death, 9 per cent were under the age of 15. About 33 per cent of these children did not respond to the question on survival at age 5, while of the remaining 91 per cent of respondents (adults), 10 per cent did not answer the question on the survival status of the father at age 5 and 15. On the other hand, of the proportion of respondents who did not have a living mother, 32 per cent did not know the year of mother's death. About 4 per cent of these respondents were under the age of 15 (i.e. they are classified as children) and the remaining 96 per cent over the age of 15 are classified as adults. An estimated 33 per cent of these children did not respond to the question on survival of mother at age 5 while an estimated 9 per cent of adults failed to respond to the question on survival of mother at age 5 and 15.

Table 3.3 Distribution of respondents who did not know father/mother year of death and whether father/mother was alive at age 5 and 15

%	<i>Respondents under 15 yrs</i>	<i>Respondents over 15 yrs</i>
<i>respondents who did not know father's year of death</i>	9	91
<i>respondents who did not know if father was alive at age 5 or 15</i>	33	10
<i>respondents who did not know mother's year of death</i>	4	96
<i>respondents who did not know if mother was alive at age 5 or 15</i>	33	9

3.3 The conventional orphanhood method

Using data from the 1996 and 2001 censuses, the 2007 Community Survey (CS) and the 2008 National Income Dynamics Survey (NIDS) the conventional orphanhood method was applied using regression coefficients from Timæus (1992). In the case of males, unpublished regression coefficients produced by Timæus were used for children in age group of 5-9 (personal communication, Professor R E Dorrington)¹. In order to calculate the conditional survival probabilities the average age of the parents at the time of birth/conception of their children is needed. To estimate these average ages, the estimated number of births (in the last 12 months) not imputed or edited in the 2001 census as provided by Moultrie and Dorrington (2004) was used. For mothers the average age was estimated as 27 years while for fathers the average was estimated to be

¹Timæus's unpublished regression coefficients: $a_0 = -0.5578$, $a_1 = 0.0004$, $a_2 = 1.4708$, $a_3 = 0.0695$

34 years. Although the average age may change over time, it is assumed, as was done by Dorrington, Moultrie and Timæus (2004) that the change was small over the period under consideration. The conditional probabilities of survival were then translated into a common mortality index ${}_{35}q_{15}$, that is, the probability of a 15 year old dying before age 50. To make this translation an appropriate life table, which takes into account the HIV/AIDS epidemic, is needed. The varying nature of the epidemic since the early 1990's makes it difficult to find a single standard life table which incorporates the impact of HIV/AIDS over time. In this research the Brass General Standard was used Dorrington, Moultrie and Timæus (2004) produced plausible results from the 2001 census data using this table. By assuming $\beta = 1$, an expression for α was derived as

follows: for women ${}_n p_{25} = \frac{l(25+n)}{l(25)} = \frac{1 + \exp(2\alpha + 2Y_s(25))}{1 + \exp(2\alpha + 2Y_s(25+n))}$ which implies that

$$l(25+n) * [1 + \exp(2\alpha + 2Y_s(25+n))] = l(25) * [1 + \exp(2\alpha + 2Y_s(25))] \text{ which}$$

simplifying and solving for α gives the following expression:

$$\alpha = -0.5 * \ln \left[1 + \left(\frac{\frac{{}_n p_{25}}{l_s(25+n)} - \frac{1}{l_s(25)}}{1 - {}_n p_{25}} \right) \right].$$

$$\text{Similarly for males we obtain } \alpha = -0.5 * \ln \left[1 + \left(\frac{\frac{{}_n p_{35}}{l_s(35+n)} - \frac{1}{l_s(35)}}{1 - {}_n p_{35}} \right) \right]$$

and calculate ${}_{35}q_{15} = 1 - {}_{35}p_{15} = 1 - \frac{l(50)}{l(15)} = 1 - \frac{1 + \exp(2\alpha + 2Y_s(15))}{1 + \exp(2\alpha + 2Y_s(50))}$. The Brass and

Bamgboye (1981) time location method was used to estimate the dates to which the estimates apply.

3.3.1 Adjusting for HIV/AIDS

The proportion of respondents aged 5-9 and 10-14 with living mother or father is adjusted for the bias introduced by HIV/AIDS. Timæus and Nunn (1997) assume that the fertility ratio of HIV positive women to HIV negative women is 0.8 and that the perinatal vertical transmission rate is 0.25, however, these assumptions do not necessarily apply to South Africa. The fertility ratio of HIV positive women to HIV negative women in South Africa appears to be a bit higher than this because young

women have a higher fertility rate than young HIV negative women to an extent that offsets the fact that older HIV positive women are less fertile than older HIV negative women. The Actuarial Society of South Africa (2005), ASSA 2003 model assumes this ratio to be above one while the UN spectrum model assumes this ratio to be slightly below one. Thus in this research it was decided to use one as between the assumptions of the two models. On the other hand the rate of perinatal transmission in South Africa is lower than that suggested by Timæus and Nunn (1997) mainly because of the introduction of Prevention of Mother to Child Transmission (PMTCT) and antiretroviral therapy programs. A perinatal transmission rate of 0.2 is assumed in this research as was assumed by the ASSA 2003 model. These assumptions are substituted

into the expression $\frac{S(a)}{S^*(a)} = \frac{1-hP}{1+\frac{1-F}{F}P}$ to get ${}_5S_x' = (1-0.2P){}_5S_x^*$, where $S(a)$ is

the proportion still alive of women who would have given birth a years ago in the absence of HIV, $S^*(a)$ represents the proportion of respondents aged a who report that their mother is still alive, h represents the vertical transmission, F is the reduction in fertility of HIV positive women and P represents prevalence at the time of the children's birth (Timæus and Nunn 1997). The proportion of respondents aged 10-14 with mother alive is thus reduced by 20 per cent of the estimated prevalence at the time the children were born and of respondents aged 5-9 is reduced by 10 per cent of the estimated prevalence.

The proportion of children aged 10-14 with father alive is reduced by 60 per cent of the 20 per cent reduction of the estimated prevalence at the time when the children were born and the proportion age 5-9 with father alive is reduced by 60 per cent of the 10 per cent reduction of the estimated prevalence (Timæus and Jasseh 2004). The prevalence P given in Table 3.1 was estimated from ASSA (2003) model as the antenatal clinic HIV prevalence.

Table 3.4 Antenatal HIV prevalence, South Africa

<i>year</i>	<i>prev</i>	<i>year</i>	<i>prev</i>
1989	0.2%	1999	20.0%
1990	0.4%	2000	22.3%
1991	0.9%	2001	24.1%
1992	1.6%	2002	25.5%
1993	2.9%	2003	26.5%
1994	4.8%	2004	27.3%
1995	7.4%	2005	27.9%
1996	10.5%	2006	28.3%
1997	13.9%	2007	28.7%
1998	17.2%	2008	28.9%

Source: ASSA 2003 model, (Actuarial Society of South Africa 2005)

Regression Coefficients from Timæus (1992) were estimated at a time when the prevalence of HIV was very low and when applied in populations with a high prevalence of HIV they tend to overestimate life table survivorship of women between 25 and 35 years (Timæus and Nunn 1997). To avoid such problems, revised regression coefficients from Timæus and Nunn for mothers with children aged 5-9 and 10-14 (i.e. born after 1995) were used. For the rest of the age groups, coefficients from Timæus (1992) were used because these children were born at a time when the prevalence was still low (i.e. lower than 2 per cent) or non-existent.

3.4 The conventional synthetic cohort method

To produce estimates of mortality for the inter-census and inter-survey periods the synthetic cohort method developed by Timæus (1991b) was applied. This variation of the orphanhood method was used because it minimises the effects of the 'adoption effect' by concentrating on respondents who are aged 20 and above. Regression coefficients from Timæus (1991b) were used while the average age of the parent at the time the children were born was assumed to be 27 and 34 years for mothers and fathers respectively. This information was used to calculate conditional probabilities of survival $l(25+x)/l(45)$ for females and $l(35+x)/l(55)$ for males.

To translate the conditional probabilities into a common index of mortality there is the need again to select a suitable standard life table which takes into account the AIDS epidemic. The United Nations Population Division (2009) has constructed life tables which take into account the AIDS epidemic. To construct these life tables the UNPD collects the most recent population data from censuses and population registers and if

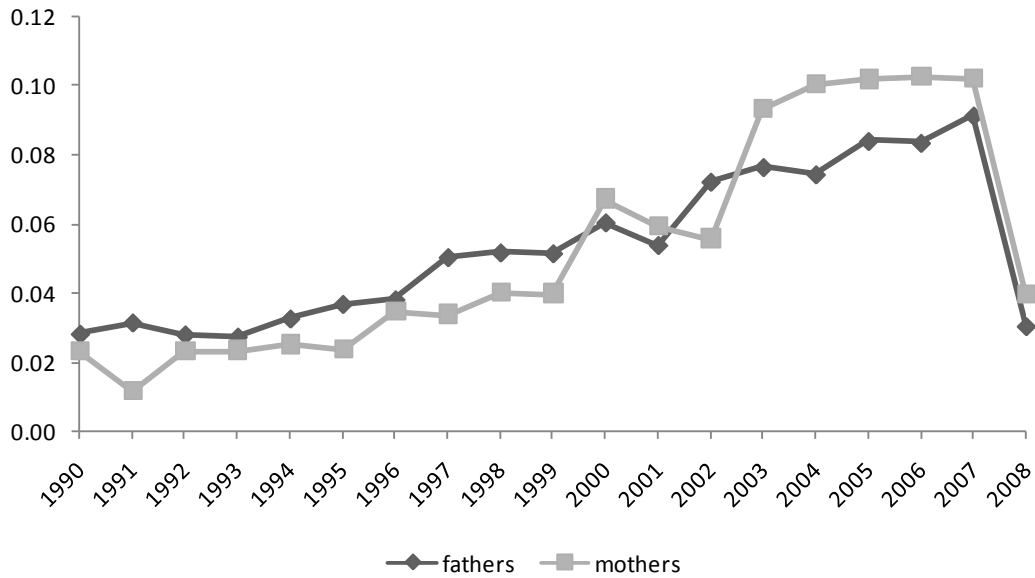
necessary the collected data is adjusted for completeness, accuracy and consistency. Using the collected data and a set of assumptions on the future fertility, migration and mortality patterns, the UNPD estimates past and future fertility, mortality (including HIV related deaths), migration and overall population. The UNPD used deaths and populations calculated from this information to construct five year period life tables from 1995 to 2010 which take into consideration the AIDS epidemic. In this research simple linear interpolation was used to construct period life tables that correspond to the dates of the censuses and surveys in this research. Using these life tables and the Brass Logit transformation and assuming $\beta = 1$, conditional survival probabilities were translated to ${}_{35}q_{15}$, the probability of a person aged 15 dying before reaching age 50.

3.5 Using year of death data

The starting point for using year of death data as suggested by Chackiel and Orellana (1985) is to estimate the mean year of death and the proportion with father and mother alive 5 and 10 years before the survey. In this research the mean year of death was calculated from respondents who answered 'no' to the question 'is your mother/father alive'. An obstacle to the use of the year of death data in this research is the presence of a high non-response to the question on year of death. Chackiel and Orellana did not mention having any non-response or how to treat such a situation if it occurs.

Before proceeding to applying the method proposed by Chackiel and Orellana there is the need to decide on how to deal with the missing data. The starting point is to look at the distribution of reported years of death. Figure 3.7 below shows the distribution of these years of death for both fathers and mothers. The majority of deaths (parents) occurred after 1990 so for ease of interpretation Figure 3.7 shows the distribution of these deaths from 1990 onwards (although deaths before 1990 were included in the analysis). As shown in Figure 3.7 the deaths reported for fathers and mothers increase with time. This is what one would expect given the fact that in South Africa mortality has been on the increase since the mid 1990s, increasing rapidly from the year 2000.

Figure 3.4 Distribution of reported years of deaths



Having accepted the distributions shown in Figure 3.4, we proceed to use this distribution to distribute deaths with missing year of death. Here the assumption is the distribution of deaths with missing year of death is the same as the deaths that were reported. Respondents aged x with an implausible year of death of father/mother were treated as respondents aged x with missing year of death. This assumption will not have an significant effect on the results since the population of those with implausible responses constitute 0.91 and 0.30 per cent of those who responded to the question on father/mother year of death respectively.

The mean year of death of the mothers and fathers was calculated as follows:

$$M_k = \frac{\sum_{i=0}^4 \sum_j d_{kij} Y_{kij}}{\sum_{i=0}^4 \sum_j d_{kij}}$$

where M_k is the mean year of death for each child's age group k to

$k+4$, where $k = 5, 10, \dots, 60$ and d_{kij} represents the number of dead parents of children aged $k+i$ and Y_{kij} represents the year of death of dead parents of children aged $k+i$.

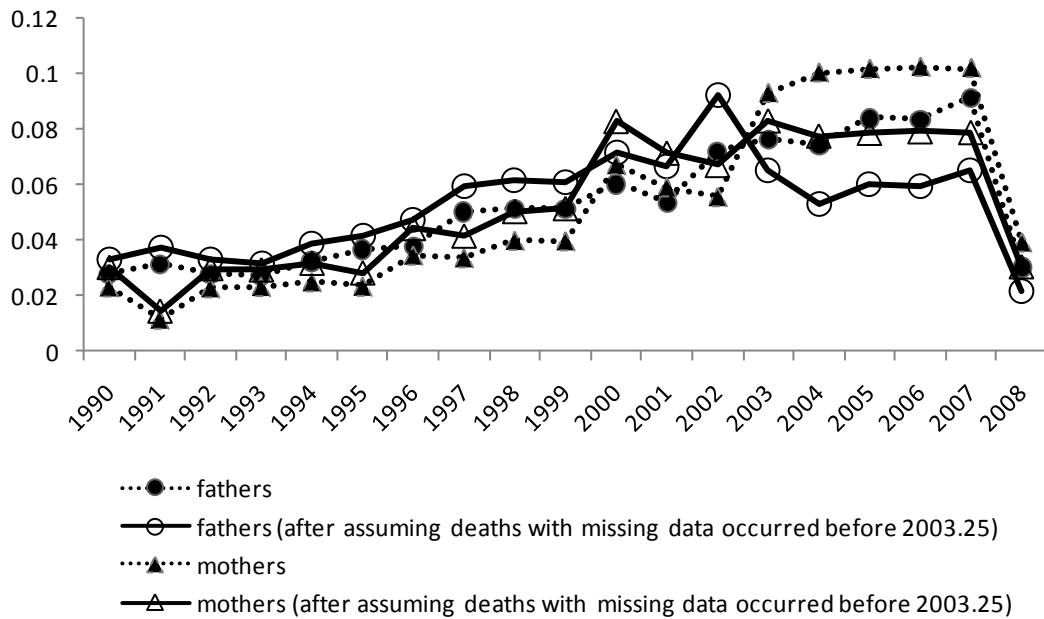
Having calculated the mean year of death, proportions with mother/father alive 5 and 10 years prior to the 2008 NIDS survey were calculated by adding back all the deaths that occurred for the period 1998.25 to 2003.25 and 2003.25 to 2008.25 (assuming that the deaths are uniformly distributed over calendar years) to those respondents with father and mother alive in age group x (where x represents age groups from 15-19 up to

age group 50-54) at the time of the survey and then dividing by the total number of respondents who responded in that particular age group x .

Having obtained the proportions with mother and father alive for the two synthetic surveys at 1998.25 and 2003.25, the orphanhood method was applied as described in section 3.3 to produce estimates of mortality. To produce estimates of mortality for the periods between the two synthetic surveys and the actual NIDS survey the synthetic cohort variation of the orphanhood method was applied using coefficients from Timæus (1991b) which utilizes data from respondents who are aged 20 years and older.

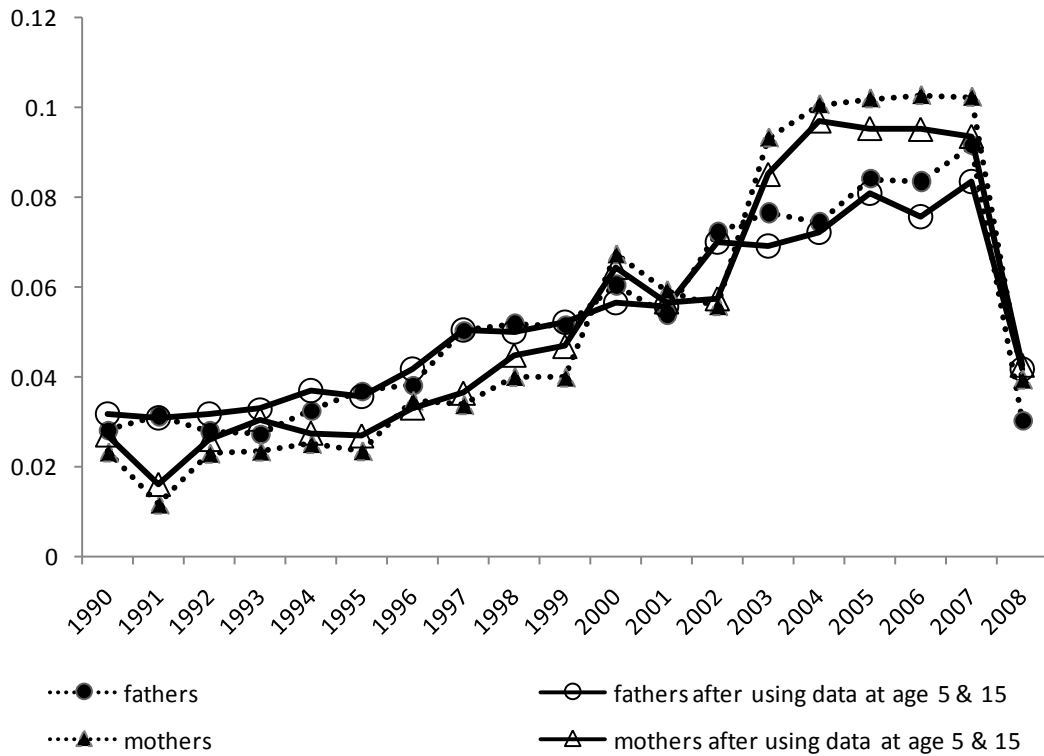
The quality of estimates that are produced using the year of death data are affected by the high non-response to the question on year of death. In a bid to improve the quality of estimates that were produced especially between the synthetic surveys of 2003 and 2008 it was assumed that all deaths of fathers and mothers with unknown year of death occurred before 2003.25. This assumption seems reasonable since respondents are more likely to remember the year of death of a death that occurred in the recent past than in the distant past. Figure 3.5 below shows the distributions of reported years of death before and after making the assumption that deaths with missing year of death occurred before 2003.25. The distributions are bound not to be the same after making this assumption but what is worrying is there are spikes, one in the year 2000 and the other in year 2002. These could affect the quality of estimates that are produced using the single survey orphanhood method. For females these spikes might not have much of an effect because they occur at roughly the time that mortality was increasing at a rapid rate but for males the spike is roughly between 2002 and 2003 when mortality was levelling off. Despite the assumption not producing a distribution similar to the original distribution of reported deaths it might still give useful results after applying the synthetic cohort method. After making this assumption, all deaths with unknown years of death were distributed proportionately for all the years from 2003.25 going backwards. The mean year of death and the proportion of children with mother surviving for the 2 synthetic surveys were calculated.

Figure 3.5 Distribution of reported years of death, after assuming that deaths with missing years of death occurred before 2003.25



The National Income Dynamics Study (NIDS) collected data on survival of the parent at age 5 and 15 of the child. These data can be of great use with regards to estimating the timing of death of those without year of death. In this research these data were used to distribute deaths with unknown years of death. Deaths of parents without year of death were distributed uniformly between the child's year of birth and the year the child turned 5 or 15 depending on whether the parent was alive at age 5 or 15. Those who did not respond to the question on orphanhood had their parents deaths distributed between year of birth and year of survey. The distributions of the reported deaths before and after the imputation are compared. Figure 3.6 shows that there is not much difference between the two sets of distributions. The two distributions are similar but closer inspection reveals that the distribution after imputation has fewer deaths in the recent past and more in the distant past than before the imputation. If the distribution of the reported deaths after the imputation was completely irregular and not showing any resemblance to the distribution before the imputation it would cast doubt on the quality of the data on orphanhood at age 5 and 15. If the data is not of reasonable quality it might introduce bias onto the estimates that we produce. Having accepted the data as credible we proceed to apply Chackiel and Orellana's proposed method.

Figure 3.6 Distribution of reported deaths before and after using orphanhood data at age 5 and 15 of the respondent



3.6 The ASSA 2003 model and how it was used

The ASSA 2003 model is a mathematical model developed by the Actuarial Society of South Africa (ASSA) to make projections on the future course of the HIV/AIDS epidemic. Through the use of various assumptions on fertility, migration and mortality the model is designed so as to reproduce the past and current number of HIV and non-HIV related deaths. A more detailed description of the model and its various assumptions can be found in the ASSA2003 user guide which can be downloaded from the Actuarial Society of South Africa website www.assa.org.za. In this research estimates of mortality ${}_{35}q_{15}$ from the ASSA model were selected to be the benchmark. Table 3.2 below presents ${}_{35}q_{15}$ estimated by the ASSA model from 1986 to 2008. These estimates help in deciding whether estimates derived using year of death data are any better than estimates derived using conventional orphanhood methods.

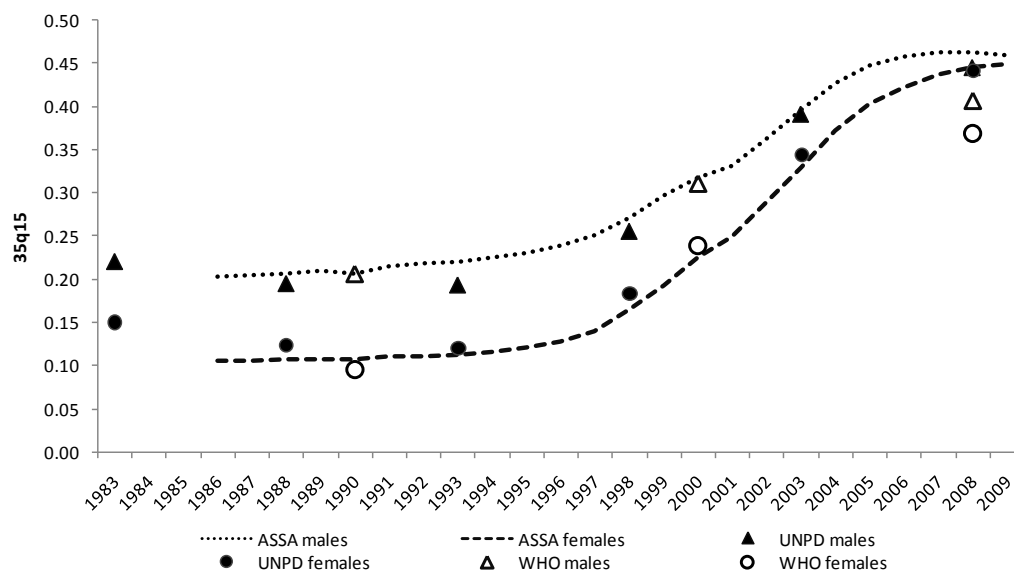
Table 3.5 Estimates of adult mortality in South Africa, 35q15

Year	Males	Females	Year	Males	Females	Year	Males	Females
1986	0.202	0.105	1994	0.224	0.115	2002	0.362	0.288
1987	0.204	0.106	1995	0.230	0.120	2003	0.395	0.330
1988	0.206	0.107	1996	0.238	0.127	2004	0.426	0.371
1989	0.209	0.108	1997	0.251	0.139	2005	0.447	0.403
1990	0.207	0.106	1998	0.272	0.165	2006	0.457	0.422
1991	0.214	0.110	1999	0.297	0.193	2007	0.462	0.437
1992	0.217	0.111	2000	0.317	0.225	2008	0.463	0.445
1993	0.220	0.112	2001	0.331	0.249			

Source: ASSA 2003 model, (Actuarial Society of South Africa 2005)

The selection of ASSA (2005) estimates as benchmark is just a matter of preference and accessibility, there are other alternatives that are accessible and could have been used as benchmark, such as the UNPD (2009) estimates. Figure 3.7 below compares the ASSA estimates and estimates from UNPD (2009) and WHO (2010). Comparison of the estimates from all the three organisations show that the estimates are in agreement except for the recent time periods where point estimates from WHO seem to be much lower than those from ASSA and UNPD. The difference in estimates for the recent time periods could be the result of differences in methodology or assumptions that might or might not have been taken into consideration by WHO. Despite these differences the estimates are consistent for all the other time periods. This gives credibility to the use of the ASSA estimates as a benchmark.

Figure 3.7 The probability of dying between ages 15 and 50, South Africa



4.1 Introduction

This chapter is divided into three sections. The first section, 4.2 presents the results produced using year of death data to estimate time reference points and compares these estimates with estimates produced using the Brass and Bamgboye (1981) method. The second section, 4.3 compares estimates of mortality produced using year of death data and those produced using the conventional orphanhood method. The last section presents results produced by applying the synthetic cohort method to the synthetic surveys and compares these results with those produced by the conventional synthetic cohort method.

4.2 Time location of estimates

The results of using year of death data and the Brass and Bamgboye method to estimate time reference points are presented in Tables 4.1 and 4.2 for paternal and maternal estimates respectively. The results presented in both tables are very difficult to interpret because of the erratic nature of the difference between the two sets of time reference points. With the exception of the 1998 synthetic survey, the difference between the two sets of reference points seem to increase with age up to the middle age group of 25-29 and thereafter appear to decrease. It is not very clear why the difference between the time reference points is like this. What is clear is that time reference points derived from the year of death data are located more into the near past than time reference points derived from the Brass and Bamgboye method with an average difference of about 2 years. Chackiel and Orellana in their study showed that both methods of calculating time location produce similar results for younger age groups with reference points from year of death data stretching back in time at older age groups. In this research this is clearly not the case. One has to suspect the year of death data to be the cause of this erratic trend (of the difference column). This suspicion is driven by the fact that year of death data contain a high proportion of missing data which had to be redistributed. But then again the reference points derived from the Brass and Bamgboye method have a problem of their own. A look at the column of theoretical reference points for females shows that these fail to go back in time which should not be the case. The problem may not lie with the theoretical method but rather with the proportion of orphans in the higher age groups of 40-44 and 45-49. Some deaths of these mothers would have taken place many years in the past thus respondents with these mothers are prone to recall

bias. Brass (1975) mentions this problem and it seems to be what is causing the theoretical points to fail to extend sufficiently back in time. This problem of recalling dates is common in Africa where the knowledge of dates is not as widespread as it is in South America. Chakiel and Orellana do not seem to have encountered such a problem because their field work on the question on year of death was done in South America. There is therefore a need for more information so as to be able to judge which reference points are more credible. This can be done by plotting mortality estimates with both sets of reference points and comparing them with our benchmark.

Table 4.1 Comparison of empirical and theoretical time reference points, paternal

Age group	2008.25			2003.25			1998.25		
	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff
5-9	2004.6	2003.2	-1.5	1999.9	1998.3	-1.6	1995.0	1993.3	-1.7
10-14	2003.4	2001.0	-2.4	1998.6	1996.2	-2.4	1992.7	1991.4	-1.3
15-19	2001.8	1999.0	-2.8	1996.8	1994.4	-2.4	1991.4	1989.5	-2.0
20-24	2000.6	1997.5	-3.1	1994.7	1992.8	-2.0	1989.4	1988.0	-1.3
25-29	1998.6	1996.0	-2.6	1993.6	1991.4	-2.2	1987.1	1986.8	-0.3
30-34	1997.5	1994.4	-3.1	1991.3	1990.3	-0.9	1985.9	1985.9	0.0
35-39	1994.3	1994.3	0.0	1989.5	1990.1	0.6	1983.4	1985.9	2.5

Table 4.2 Comparison of empirical and theoretical time reference points, maternal

Age group	2008.25			2003.25			1998.25		
	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff
5-9	2005.0	2004.8	-0.2	2000.2	1999.8	-0.4	1995.8	1994.8	-1.0
10-14	2003.6	2002.4	-1.2	2000.0	1997.5	-2.5	1992.7	1992.6	-0.1
15-19	2003.2	2000.3	-3.0	1997.8	1995.4	-2.4	1990.5	1990.5	0.0
20-24	2001.6	1998.4	-3.2	1995.8	1993.7	-2.1	1991.2	1988.8	-2.4
25-29	2000.0	1997.0	-3.0	1995.0	1992.3	-2.7	1990.3	1987.4	-2.9
30-34	1998.4	1996.0	-2.5	1994.4	1991.2	-3.1	1987.4	1986.1	-1.3
35-39	1997.9	1995.2	-2.7	1991.8	1990.1	-1.7	1986.4	1985.9	-0.4
40-44	1995.2	1994.1	-1.1	1991.5	1990.4	-1.1	1985.5	1985.3	-0.2
45-49	1994.9	1994.8	-0.1	1989.5	1990.2	0.7	1984.5	1986.3	1.8

Figure 4.1 and 4.2 below present results of estimates of mortality derived from the NIDS data set using the conventional orphanhood method for paternal and maternal mortality respectively. Both figures use the same estimates of mortality, ${}_{35}q_{15}$, which are plotted using reference points from the Brass and Bamgboye method and from year of death data and compares these estimates to the ASSA benchmark estimates. Figure 4.1 shows, for males, that estimates plotted using reference points from the year of death data (2008 Survey (YOD)) are consistent with the benchmark estimates while those plotted using the Brass and Bamgboye method (2008 NIDS) overestimate mortality. Figure 4.2 on the other hand shows, for females, that estimates plotted using reference points derived from year of death data underestimate mortality while estimates plotted using reference points from the Brass and Bamgboye method are consistent

with the ASSA benchmark estimates. From the evidence gathered from the two figures it is quite difficult to make a judgement in favour of one set of reference points as representing the correct time location. There is therefore a need to make further assessments to see which set of reference points is credible. At this stage we take a look at the results produced by using data on orphanhood at ages 5 and 15 of the respondent to distribute deaths with missing year of death.

Figure 4.1 Male mortality estimates derived from the 2008 NIDS survey plotted using reference points from the Brass & Bamgboye method and from year of death data (YOD)

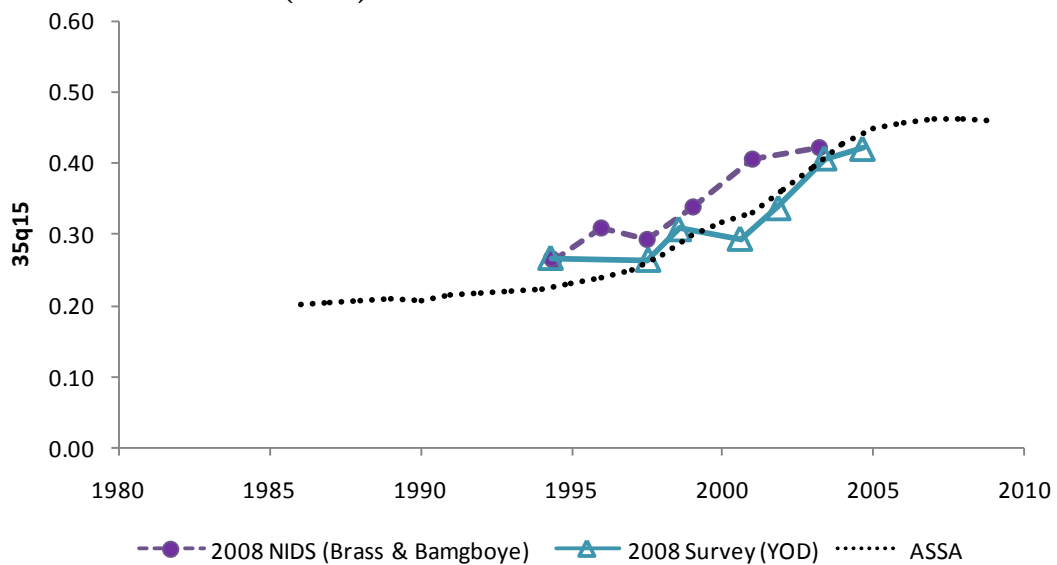
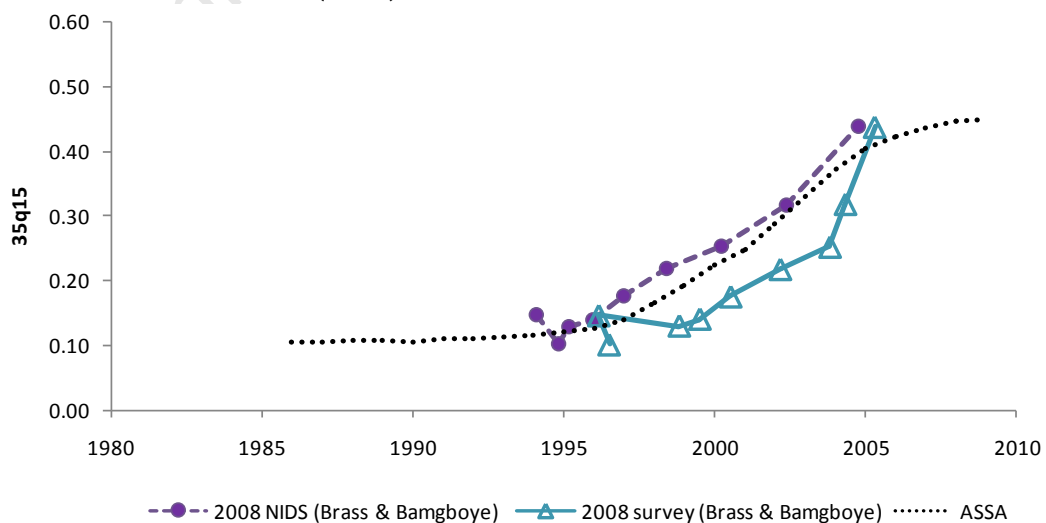


Figure 4.2 Female mortality estimates derived from the 2008 NIDS survey plotted using reference points from the Brass & Bamgboye method and from year of death data (YOD)



Tables 4.3 and 4.4 below presents the time reference points produced after using data on orphanhood at age 5 and 15 of the respondent for fathers and mothers respectively. A look at the difference columns in both tables shows that the reference points are now closer to each other than before. This is particularly significant with the 2003 and 1998 synthetic surveys. Another apparent feature in both tables is that the reference points are stretching much further back into time than before. This second feature is what Chackiel and Orellana observed in their study in South America. This evidence casts doubts on the credibility of the reference points derived using year of death data and the indications are that these points are biased towards the more recent past.

Table 4.3 Comparison of empirical and theoretical time reference points, empirical time location calculated from year of death data imputed by using data on survival of father at age 5 and 15

Age group	2008.25			2003.25			1998.25		
	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff
5-9	2004.6	2003.2	-1.4	1999.6	1998.3	-1.3	1994.7	1993.3	-1.4
10-14	2002.9	2001.0	-1.9	1997.9	1996.1	-1.8	1991.7	1991.3	-0.4
15-19	2001.2	1999.0	-2.2	1995.9	1994.4	-1.5	1990.7	1989.3	-1.3
20-24	1999.8	1997.5	-2.3	1993.7	1992.6	-1.0	1988.4	1987.9	-0.6
25-29	1997.3	1996.0	-1.3	1992.3	1991.3	-1.0	1986.2	1986.5	0.3
30-34	1996.3	1994.4	-1.9	1989.9	1990.2	0.3	1984.2	1985.5	1.4
35-39	1993.1	1994.3	1.2	1987.6	1989.7	2.1	1982.2	1985.5	3.3

Table 4.4 Comparison of empirical and theoretical time reference points, empirical time location calculated from year of death data imputed by using data on survival of mother at age 5 and 15

Age group	2008.25			2003.25			1998.25		
	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff	Empirical	Theoretical	Diff
5-9	2005.3	2004.8	-0.5	1999.9	1999.8	-0.1	1994.7	1994.8	0.1
10-14	2003.8	2002.4	-1.4	1999.2	1997.5	-1.7	1991.7	1992.5	0.9
15-19	2003.3	2000.3	-3.1	1997.1	1995.4	-1.7	1990.7	1990.5	-0.2
20-24	2001.5	1998.4	-3.1	1995.1	1993.7	-1.5	1988.4	1988.8	0.3
25-29	1999.9	1997.0	-2.9	1994.1	1992.2	-1.9	1986.2	1987.3	1.1
30-34	1998.4	1996.0	-2.4	1991.4	1991.2	-0.2	1984.2	1985.9	1.8
35-39	1996.4	1995.2	-1.2	1989.8	1989.9	0.1	1982.2	1985.6	3.3
40-44	1994.1	1994.1	0.0	1988.6	1990.1	1.6	1979.7	1984.6	4.9
45-49	1993.5	1994.8	1.3	1986.2	1989.6	3.3	1977.6	1984.6	7.0

To get a feel for what is going on we plot estimates of mortality derived from the 2008 NIDS survey with reference points before and after imputing missing data using data of orphanhood at age 5 and 15 of the respondent. Figures 4.3 and 4.4 below illustrate the effect of using the data on orphanhood at age 5 and 15 of the respondent on the time reference points derived from year of death data. As shown in both figures the extra information helps the estimates to stretch back in time and also brings the estimates closer to the benchmark estimates. This shows that the time reference points derived

from year of death data before using data on orphanhood at age 5 and 15 of the respondent are biased towards the more recent past.

Figure 4.3 Estimates of paternal mortality from the 2008 NIDS survey plotted using different reference points

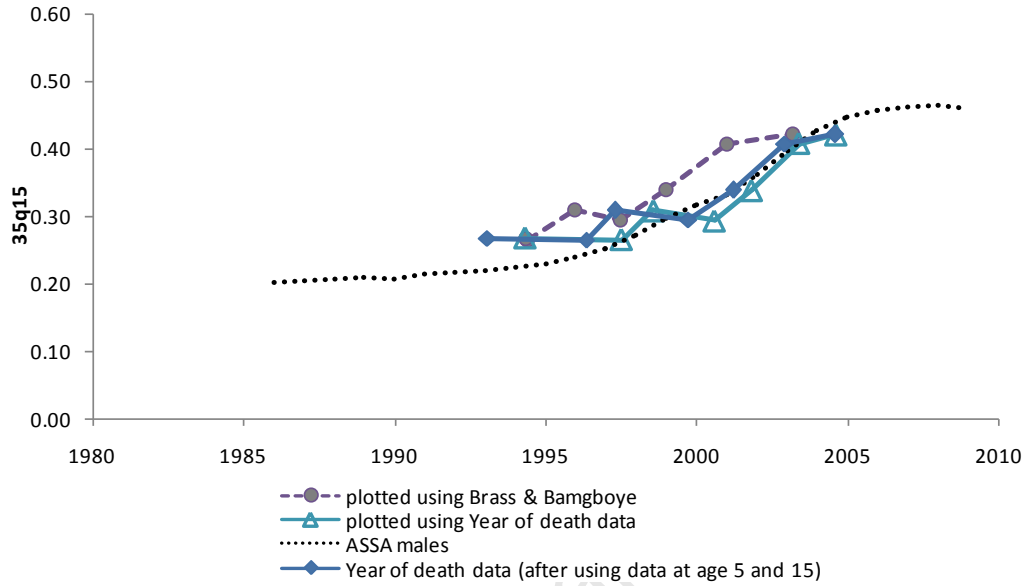
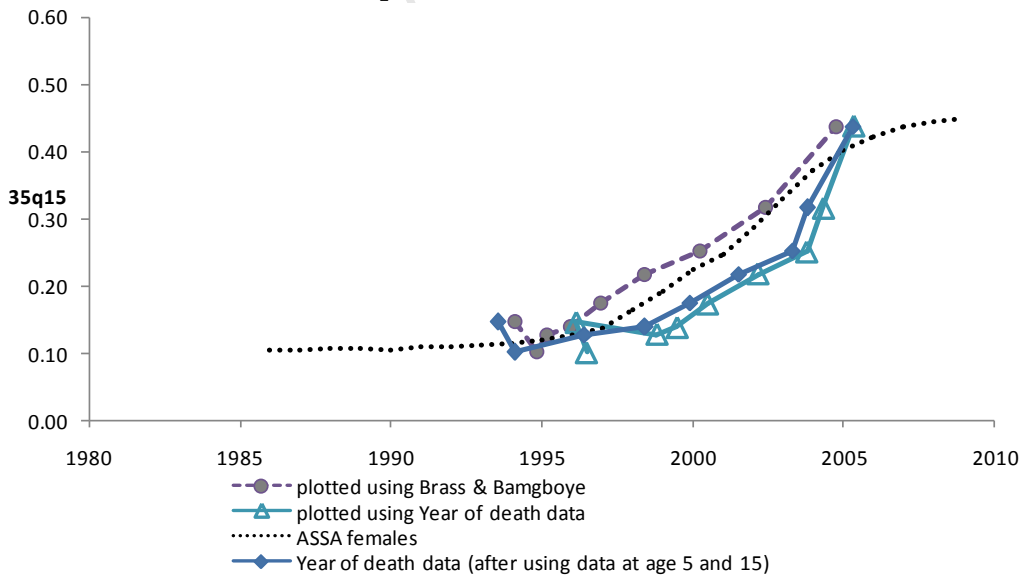


Figure 4.4 Estimates of maternal mortality from the 2008 NIDS survey plotted using different reference points



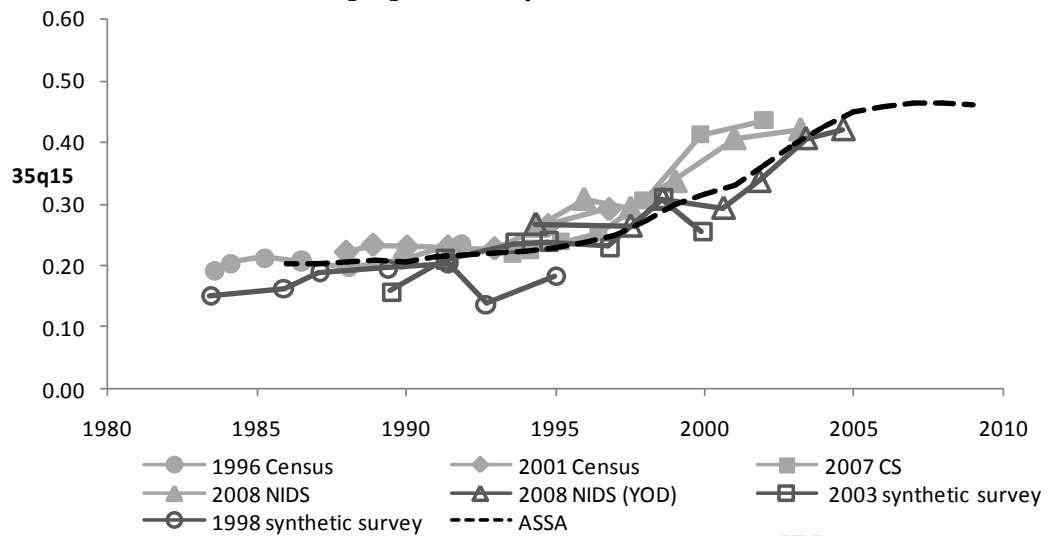
4.3 Single survey estimates of maternal and paternal orphanhood

4.3.1 Single survey estimates of maternal and paternal orphanhood after distributing deaths with missing year of death proportionately up to 2008.23

In this section estimates of mortality derived from using the year of death data are presented and compared to estimates derived from the conventional orphanhood method. Figures 4.5 and 4.6 below present estimates from the two approaches for males and females respectively. In both figures mortality estimates derived using year of death data are plotted using reference points calculated from the year of death data itself while estimates from the conventional method are plotted using reference points from the Brass and Bamgboye method.

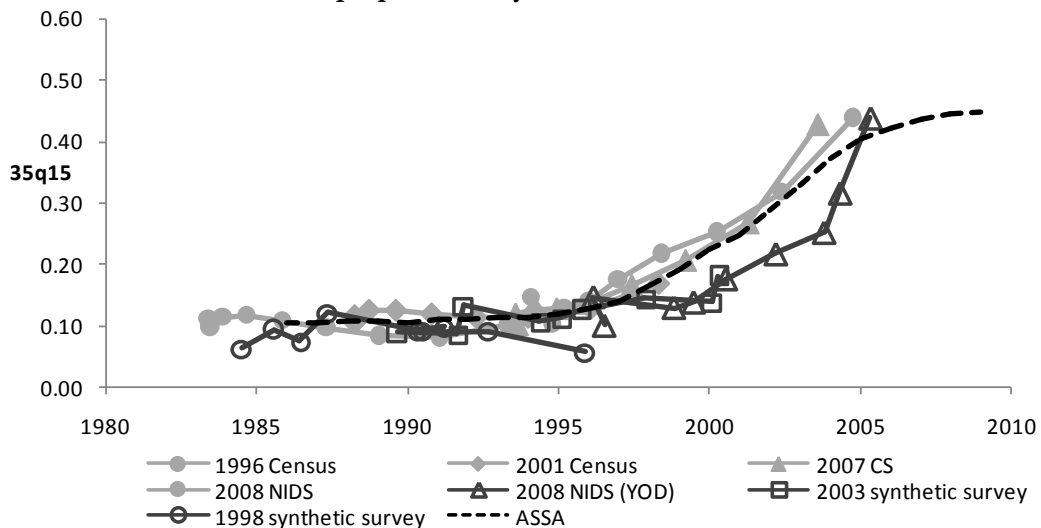
As shown in Figure 4.5 estimates of male mortality derived from year of death data are consistent with estimates from the ASSA model except for estimates derived from the 1998 synthetic survey which seem to underestimate mortality. This underestimating of mortality also indicates that some deaths which belong in the distant past might have been distributed into the more recent past. The argument is further strengthened by the fact that estimates derived from the conventional method for the same period are consistent with the ASSA model. Comparing the estimates from the two approaches we observe that estimates derived from the year of death data have a better fit from the year 1995 and onwards while those from the conventional approach seem to overestimate mortality. The reason why estimates from the conventional method overestimate mortality for the recent past is not obvious. The problem can be attributed to the Brass and Bamgboye dating procedure or it could be attributed to the problem of missing fathers who were reported as not being alive.

Figure 4.5 Estimates of male mortality derived using the conventional orphanhood method and using year of death data, after distributing deaths with unknown YOD proportionately



Unlike male estimates, female estimates derived from using year of death data underestimate mortality. As shown in Figure 4.6 below the use of year of death data results in lighter mortality than is portrayed by estimates derived by the ASSA model, while estimates from the conventional method are consistent with those from the ASSA model for almost all the reference years. Doubts have to be cast on the way deaths with missing years of death were distributed. As illustrated in Figure 4.4, distributing the deaths with missing year of death proportionately will result in an underestimation of mortality especially for the recent past.

Figure 4.6 Estimates of female mortality derived using the conventional orphanhood method and using year of death data, after distributing deaths with unknown YOD proportionately



4.3.2 Single survey estimates of maternal and paternal orphanhood after assuming that deaths with missing year of deaths occurred before 2003.25

The proportion of deaths with missing year of death seem to pose a big problem since distributing these deaths proportionately results in underestimation of mortality. The main problem is that most of these deaths belong in the distant past and distributing these deaths proportionately would bring some of these deaths into the recent past. In this subsection we present the results obtained when all these deaths with missing year of death are assumed to have occurred before 2003.25.

Figures 4.7 and 4.8 below present the estimates produced after making this assumption for males and females respectively. In both figures the deaths with missing year of death were distributed proportionately up to the year 2003.25. Looking at Figure 4.7 one can observe that estimates from the 1998 synthetic survey are now consistent with the ASSA model which is an improvement from the first approach of distributing deaths up to the survey date. However, this approach leads to an unfortunate overestimating of mortality by the estimates derived from the 2003 synthetic survey.

Results shown in Figure 4.8 on the other hand are more positive in that most estimates compare well with the ASSA estimates. Although estimates in the recent past still depict an underestimation it is an improvement from the case when we distributed deaths proportionately up to the survey date.

Figure 4.7 Estimation of male mortality derived from orphanhood data and parents year of death, assuming all unknown years of death occurred 5 years prior to survey

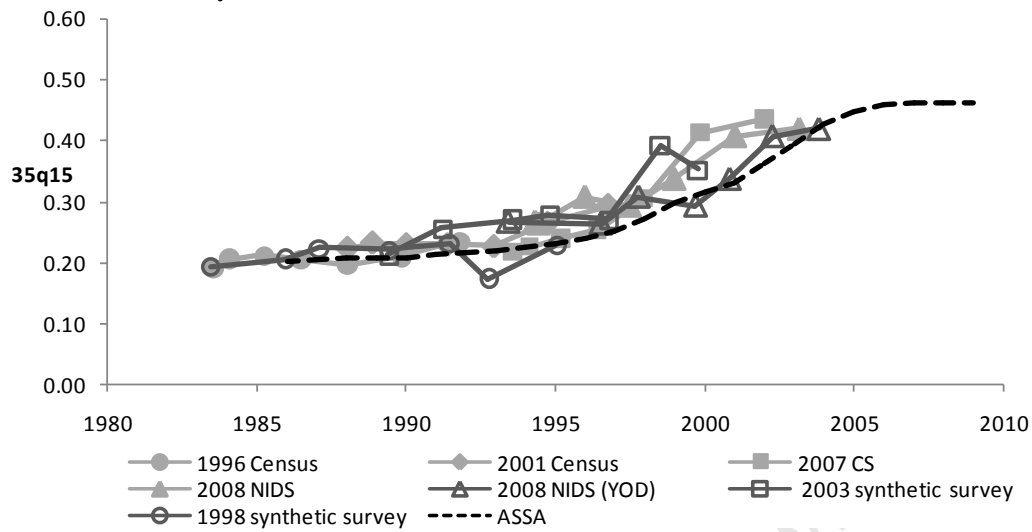
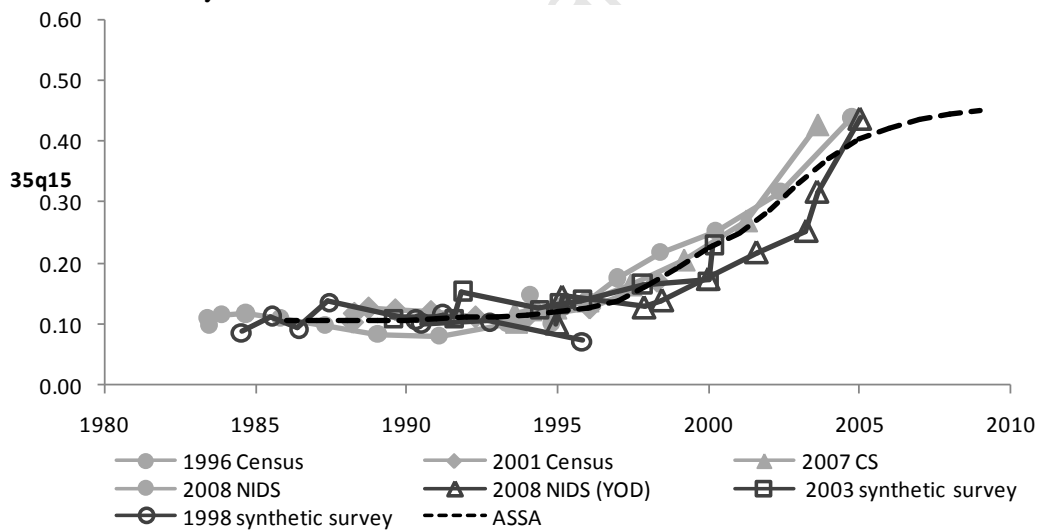


Figure 4.8 Estimation of female mortality derived from orphanhood data and parents year of death, assuming all unknown years of death occurred 5 years prior to survey



4.3.3 Single survey estimates of maternal and paternal orphanhood after imputing year of death data using data on orphanhood at age 5 and 15 of the respondent

Figures 4.9 and 4.10 below show estimates of mortality produced from the conventional method and from the year of death data after imputing the data using data on orphanhood at ages 5 and 15 of the respondent. As shown in Figure 4.9 male estimates derived from the year of death data are internally consistent and consistent with the ASSA estimates. Comparing with the estimates produced in sub-sections 4.3.1 and 4.3.2

these estimates have a much better fit. In addition to comparing well with the ASSA model the estimates stretch back for a much wide span of years compared to the cases in sub-sections 4.3.1 and 4.3.2. Female estimates from the year of death data also stretch for a wider span of years but estimates in the recent past do not compare well with the ASSA estimates for the same time period. This could possibly indicate that respondents did not answer well the question on mother's year of death. Despite this lack of fit for the recent estimates, female estimates from the year of death data are consistent internally and are much closer to the estimates from the ASSA model compared to similar estimates presented in sub-sections 4.3.1 and 4.3.2.

Figure 4.9 Estimation of male mortality derived from orphanhood data and father's year of death, with unknown year of death distributed using data on survival of father at age 5/15 of the child

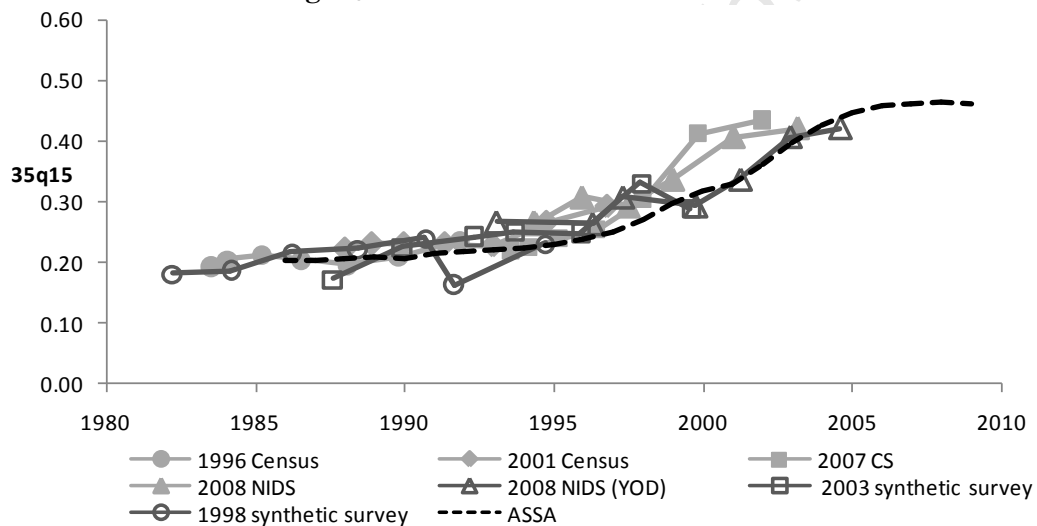
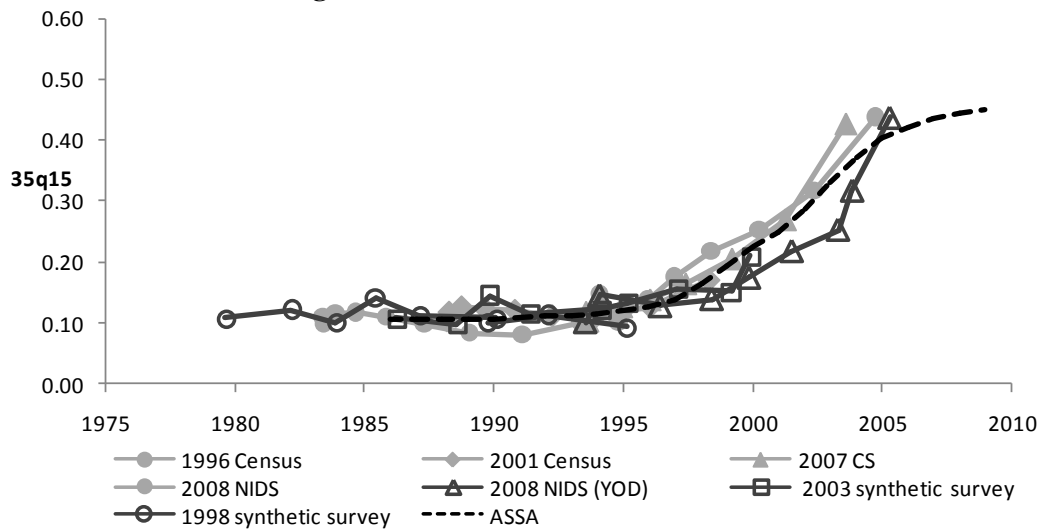


Figure 4.10 Estimation of female mortality derived from orphanhood data and mother's year of death, with unknown year of death distributed using data on survival of mother at age 5/15 of the child



4.4 Synthetic Cohort Estimates of Maternal and Paternal orphanhood

Having looked at estimates of mortality derived using the single survey method we now present the results produced using the synthetic cohort orphanhood method.

Figures 4.11 and 4.12 below presents paternal and maternal point estimates respectively derived using the conventional synthetic cohort method from the 1996 census, 2001 census, 2007 Community survey and the 2008 NIDS survey together with point estimates derived from year of death data (before and after assuming that deaths with missing year occurred before 2003.25 and after using data on age 5 and 15 to distribute the deaths with missing years).

With the exception of the estimate derived from the 2001 census and the 2008 NIDS survey, estimates derived using the conventional orphanhood method are remarkably consistent with the ASSA model. As shown in both figures the point estimates derived from the 2001 census and NIDS survey are too high and thus do not portray the correct level of mortality. The reason why the method produced such a high estimate is probably due to a problem of consistency between the two surveys with a higher than expected proportion of mother/father not alive in the NIDS survey. The other important finding is that the method failed to produce a point estimate between the Community Survey and the 2008 NIDS survey. This again points to a problem of consistency in reporting between the two surveys.

When compared to our ASSA benchmark estimates the point estimates derived from year of death data are generally high. One major exception to this are the female point estimates obtained from the 1998 and 2003 synthetic surveys. Even after using data on orphanhood to distribute the deaths with missing years the point estimates still remain high. This is a bit of a surprise because we expected the point estimates from year of death data to be consistent with the ASSA estimates because these point estimates are obtained using the NIDS survey only and are therefore not affected by the problem of consistency between surveys. The possible explanation for this is that the estimates are affected by the poor quality of these data. The synthetic orphanhood method which uses data from age 20 of respondents is sensitive to reporting errors and because year of death data contained a high proportion of missing data it has compromised the quality of estimates. We therefore accept the point estimates from the conventional method as representing the correct level of mortality.

Figure 4.11 Paternal synthetic cohort estimates produced using orphanhood data and year of death data

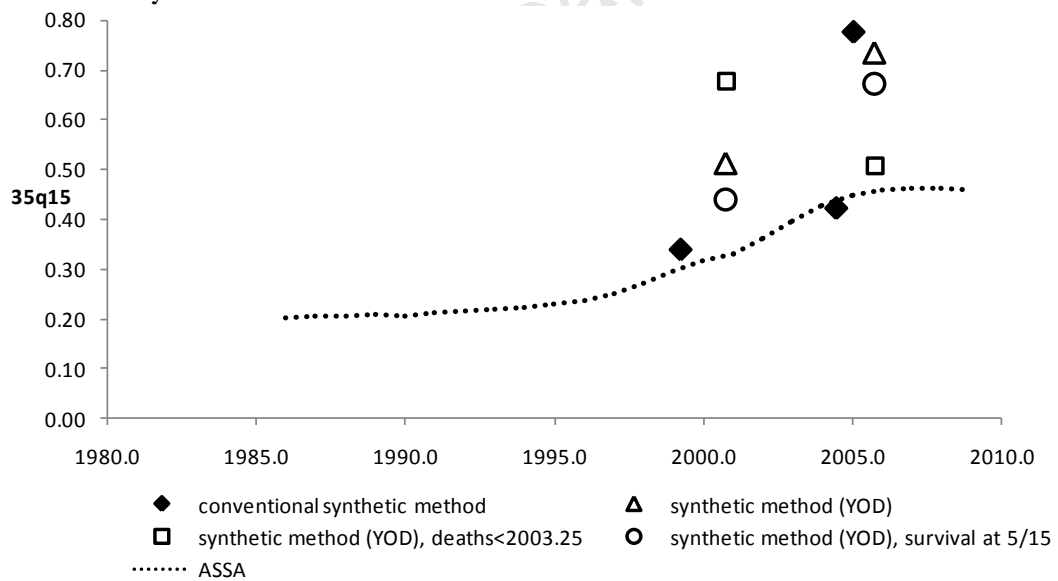
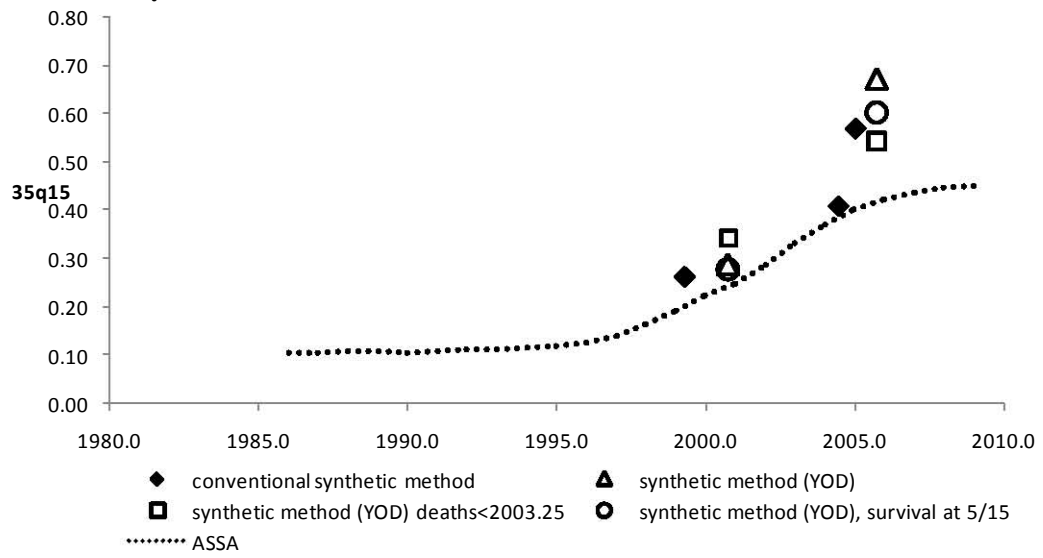


Figure 4.12 Maternal synthetic cohort estimates produced using orphanhood data and year of death data



5.1 Introduction

The purpose of this research was twofold. The first objective was to investigate whether estimates of time location produced from using year of death data are more accurate than estimates produced by using the Brass and Bamgboye (1981) method. The second objective was to investigate whether estimates of mortality produced by using year of death data are any better than those produced by the conventional orphanhood method in terms of level and consistency. This chapter discusses the results obtained from the data analysis and looks at whether the two objectives were met. Section 5.2 looks at the problem of missing data encountered in the use of year of death data while section 5.3 discusses the estimates of time location in relation with estimates of mortality. Section 5.2 looks at the results produced by using the synthetic cohort method. Finally section 5.3 and 5.4 give the limitations of the study and the conclusions for the study respectively.

5.2 The problem of missing data

In order to achieve our objectives, data from the 2008 NIDS survey were used because a question on survival of parents and on year of death of the parent was asked in the survey. The conventional orphanhood method and the synthetic cohort method were applied to the 1996 and 2001 census, the 2007 Community survey and the 2008 NIDS survey so as to produce levels and trends of mortality in South Africa. Having obtained these levels and trends, synthetic surveys were created using year of death data as suggested by Chackiel and Orellana (1985). The research encountered a significant problem in using the year of death data, that is, what to do with the high levels of subject non-response to the question on year of death. One of the reasons why the use of year of death data has attracted criticism is the existence of high subject non-response. Chackiel and Orellana (1985) fail to mention this problem or what is to be done about it. As was pointed out in Chapter 3, the NIDS data contains a high percentage of subject non-response, 36 per cent for fathers and 32 per cent for mothers. These data cannot simply be ignored since ignoring them would result in an underestimation of mortality. Even if the percentage of non-responses is relatively low it is still important information that cannot be ignored. Timæus (1991a) acknowledges encountering this problem in East Africa and had to impute about 25 per cent of the data so as to be able to use the year of death data. Unfortunately, in his paper he does

not explain the details of the imputation method he used. In this research, three approaches of dealing with these missing data were considered. The first approach is to apportion deaths with missing year of death proportionately. The second approach is to simply distribute all deaths with missing year of death proportionately up to 2003.25, i.e. assuming that all deaths with missing year of death occurred before 2003.25. The third approach is to use data on orphanhood at ages 5 and 15 of the child to impute the year of death data. These three approaches allowed the year of death data to be used as suggested by Chackiel and Orellana (1985).

5.3 Estimates of time location and mortality

To assess the performance of using year of death data to estimate mortality we have to look at time reference points in conjunction with the mortality estimates that are produced. Having apportioned deaths of those without a year of death, the mean year of death was calculated for each of the age groups and compared to the time reference points produced using the theoretical method developed by Brass and Bamgboye (1981) to establish whether the empirical time reference points are closer to the benchmark estimates than those calculated theoretically. As we observed in the results chapter, time reference points calculated from year of death data are located in the more recent past than those calculated theoretically.

There are three reasons to doubt the use of the year of death data in estimating time reference points and in estimating mortality. The first reason is that estimates of female mortality derived from the 2008 NIDS survey when plotted with time reference points calculated from year of death data are located in the more recent past and are not consistent with mortality estimates from the ASSA model. On the other hand female mortality estimates plotted using theoretical time reference points are closer to the ASSA estimates implying that time location calculated theoretically is closer to reality than that calculated using year of death data. However, what is of interest is the fact that estimates of male mortality derived from the 2008 survey when plotted with the time reference points, are consistent with the ASSA model while estimates plotted using theoretical time reference points seem to overestimate mortality. What seems to be happening is that a portion of missing fathers were reported as not alive by the respondents while in actual fact they were alive at the time of the survey, and the shifting of the estimates to the more recent past by using year of death data has compensated for this bias. As we pointed out earlier in this research, the problem of

reporting missing fathers as not alive has been known to exist in African settings since the advent of the orphanhood method. With the development of society the problem might be less severe but it is still known to exist and in this research it seems to be having an effect on our estimates. This has resulted in male estimates having a better fit than female estimates. Another interesting finding with regards to estimates derived using the single survey orphanhood method in this research is the absence (or insignificant effect) of the adoption effect. The problem of respondents in young age groups mistaking their foster parents as their biological parents is a phenomenon known in African culture. Udjo (2005) identified the adoption effect as the reason why he produced low estimates of female mortality from the 1996 census. However, in this research there is no evidence of the adoption effect in the estimates of female mortality derived using the conventional orphanhood method. Although estimates that were derived from using the year of death data underestimate mortality (compared to our gold standard), evidence does not point to the adoption effect, but rather to a problem with the time location.

The second reason to doubt the use of year of death data is that theoretical and empirical time reference points are not significantly different for the oldest age groups of children for both males and females. Assuming that year of death data are not severely affected by reporting errors, time reference points calculated from year of death data are expected to stretch further into the past than theoretically calculated reference points (Chackiel and Orellana 1985). For instance Chackiel and Orellana observe that empirically calculated time stretched as much as 5.7 years further back than theoretical time reference points. In this research time reference points produced using year of death data corrected by apportionment do not stretch back any further than theoretical time reference points. The discrepancy between the time reference points seem to be exaggerated by the apportioning of such a high proportion of non-response to the question on the year of death of parent.

The third reason to doubt the use of these year of death data is that the mortality estimates produced by the two synthetic surveys for both males and females are not internally consistent and are not consistent with the ASSA estimates. The underestimation of mortality by the 1998 synthetic survey for both males and females

show that deaths that belong in the past have been distributed in the more recent past and there is thus a need to correct for this bias.

Assuming, instead, that all deaths of parents with missing year of death occurred before 2003.25 has the effect of shifting the mean year of death backward in time by an average of about one year for both males and females. In the case of fathers it is difficult to say that this improves the time location of the estimates since there are two forces at play, the bias caused by missing fathers who are reported as not alive and the bias caused by distributing death proportionately. However, what is clear from using this assumption is that the 2003 synthetic survey overestimates male mortality. However in the case of mothers the assumption clearly leads to an improvement in the time location of the mortality estimates from the 2008.25 survey. Estimates of female mortality produced using this assumption is better than those produced by simple apportionment in terms of level and internal consistency.

As was highlighted in the results section, imputing the year of death data by using data on the survival of parents at ages 5 and 15 leads to a significant improvement in the quality of estimates of time location and mortality for both males and females. Probably the most important contribution by the data on orphanhood at age 5 and 15 is that they have allowed the mean year of death calculated using year of death data to extend further back in time, with mean year of death calculated from using year of death data stretching a further 7 years for females and 3.3 years for males for the 1998.25 synthetic survey. It is possible that the data on orphanhood at ages 5 and 15 could have biased the estimates of time location back into the past, but that is unlikely to be the case since estimates of mortality that have been produced are internally consistent and consistent with the ASSA benchmark estimates. The data on survival at age 5 and 15 of the child reveal that if there is better quality data on year of death one can obtain quality estimates that might be even better than those produced using the standard methods.

5.4 Synthetic cohort estimates of maternal and paternal orphanhood

Judging from the estimates of mortality produced by Dorrington, Moultrie and Timæus(2004), the level of mortality portrayed by the estimates from year of death data is unrealistically high. It is clear from the high level of estimates that simple apportionment of deaths with unknown year of death leads to an exaggeration of mortality especially for the recent estimate. This is similar to the case of Burundi where

year of death data were used by Timæus (1991a) to estimate synthetic cohort estimates, where 25 per cent of the data on the year of death had to be estimated, this resulted in an upward bias of the recent estimate and as with this research male estimates are more affected than female estimates. Assuming instead that deaths without year of death occurred before 2003.25 results, as expected in lower estimates of mortality for the recent period 2003.25-2008.25. What is surprising is that the male estimate is lower than the female estimate. This probably indicates that of the respondents who did not know year of death, a high percentage of them did not lose their father within the five years before the survey. Results produced by using data on survival of parents when the child was age 5 and 15 are disappointing. Although there was an improvement of the estimate of female mortality over the 2008.25-2003.25 period, the estimates are still too high. The question that needs to be answered is why the single survey but not the synthetic cohort method produced reasonable results. The most likely explanation for this is as suggested by Timæus (1991b), that the synthetic cohort method is sensitive to errors in reporting of dates of orphanhood.

5.5 Limitations of the Study

The major limitation of this research is the high level of subject non-response to the question on year of death. This then forces us to estimate the year of death of the parents with missing year of death, which as we pointed out leads to a bias in estimates of time location and mortality. In addition to this problem the data might contain a high percentage of respondents who fail to recall accurately the year of death of parent especially if the parent died a long time ago, older respondents in particular may be prone to this problem of recalling. The accurate recording of these data might have been further compromised by a possible lack of training of field workers in recording the year of death of the parent. In addition to the above problems the research was also limited by the small sample size of the survey. Small sample sizes can give rise to national estimates which are not representative of reality and this might in turn result in incorrect estimates of mortality. This might help explain the inconsistencies observed in the synthetic cohort estimates. A further weakness in this research is the use of estimates derived from the ASSA model as benchmark. The ASSA model has its own weaknesses, which includes the fact that there is some uncertainty as to the exact level of HIV prevalence in South Africa, and thus using the model to validate the accuracy of mortality estimates in South Africa might lead to a wrong conclusion.

5.6 Conclusions and Scope for the Future

The overall conclusion from this research is that it is possible to produce estimates of mortality using year of death data but without quality data it is difficult to produce quality estimates of time location and mortality that can match or surpass those produced using the conventional orphanhood method. Year of death data gives us the possibility of documenting the history of mortality since the estimates produced can go back as far as 30 years prior to the survey, unfortunately the method is affected by a high degree of subject non-response which ends up distorting the estimates of mortality as was the case in this research and in research done in East Africa. The poor quality of estimates produced using year of death data in this research or in East Africa does not necessarily mean that the data are useless, there is still need for the question on year of death data to be tested in larger surveys like the Community Survey or census and also to test it in other countries in Sub-Saharan Africa. There is therefore a need for research into how best to treat these missing data, which would probably include researching into the distributions of year of death and the accuracy of the distributions of deaths by year of death of those reporting year of death and the level of non response that can be tolerated in estimating mortality using year of death data. There is also a need to investigate the advantages of including the question on whether the parent died in the last five years prior to a survey. This question would allow one to create synthetic survey data as at a point five years prior to the actual survey and thus the estimation of more recent mortality rates that are possible using only a single survey.

REFERENCE

- Actuarial Society of South Africa. 2005. *AIDS and Demography Model 2003*. Cape Town: Actuarial Society of South Africa. www.actuarialsociety.org.za.
- Bhat, P. N. M. 2004. *A Generalized Relational Model for Demographic Analysis*. Delhi: Institute of Economic Growth.
- Blacker, J. G. C. 1977. "The Estimation of Adult Mortality in Africa from Data on Orphanhood", *Population Studies* **31** (1):107-128.
- Bradshaw, D. and Timæus, I. 2006. "levels and Trends of Adult Mortality," Jamison, D. T., Feachem, R. G., Makgoba, M. W., Bos, E. R., Baingana, F. K., Holfman, K. J. and Rogo, K. O. *Disease and Mortality in Sub-Saharan Africa*. Washington DC: World Bank Publications, pp. 31-42.
- Brass, W. 1975. *Methods for estimating fertility and mortality from limited and defective data*. North Caroline: Caroline Population Center.
- Brass, W. and Bamgboye, E. A. 1981. *The time location of survivorship: estimates for maternal and paternal orphanhood and the ever-widowed*. Center for Population Studies Research Papers, 81-1. London: London School of Hygiene and Tropical Medicine.
- Brass, W. and Hill, K. 1973. "Estimating adult mortality from orphanhood," Paper presented at Proceedings of the International Population Conference Liege, 1973. International Union for the scientific study of population, Liege.
- Chackiel, J. and Orellana, H. 1985. "Adult female mortality trends from restropective questions about maternal orphanhood included in censuses and surveys," Paper presented at International Population Conference. Florence, 1985. International Union for the Scientific Study of Population, Liege.
- Dorrington, R. and Moultrie, T. A. 2009. *Demography: Analysis of the NIDS Wave 1 Dataset Discussion Paper No. 9*. Cape Town: National Income Dynamics Study.
- Dorrington, R., Moultrie, T. A. and Timæus, I. 2004. *Estimation of mortality using the South African Census 2001 data*. Cape Town: Center for Actuarial Research for Statistics South Africa.
- Feeney, G. 2001. "The Impact of HIV/AIDS on Adult Mortality in Zimbabwe", *Population and Development Review* **27**:771-780.
- Hill, K. and Trussell, J. 1977. "Further Developments in Indirect Mortality Estimation", *Population Studies* **31** (2):313-334.
- Luy, M. 2009. *Estimating Mortality Differentials in Developed Populations from Survey Information on Maternal and Paternal Orphanhood*. European Demographic Research Working Papers. Vienna: Vienna Institute of Demography.
- Moultrie, T. A. and Dorrington, R. 2004. *Estimation of fertility from the 2001 South Africa Census data*. Cape Town: Center for Actuarial Research for Statistics South Africa.
- Palloni, A., Massagli, M. and Marcotte, J. 1984. "Estimating Adult Mortality with Maternal Orphanhood Data: Analysis of Sensitivity of the Techniques", *Population Studies* **38** (2):255-279.
- Timæus, I. 1986. "An assessment of methods for estimating adult mortality from two sets of data on maternal orphanhood", *Demography* **23** (3):435-450.
- Timæus, I. 1991a. "Estimation of adult mortality from orphanhood before and since marriage", *Population Studies* **45** (3):455-472.
- Timæus, I. 1991b. "Estimation of Mortality from Orphanhood in Adulthood", *Population Association of America* **28** (2):213-227.
- Timæus, I. 1992. "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of the United Nations* **33**:47-63.
- Timæus, I. and Graham, W. 1989. *Measuring adult mortality in developing countries. A review and assessment of methods*. Planning Policy and Research Working Papers, WPS 155. Washington DC: The World Bank.
- Timæus, I. and Jasseh, M. 2004. "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys", *Population Association of America* **41** (4):757-772.

- Timæus, I. and Nunn, A. J. 1997. "Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method", *Health Transition Review* 7:23-43.
- Tollman, S. M., Kahn, K., Garenne, M. and Gear, J. S. S. 1999. "Reversal in mortality trends: evidence from the Agincourt field site, South Africa, 1992-1995", *AIDS* 13:1091-1097.
- Udjo, E. O. 2005. "An Examination of Recent Census and Survey Data on Mortality," Zuberi, T., Sibanda, A. and Udjo, E. O. *The Demography of South Africa*. New York: M.E. Sharpe, pp. 90-120.
- Udjo, E. O. 2007. *Estimating Trends in Orphanhood Among Younger Persons In the Era of HIV/AIDS in South Africa, 2001-2006*. Pretoria: University of South Africa.
- United Nations. 1983. *Manual X. Indirect techniques for demographic estimation*. New York: United Nations.
- United Nations, D. O. E. A. S. A., Population Division. 2009. *World Population Prospects: The 2008 Revision*. Accessed:
- United Nations Population Division. 2009. *World Population Prospects: The 2008 Revision. CD-ROM Edition - Comprehensive Dataset in Excel format*. New York: United Nations Publications, ST/ESA/SER.A/282.
- World Health Organisation. 2010. *Life tables for WHO Member States*. http://www.who.int/healthinfo/statistics/mortality_life_tables/en.htm. Accessed: 5 May 2011.
- Zlotnik, H. and Hill, K. 1981. "The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality", *Demography* 18:103-122.

APPENDIX A: TABLES

Table A.1 Interpolation table for estimating time location of estimates derived from information on survival of mothers and fathers

<i>Adjusted age</i>	<i>Standard function</i>	<i>Adjusted age</i>	<i>Standard function</i>	<i>Adjusted age</i>	<i>Standard function</i>	<i>Adjusted age</i>	<i>Standard function</i>
<i>N+M</i>	<i>f(N+M)</i>	<i>N+M</i>	<i>f(N+M)</i>	<i>N+M</i>	<i>f(N+M)</i>	<i>N+M</i>	<i>f(N+M)</i>
26	0.090	39	0.099	52	0.218	65	0.431
27	0.090	40	0.104	53	0.231	66	0.452
28	0.090	41	0.109	54	0.245	67	0.473
29	0.090	42	0.115	55	0.259	68	0.495
30	0.090	43	0.122	56	0.274	69	0.518
31	0.090	44	0.130	57	0.289	70	0.542
32	0.090	45	0.139	58	0.305	71	0.568
33	0.090	46	0.149	59	0.321	72	0.595
34	0.090	47	0.160	60	0.338	73	0.622
35	0.091	48	0.171	61	0.356	74	0.650
36	0.092	49	0.182	62	0.374	75	0.678
37	0.093	50	0.193	63	0.392		
38	0.095	51	0.205	64	0.411		

Table A.2 Estimates of male mortality, 35q15, derived using the conventional orphanhood method with reference points derived using the Brass and Bamgboye method

	<i>1996 Census</i>	<i>date</i>	<i>2001 Census</i>	<i>date</i>	<i>2007 CS</i>	<i>date</i>	<i>2008 NIDS</i>	<i>date</i>
5-9	0.236	1991.8	0.292	1996.8	0.434	2002.0	0.421	2003.2
10-14	0.209	1989.8	0.265	1994.7	0.412	1999.9	0.407	2001.0
15-19	0.197	1988.0	0.227	1992.9	0.304	1998.0	0.338	1999.0
20-24	0.206	1986.5	0.230	1991.4	0.253	1996.4	0.294	1997.5
25-29	0.213	1985.2	0.231	1990.0	0.239	1995.2	0.308	1996.0
30-34	0.205	1984.1	0.232	1988.9	0.225	1994.2	0.264	1994.4
35-39	0.192	1983.5	0.223	1988.0	0.218	1993.6	0.267	1994.3

Table A.3 Estimates of female mortality, 35q15, derived using the conventional orphanhood method with reference points derived using the Brass and Bamgboye method

	<i>1996 Census</i>	<i>date</i>	<i>2001 Census</i>	<i>date</i>	<i>2007 CS</i>	<i>date</i>	<i>2008 NIDS</i>	<i>date</i>
5-9	0.101	1993.3	0.170	1998.3	0.428	2003.6	0.438	2004.8
10-14	0.081	1991.1	0.128	1996.1	0.269	2001.3	0.317	2002.4
15-19	0.083	1989.1	0.113	1994.0	0.206	1999.2	0.252	2000.3
20-24	0.098	1987.3	0.113	1992.3	0.166	1997.4	0.218	1998.4
25-29	0.109	1985.8	0.120	1990.8	0.138	1996.0	0.175	1997.0
30-34	0.117	1984.7	0.125	1989.6	0.128	1994.9	0.140	1996.0
35-39	0.115	1983.9	0.126	1988.7	0.124	1994.1	0.129	1995.2
40-44	0.111	1983.4	0.117	1988.3	0.118	1993.6	0.102	1994.8
45-49	0.098	1983.5	0.105	1988.2	0.102	1993.6	0.147	1994.1

Table A.4 Estimates of male mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after distributing deaths with missing year of death proportionately)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.184	1995.0	0.256	1999.9	0.421	2004.6
10-14	0.139	1992.7	0.307	1998.6	0.407	2003.4
15-19	0.203	1991.4	0.231	1996.8	0.338	2001.8
20-24	0.196	1989.4	0.238	1994.7	0.294	2000.6
25-29	0.189	1987.1	0.235	1993.6	0.308	1998.6
30-34	0.163	1985.9	0.209	1991.3	0.264	1997.5
35-39	0.152	1983.4	0.159	1989.5	0.267	1994.3

Table A.5 Estimates of female mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after distributing deaths with missing year of death proportionately)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.058	1995.9	0.181	2000.3	0.438	2005.3
10-14	0.092	1992.7	0.139	2000.1	0.317	2004.3
15-19	0.092	1990.4	0.145	1997.9	0.252	2003.8
20-24	0.098	1991.2	0.126	1995.7	0.218	2002.2
25-29	0.093	1990.3	0.114	1995.1	0.175	2000.5
30-34	0.121	1987.3	0.108	1994.3	0.140	1999.5
35-39	0.076	1986.4	0.133	1991.8	0.129	1998.8
40-44	0.094	1985.6	0.087	1991.6	0.147	1996.1
45-49	0.064	1984.5	0.090	1989.7	0.102	1996.5

Table A.6 Estimates of male mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after assuming that deaths with missing year of death occurred before 2003.25)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.229	1995.0	0.353	1999.8	0.421	2003.8
10-14	0.173	1992.7	0.393	1998.5	0.407	2002.3
15-19	0.232	1991.4	0.271	1996.7	0.338	2000.8
20-24	0.221	1989.4	0.279	1994.8	0.294	1999.7
25-29	0.224	1987.1	0.269	1993.5	0.308	1997.8
30-34	0.205	1985.9	0.257	1991.2	0.264	1996.5
35-39	0.194	1983.4	0.212	1989.4	0.267	1993.4

Table A.7 Estimates of female mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after assuming that deaths with missing year of death occurred before 2003.25)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.073	1995.8	0.230	2000.2	0.438	2005.0
10-14	0.105	1992.7	0.173	2000.0	0.317	2003.6
15-19	0.101	1990.5	0.166	1997.8	0.252	2003.2
20-24	0.115	1991.2	0.139	1995.8	0.218	2001.6
25-29	0.107	1990.3	0.134	1995.0	0.175	2000.0
30-34	0.137	1987.4	0.126	1994.4	0.140	1998.4
35-39	0.094	1986.4	0.153	1991.8	0.129	1997.9
40-44	0.113	1985.5	0.111	1991.5	0.147	1995.2
45-49	0.089	1984.5	0.112	1989.5	0.102	1994.9

Table A.8 Estimates of male mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after using data on orphanhood at age 5 and 15 to distribute the deaths with missing year of death)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.232	1994.7	0.289	1999.6	0.421	2004.6
10-14	0.162	1991.7	0.331	1997.9	0.407	2002.9
15-19	0.240	1990.7	0.248	1995.9	0.338	2001.2
20-24	0.223	1988.4	0.252	1993.7	0.294	1999.8
25-29	0.217	1986.2	0.242	1992.3	0.308	1997.3
30-34	0.186	1984.2	0.225	1989.9	0.264	1996.3
35-39	0.181	1982.2	0.173	1987.6	0.267	1993.1

Table A.9 Estimates of female mortality, 35q15, derived using year of death data with reference points derived using the year of death data (after using data on orphanhood at age 5 and 15 to distribute the deaths with missing year of death)

	1998		2003		2008 NIDS	
	<i>Synthetic Survey</i>	<i>date</i>	<i>Synthetic Survey</i>	<i>date</i>	<i>Survey</i>	<i>date</i>
5-9	0.093	1995.1	0.211	1999.9	0.438	2005.3
10-14	0.113	1992.1	0.152	1999.2	0.317	2003.8
15-19	0.102	1989.7	0.157	1997.1	0.252	2003.3
20-24	0.109	1990.1	0.134	1995.1	0.218	2001.5
25-29	0.113	1987.1	0.123	1994.1	0.175	1999.9
30-34	0.142	1985.4	0.116	1991.4	0.140	1998.4
35-39	0.102	1983.9	0.145	1989.8	0.129	1996.4
40-44	0.121	1982.2	0.099	1988.6	0.147	1994.1
45-49	0.108	1979.6	0.108	1986.2	0.102	1993.5

Stata code for imputing year of death data using data on survival of parent at age 5 and 15 of the respondent

1. Stata code for fathers

```
// w1_a stands for wave 1 of the NIDS survey, adults
// w1_a_dob_y stands for year of birth
// w1_a_fthali stands for father alive
// w1_a_fthdthy stands for father's year of death
// w1_a_fthdth15 stands for father dead when responded was aged 15
// w1_a_intrv_y stands for year of interview

// this line generates a variable which represents year of death of the father given that
// the respondent didnt answer the question "alive at age 5 or age 15"
gen age_adult_missing_y= w1_a_dob_y+int(( w1_a_intrv_y-
w1_a_dob_y+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdthy>2008 &
w1_a_fthdth15!=1 & w1_a_fthdth15!=2 & w1_a_fthdth5!=1 &
w1_a_fthdth5!=2

// this line generates a variable which represents year of death of the father given that
// the respondent answered that father was not alive at age 5
gen age_5only_missing_y= w1_a_dob_y+int(( w1_a_dob_y+5-
w1_a_dob_y+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdthy>2008 &
w1_a_fthdth5==1

// this line generates a variable which represents year of death of the father given that
// the respondent answered that father was alive at age 5 but did not respond
// for age 15
gen age_after5_missing15_missing_y= w1_a_dob_y+5+int(( w1_a_intrv_y-
w1_a_dob_y-5+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdthy>2008 &
w1_a_fthdth5==2 & w1_a_fthdth15!=1 & w1_a_fthdth15!=2

// this line generates a variable which represents year of death of the father given that
// the respondent answered that father was not alive at age 15 but did not respond
// for age 5
gen age_missing5_yes15_missing_y= w1_a_dob_y+int(( w1_a_dob_y+15-
w1_a_dob_y+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdthy>2008 &
w1_a_fthdth15==1 & w1_a_fthdth5!=1 & w1_a_fthdth5!=2

// this line generates a variable which represents year of death of the father given that
// the respondent answered that father was alive at age 5 but was not alive at age
// 15
gen age_5to15_missing_y= w1_a_dob_y+5+int(( w1_a_dob_y+15- w1_a_dob_y-
5+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdthy>2008 &
w1_a_fthdth15==1 & w1_a_fthdth5==2
```

```

/ this line generates a variable which represents year of death of the father given that
the respondent answered that father was alive at age 15
gen age_after15_missing_y= w1_a_dob_y+15+int(( w1_a_intrv_y- w1_a_dob_y-
15+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdth5>2008 &
w1_a_fthdth15==2

// this section of code deals with situations where the respondent answered by giving
an impossible year of death i.e greater than year of birth
gen implausible_y1=w1_a_dob_y+int(( w1_a_intrv_y- w1_a_dob_y+1)*runiform()) if
w1_a_fthali==2 & w1_a_fthdth15!=1 & w1_a_fthdth15!=2 &
w1_a_fthdth5!=1 & w1_a_fthdth5!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_fthdth5
gen implausible_y2= w1_a_dob_y+int(( w1_a_dob_y+5- w1_a_dob_y+1)*runiform())
if w1_a_fthali==2 & w1_a_fthdth5==1 & (2008-w1_a_best_age_yrs-
1)>w1_a_fthdth5
gen implausible_y3= w1_a_dob_y+5+int(( w1_a_intrv_y-w1_a_dob_y-
5+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdth15!=1 & w1_a_fthdth15!=2
& (2008-w1_a_best_age_yrs-1)>w1_a_fthdth5
gen implausible_y4= w1_a_dob_y+int(( w1_a_dob_y+15-
w1_a_dob_y+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdth15==1 &
w1_a_fthdth5!=1 & w1_a_fthdth5!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_fthdth5
gen implausible_y5= w1_a_dob_y+5+int(( w1_a_dob_y+15- w1_a_dob_y-
5+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdth15==1 &
w1_a_fthdth5==2 & (2008-w1_a_best_age_yrs-1)>w1_a_fthdth5
gen implausible_y6= w1_a_dob_y+15+int(( w1_a_intrv_y- w1_a_dob_y-
15+1)*runiform()) if w1_a_fthali==2 & w1_a_fthdth15==2 & (2008-
w1_a_best_age_yrs-1)>w1_a_fthdth5

// this section is where i replace the missing years of death with the ones i generated
from above
gentaku_adult_fathers= w1_a_fthdth5
replace taku_adult_fathers= age_adult_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth15!=1 & w1_a_fthdth15!=2 &
w1_a_fthdth5!=1 & w1_a_fthdth5!=2
replace taku_adult_fathers= age_5only_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth5==1
replace taku_adult_fathers= age_after5_missing15_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth5==2 & w1_a_fthdth15!=1 &
w1_a_fthdth15!=2
replace taku_adult_fathers= age_missing5_yes15_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth15==1 & w1_a_fthdth5!=1 &
w1_a_fthdth5!=2
replace taku_adult_fathers= age_5to15_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth15==1 & w1_a_fthdth5==2
replace taku_adult_fathers= age_after15_missing_y if w1_a_fthali==2 &
w1_a_fthdth5>2008 & w1_a_fthdth15==2

replace taku_adult_fathers= implausible_y1 if w1_a_fthali==2 & w1_a_fthdth15!=1 &
w1_a_fthdth15!=2 & w1_a_fthdth5!=1 & w1_a_fthdth5!=2 & (2008-
w1_a_best_age_yrs-1)>w1_a_fthdth5

```

```

replace taku_adult_fathers= implausible_y2 if w1_a_fthali==2 & w1_a_fthdth5==1 &
(2008-w1_a_best_age_yrs-1)>w1_a_fthdthy
replace taku_adult_fathers=implausible_y3 if w1_a_fthali==2 & w1_a_fthdth5==2 &
w1_a_fthdth15!=1 & w1_a_fthdth15!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_fthdthy
replace taku_adult_fathers=implausible_y4 if w1_a_fthali==2 & w1_a_fthdth15==1
& w1_a_fthdth5!=1 & w1_a_fthdth5!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_fthdthy
replace taku_adult_fathers=implausible_y5 if w1_a_fthali==2 & w1_a_fthdth15==1
& w1_a_fthdth5==2 & (2008-w1_a_best_age_yrs-1)>w1_a_fthdthy
replace taku_adult_fathers=implausible_y6 if w1_a_fthali==2 & w1_a_fthdth15==2
& (2008-w1_a_best_age_yrs-1)>w1_a_fthdthy

// this code deals with situations where the respondent answered by giving an
impossible year of death i.e greater than year of birth
gen implausible_y7 =(w1_a_intrv_y-w1_a_best_age_yrs-1)+int((w1_a_intrv_y-
w1_a_intrv_y+w1_a_best_age_yrs+1+1)/uniform()) if
taku_adult_fathers<(w1_a_intrv_y-w1_a_best_age_yrs-1)

gentaku_father_a= taku_adult_fathers
replacetaku_father_a=implausible_y7 if taku_adult_fathers<(w1_a_intrv_y-
w1_a_best_age_yrs-1)

collapse (sum) w1_wgt, by( w1_a_best_age_yrs w1_a_fthali w1_a_fthdthy
w1_a_fthdth15 w1_a_fthdth5 age_adult_missing_y age_5only_missing_y
age_after5_missing15_missing_y age_missing5_yes15_missing_y
age_5to15_missing_y age_after15_missing_y taku_father_a)
outsheet w1_a_best_age_yrs taku_father_a w1_wgt if w1_a_fthali==2 using
"F:\taku_a_fathers.csv", c

// uniformly distributes between DOS and DOB if respondent didn't answer the
question on alive at age 5
genage_child_missing_y= w1_c_dob_y+int(( w1_c_intrv_y-
w1_c_dob_y+1)*runiform()) if w1_c_fthali==2 & w1_c_fthdthy>2008 &
w1_c_fthdth5!=1 & w1_c_fthdth5!=2

// uniformly distributes between DOS and DOB if respondent is less than 5 years old
genage_child_under_missing_y= w1_c_dob_y+int(( w1_c_intrv_y-
w1_c_dob_y+1)*runiform()) if w1_c_fthali==2 & w1_c_fthdthy>2008 &
w1_c_best_age_yrs<5

// uniformly distributes between DOB+5 and DOB if respondent answered that father
was not alive by age 5
gen age_5only_missing_y= w1_c_dob_y+int(( w1_c_dob_y+5-
w1_c_dob_y+1)*runiform()) if w1_c_fthali==2 & w1_c_fthdthy>2008 &
w1_c_fthdth5==1 & w1_c_best_age_yrs>4

// uniformly distributes between DOS and DOB+5 if father was alive at age 5 and
respondent was greater than age 5

```

```

gen age_5after_missing_y = w1_c_dob_y+5+int(( w1_c_intrv_y - w1_c_dob_y-
5+1)*runiform()) if w1_c_fthali==2 & w1_c_fthdthy>2008 & w1_c_fthdth5==2 &
w1_c_best_age_yrs>4

// this section deals with implausible years of death
gen implausible_y1 = w1_c_dob_y+int(( w1_c_intrv_y- w1_c_dob_y+1)*runiform()) if
w1_c_fthali==2 & w1_c_fthdth5!=1 & w1_c_fthdth5!=2 &int(2008-
w1_c_best_age_yrs-1)>w1_c_fthdthy
gen implausible_y2 = w1_c_dob_y+int(( w1_c_intrv_y- w1_c_dob_y+1)*runiform()) if
w1_c_fthali==2 & w1_c_best_age_yrs<5 &int(2008-w1_c_best_age_yrs-
1)>w1_c_fthdthy
gen implausible_y3 = w1_c_dob_y+int(( w1_c_dob_y+5- w1_c_dob_y+1)*runiform())
if w1_c_fthali==2 & w1_c_fthdth5==1 & w1_c_best_age_yrs>4 &int(2008-
w1_c_best_age_yrs-1)>w1_c_fthdthy
gen implausible_y4 = w1_c_dob_y+5+int(( w1_c_intrv_y - w1_c_dob_y-
5+1)*runiform()) if w1_c_fthali==2 & w1_c_fthdth5==2 & w1_c_best_age_yrs>4
&int(2008-w1_c_best_age_yrs-1)>w1_c_fthdthy

gentaku_fathers_c = w1_c_fthdthy
replacetaku_fathers_c = age_child_missing_y if w1_c_fthali==2 & w1_c_fthdthy>2008
& w1_c_fthdth5!=1 & w1_c_fthdth5!=2
replacetaku_fathers_c = age_child_under_missing_y if w1_c_fthali==2 &
w1_c_fthdthy>2008 & w1_c_best_age_yrs<5
replacetaku_fathers_c = age_5only_missing_y if w1_c_fthali==2 & w1_c_fthdthy>2008
& w1_c_fthdth5==1 & w1_c_best_age_yrs>4
replacetaku_fathers_c = age_5after_missing_y if w1_c_fthali==2 & w1_c_fthdthy>2008
& w1_c_fthdth5==2 & w1_c_best_age_yrs>4

replacetaku_fathers_c = implausible_y1 if w1_c_fthali==2 & w1_c_fthdth5!=1 &
w1_c_fthdth5!=2 &int(2008-w1_c_best_age_yrs-1)>w1_c_fthdthy
replacetaku_fathers_c = implausible_y2 if w1_c_fthali==2 & w1_c_best_age_yrs<5
&int(2008-w1_c_best_age_yrs-1)>w1_c_fthdthy
replacetaku_fathers_c = implausible_y3 if w1_c_fthali==2 & w1_c_fthdth5==1 &
w1_c_best_age_yrs>4 &int(2008-w1_c_best_age_yrs-1)>w1_c_fthdthy
replacetaku_fathers_c = implausible_y4 if w1_c_fthali==2 & w1_c_fthdth5==2 &
w1_c_best_age_yrs>4 &int(2008-w1_c_best_age_yrs-1)>w1_c_fthdthy

gen implausible_y5 = (w1_c_intrv_y-w1_c_best_age_yrs-1)+int((w1_c_intrv_y-
w1_c_intrv_y+w1_c_best_age_yrs+1+1)*unifom()) if taku_fathers_c<(w1_c_intrv_y-
w1_c_best_age_yrs-1)

gentaku_father_c = taku_fathers_c
replacetaku_father_c = implausible_y5 if taku_fathers_c<(w1_c_intrv_y-
w1_c_best_age_yrs-1)

collapse (sum) w1_wgt, by( w1_c_best_age_yrs w1_c_fthali w1_c_fththa
w1_c_fthdthy w1_c_fthdth5 age_child_missing_y age_child_under_missing_y
age_5only_missing_y age_5after_missing_y taku_father_c)
outsheet w1_c_best_age_yrs taku_father_c w1_wgt if w1_c_fthali==2 using
"F:\taku_C_fathers.csv", c

```

2. Stata code for mothers

```
// w1_a stands for wave 1 of the NIDS survey, adults
// w1_a_dob_y stands for year of birth
// w1_a_mthali stands for mother alive
// w1_a_mthdthy stands for mother's year of death
// w1_a_mthdth15 stands for mother dead when responded was aged 15
// w1_a_intrv_y stands for year of interview

// this line generates a variable which represents year of death of the mother given that
// the respondent didnt answer the question "alive at age 5 or age 15"
gen age_adult_missing_y= w1_a_dob_y+int(( w1_a_intrv_y-
w1_a_dob_y+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth15!=1 & w1_a_mthdth15!=2 & w1_a_mthdth5!=1 &
w1_a_mthdth5!=2

// this line generates a variable which represents year of death of the mother given that
// the respondent answered that mother was not alive at age 5
gen age_5only_missing_y= w1_a_dob_y+int(( w1_a_dob_y+5-
w1_a_dob_y+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth5==1

// this line generates a variable which represents year of death of the mother given that
// the respondent answered that mother was alive at age 5 but did not respond
// for age 15
gen age_after5_missing15_missing_y= w1_a_dob_y+5+int(( w1_a_intrv_y-
w1_a_dob_y-5+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth5==2 & w1_a_mthdth15!=1 & w1_a_mthdth15!=2

// this line generates a variable which represents year of death of the mother given that
// the respondent answered that mother was not alive at age 15 but did not
// respond for age 5
gen age_missing5_yes15_missing_y= w1_a_dob_y+int(( w1_a_dob_y+15-
w1_a_dob_y+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth15==1 & w1_a_mthdth5!=1 & w1_a_mthdth5!=2

// this line generates a variable which represents year of death of the mother given that
// the respondent answered that mother was alive at age 5 but was not alive at age
// 15
gen age_5to15_missing_y= w1_a_dob_y+5+int(( w1_a_dob_y+15- w1_a_dob_y-
5+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth15==1 & w1_a_mthdth5==2

// this line generates a variable which represents year of death of the mother given that
// the respondent answered that mother was alive at age 15
gen age_after15_missing_y= w1_a_dob_y+15+int(( w1_a_intrv_y- w1_a_dob_y-
15+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdthy>2008 &
w1_a_mthdth15==2
```

```

// this section of code deals with situations where the respondent answered by giving
// an impossible year of death i.e greater than year of birth
gen implausible_y1=w1_a_dob_y+int(( w1_a_intrv_y- w1_a_dob_y+1)*runiform()) if
w1_a_mthali==2 & w1_a_mthdth15!=1 & w1_a_mthdth15!=2 &
w1_a_mthdth5!=1& w1_a_mthdth5!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_mthdthy
gen implausible_y2= w1_a_dob_y+int(( w1_a_dob_y+5- w1_a_dob_y+1)*runiform())
if w1_a_mthali==2 & w1_a_mthdth5==1 & (2008-w1_a_best_age_yrs-
1)>w1_a_mthdthy
gen implausible_y3= w1_a_dob_y+5+int(( w1_a_intrv_y-w1_a_dob_y-
5+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdth15!=1 &
w1_a_mthdth15!=2 & (2008-w1_a_best_age_yrs-1)>w1_a_mthdthy
gen implausible_y4= w1_a_dob_y+int(( w1_a_dob_y+15-
w1_a_dob_y+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdth15==1 &
w1_a_mthdth5!=1 & w1_a_mthdth5!=2 & (2008-w1_a_best_age_yrs-
1)>w1_a_mthdthy
gen implausible_y5= w1_a_dob_y+5+int(( w1_a_dob_y+15- w1_a_dob_y-
5+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdth15==1 &
w1_a_mthdth5==2 & (2008-w1_a_best_age_yrs-1)>w1_a_mthdthy
gen implausible_y6= w1_a_dob_y+15+int(( w1_a_intrv_y- w1_a_dob_y-
15+1)*runiform()) if w1_a_mthali==2 & w1_a_mthdth15==2 & (2008-
w1_a_best_age_yrs-1)>w1_a_mthdthy

// this section is where i replace the missing years of death with the ones i generated
// from above
gentaku_adult_mothers= w1_a_mthdthy
replace taku_adult_mothers= age_adult_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth15!=1& w1_a_mthdth15!=2 &
w1_a_mthdth5!=1& w1_a_mthdth5!=2
replace taku_adult_mothers= age_5only_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth5==1
replace taku_adult_mothers= age_after5_missing15_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth5==2 & w1_a_mthdth15!=1 &
w1_a_mthdth15!=2
replace taku_adult_mothers= age_missing5_yes15_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth15==1 & w1_a_mthdth5!=1 &
w1_a_mthdth5!=2
replace taku_adult_mothers= age_5to15_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth15==1 & w1_a_mthdth5==2
replace taku_adult_mothers= age_after15_missing_y if w1_a_mthali==2 &
w1_a_mthdthy>2008 & w1_a_mthdth15==2

replace taku_adult_mothers= implausible_y1 if w1_a_mthali==2 &
w1_a_mthdth15!=1& w1_a_mthdth15!=2 & w1_a_mthdth5!=1&
w1_a_mthdth5!=2 & (2008-w1_a_best_age_yrs-1)>w1_a_mthdthy
replace taku_adult_mothers= implausible_y2 if w1_a_mthali==2 &
w1_a_mthdth5==1 & (2008-w1_a_best_age_yrs-1)>w1_a_mthdthy
replace taku_adult_mothers=implausible_y3 if w1_a_mthali==2 &
w1_a_mthdth5==2 & w1_a_mthdth15!=1 & w1_a_mthdth15!=2 & (2008-
w1_a_best_age_yrs-1)>w1_a_mthdthy

```

```

replace taku_adult_mothers=implausible_y4 if w1_a_mthali==2 &
    w1_a_mthdth15==1 & w1_a_mthdth5!=1 & w1_a_mthdth5!=2 & (2008-
    w1_a_best_age_yrs-1)>w1_a_mthdthy
replace taku_adult_mothers=implausible_y5 if w1_a_mthali==2 &
    w1_a_mthdth15==1 & w1_a_mthdth5==2 & (2008-w1_a_best_age_yrs-
    1)>w1_a_mthdthy
replace taku_adult_mothers=implausible_y6 if w1_a_mthali==2 &
    w1_a_mthdth15==2 & (2008-w1_a_best_age_yrs-1)>w1_a_mthdthy

// this code deals with situations where the respondent answered by giving an
// impossible year of death i.e greater than year of birth
gen implausible_y7 =(w1_a_intrv_y-w1_a_best_age_yrs-1)+int((w1_a_intrv_y-
w1_a_intrv_y+w1_a_best_age_yrs+1+1)/uniform()) if
taku_adult_mothers<(w1_a_intrv_y-w1_a_best_age_yrs-1)

gentaku_mother_a= taku_adult_mothers
replacetaku_mother_a=implausible_y7 if taku_adult_mothers<(w1_a_intrv_y-
w1_a_best_age_yrs-1)

collapse (sum) w1_wgt, by( w1_a_best_age_yrs w1_a_mthali w1_a_mthdthy
    w1_a_mthdth15 w1_a_mthdth5 age_adult_missing_y age_5only_missing_y
    age_after5_missing15_missing_y age_missing5_yes15_missing_y
    age_5to15_missing_y age_after15_missing_y taku_mother_a)
outsheet w1_a_best_age_yrs taku_mother_a w1_wgt if w1_a_mthali==2 using
"F:\taku_a_mothers.csv", c

genage_child_missing_y= w1_c_dob_y+int(( w1_c_intrv_y-
w1_c_dob_y+1)*runiform()) if w1_c_mthali==2 & w1_c_mthdthy>2008 &
w1_c_mthdth5!=1 & w1_c_mthdth5!=2
genage_child_under_missing_y= w1_c_dob_y+int(( w1_c_intrv_y-
w1_c_dob_y+1)*runiform()) if w1_c_mthali==2 & w1_c_mthdthy>2008 &
w1_c_best_age_yrs<5
gen age_5only_missing_y= w1_c_dob_y+int(( w1_c_dob_y+5-
w1_c_dob_y+1)*runiform()) if w1_c_mthali==2 & w1_c_mthdthy>2008 &
w1_c_mthdth5==1 & w1_c_best_age_yrs>4
gen age_5after_missing_y= w1_c_dob_y+5+int(( w1_c_intrv_y - w1_c_dob_y-
5+1)*runiform()) if w1_c_mthali==2 & w1_c_mthdthy>2008 & w1_c_mthdth5==2
& w1_c_best_age_yrs>4

gen implausible_y1 = w1_c_dob_y+int(( w1_c_intrv_y- w1_c_dob_y+1)*runiform()) if
w1_c_mthali==2 & w1_c_mthdth5!=1 & w1_c_mthdth5!=2 & (2008-
w1_c_best_age_yrs-1)>w1_c_mthdthy
gen implausible_y2 = w1_c_dob_y+int(( w1_c_intrv_y- w1_c_dob_y+1)*runiform()) if
w1_c_mthali==2 & w1_c_best_age_yrs<5 & (2008-w1_c_best_age_yrs-
1)>w1_c_mthdthy
gen implausible_y3 = w1_c_dob_y+int(( w1_c_dob_y+5- w1_c_dob_y+1)*runiform())
if w1_c_mthali==2 & w1_c_mthdth5==1 & w1_c_best_age_yrs>4 & (2008-
w1_c_best_age_yrs-1)>w1_c_mthdthy

```

```
gen implausible_y4 = w1_c_dob_y+5+int(( w1_c_intrv_y - w1_c_dob_y-
5+1)*runiform()) if w1_c_mthali==2 & w1_c_mthdth5==2 & w1_c_best_age_yrs>4
& (2008-w1_c_best_age_yrs-1)>w1_c_mthdthy
```

```
gentaku_mothers_c= w1_c_mthdthy
replacetaku_mothers_c=age_child_missing_y if w1_c_mthali==2 &
w1_c_mthdthy>2008 & w1_c_mthdth5!=1 & w1_c_mthdth5!=2
replacetaku_mothers_c=age_child_under_missing_y if w1_c_mthali==2 &
w1_c_mthdthy>2008 & w1_c_best_age_yrs<5
replacetaku_mothers_c=age_5only_missing_y if w1_c_mthali==2 &
w1_c_mthdthy>2008 & w1_c_mthdth5==1 & w1_c_best_age_yrs>4
replacetaku_mothers_c=age_5after_missing_y if w1_c_mthali==2 &
w1_c_mthdthy>2008 & w1_c_mthdth5==2 & w1_c_best_age_yrs>4
```

```
replacetaku_mothers_c = implausible_y1 if w1_c_mthali==2 & w1_c_mthdth5!=1 &
w1_c_mthdth5!=2 & (2008-w1_c_best_age_yrs-1)>w1_c_mthdthy
replacetaku_mothers_c = implausible_y2 if w1_c_mthali==2 & w1_c_best_age_yrs<5
& (2008-w1_c_best_age_yrs-1)>w1_c_mthdthy
replacetaku_mothers_c = implausible_y3 if w1_c_mthali==2 & w1_c_mthdth5==1 &
w1_c_best_age_yrs>4 & (2008-w1_c_best_age_yrs-1)>w1_c_mthdthy
replacetaku_mothers_c = implausible_y4 if w1_c_mthali==2 & w1_c_mthdth5==2
& w1_c_best_age_yrs>4 & (2008-w1_c_best_age_yrs-1)>w1_c_mthdthy
```

```
gen implausible_y5 =(w1_c_intrv_y-w1_c_best_age_yrs-1)+int((w1_c_intrv_y -
w1_c_intrv_y+w1_c_best_age_yrs+1+1)*uniform()) if
taku_mothers_c<(w1_c_intrv_y-w1_c_best_age_yrs-1)
```

```
gentaku_mother_c= taku_mothers_c
replacetaku_mother_c=implausible_y5 if taku_mothers_c<(w1_c_intrv_y -
w1_c_best_age_yrs-1)
```

```
collapse (sum) w1_wgt, by( w1_c_best_age_yrs w1_c_mthali w1_c_mthdthy
w1_c_mthdth5 age_child_missing_y age_child_under_missing_y age_5only_missing_y
age_5after_missing_y taku_mother_c)
outsheet w1_c_best_age_yrs taku_mother_c w1_wgt if w1_c_mthali==2 using
"F:\taku_C_mothers.csv", c
```

