INVESTIGATING BIASES IN CENSUS QUESTIONS ON MORTALITY USING AGINCOURT HEALTH AND DEMOGRAPHIC SURVEILLANCE SYSTEM DATA

Polite Munyaradzi Nduru
Centre for Actuarial Research

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

November 2013

PLAGIARISM DECLARATION

This research is my original work, produced with supervisory assistance from my supervisor. I have used the Harvard convention for citation and referencing. Each contribution to this dissertation from the works of other people has been acknowledged, cited and referenced. In addition, this dissertation has not been submitted for any academic or examination purpose to any other university.

----------- Signature

----------- Date

ABSTRACT

An understanding of the errors found in census questions is important in order to assess the level of confidence in the census data and to get an appreciation of the impact of using these data in estimating mortality derived from census data. While demographic methods are often used to determine the direction of bias in the data, direct evaluation studies are required to determine the nature and extent of biases with more accuracy. Equally important is an understanding of the characteristics of the respondents who produce better responses in censuses. This can be used to inform selection of better respondents in order to improve the quality of the collected data.

This research uses census data from a survey, which is matched to the longitudinal Health Demographic Surveillance System site (HDSS) data from Agincourt, Limpopo Province, South Africa, in order to assess the biases found in data used in child and adult mortality estimation that uses indirect techniques. The research also assesses the accuracy of data used to measure mortality directly, and assesses the potential of additional questions aimed at cause of death analysis and completeness of death registration in censuses. This is done by comparing the census information to the true record as captured by the HDSS.

The results from this research show that indirect child mortality estimates are less robust than indirect adult mortality estimates because of the relatively poor quality of responses to the children ever born or children surviving (CEB/CS) questions compared to the orphanhood questions. Questions on household deaths in the year before the survey and questions on registered deaths have the potential to allow for the estimation of recent mortality levels at all ages and the completeness of death registration. However**,** there is a need to train the census enumerators to be able to elicit accurate information, in particular on ages at death. In addition, young (aged 15-29) adult female respondents generally give more accurate responses to all the questions in the census.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Dorrington, for his patience and invaluable guidance, without which this thesis would not have been possible. I would also like to thank Associate Professor Moultrie for enlightening me and answering my numerous questions about the UCT survey and the Agincourt HDSS data. I would like to express my gratitude to Zerina Matthews for her administrative support she offered throughout my studies and stay in Cape Town.

I would like to thank my family - my wife, Merlary, and my daughter, Stacey - for their patience, love and support. A special mention goes to my fellow classmates for their encouragement and moral support. I would also like to thank Tapfuma Pashapa for his assistance.

Finally, I am grateful to the Hewlett Foundation, the University of Cape Town Postgraduate Funding Office and the Centre for Actuarial Research for funding my studies and stay in Cape Town

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

Acronym	Meaning
AHDSS	Agincourt Health and Demographic Surveillance Site
AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ART	Anti-Retroviral Therapy
ASSA	Actuarial Society of South Africa
BIC	Bayesian Information Criterion
CEB/CS	Children Ever-Born/Children Surviving
DHS	Demographic and Health Surveys
DSS	Demographic Surveillance Site
EA	Enumeration Area
HDSS	Health and Demographic Surveillance Site
HHID	Household Identification Number
HIV	Human Immunodeficiency Virus
IMR	Infant Mortality Rate
INDEPTH	International Network for the Demographic Evaluation of Populations and Their Health
MDG	Millennium Development Goal
OCR	Optical Character Recognition
PES	Post Enumeration Survey
UCT	University of Cape Town
PMTCT	Prevention of Mother-to-Child Transmission
UN	United Nations
WFS	World Fertility Survey
WHO	World Health Organisation

1 INTRODUCTION

1.1 Background

The accurate measurement and estimation of mortality levels, trends, causes, and differentials is important in informing health care programmes and policy formulation in any population, particularly for developing countries because they have limited resources and experience higher mortality than developed countries. Complete vital registration systems are a preferred source of mortality data because they provide data that can be considered current. However, in many developing countries, retrospective measurement in censuses supplemented by other sources is the principal vehicle for data collection used for tracking demographic evolution and used as the basis for policy formulation in sub-Saharan Africa (Mathers and Boerma 2010, Blacker 2004, Timæus 1993) and this is likely to be the case for a number of years to come (Hill 2006).

In the absence of complete and timely vital registration systems, the incorporation of the so-called Brass-type questions in censuses that ask about the survival of relatives has been used as a substitute for collecting data on mortality (Hill 1991). The use of the Brass questions in censuses, which are usually short and simple, has enabled the estimation of both child and adult mortality levels and trends using indirect demographic techniques (Reniers, Masquelier and Gerland 2011). The mortality estimates derived using indirect techniques are usually not current; although recent/current estimates can be derived by including questions on survival of most recent birth(s) and the number of household deaths in the past calendar year or another reference period (Reniers, Masquelier and Gerland 2011). The Brass-type questions on child mortality solicit complete or summary birth histories, while for adult mortality the questions ask about the survival of close relatives, such as parents, spouses or siblings. As part of the census, other geo-socio-economic information is collected permitting detailed analysis of mortality by geographic and socio-economic differentials (Queiroz 2011, Stanton, Hobcraft, Hill *et al.* 2001, Hill 1991).

The United Nations (UN) has for many years been recommending the inclusion of questions in censuses that capture useful information on mortality in countries without functional vital registration systems. However, because the data are retrospective in nature, the data are vulnerable to reporting errors (Bangha, Diagne, Bawah *et al.* 2010, Blacker 2004, Timæus 1993). Typical potential sources of biases in these census questions include under-coverage of the universe, poor implementation of the census procedures, ambiguous question structure, respondents who cannot or do not provide accurate responses, interviewers who misunderstand the questions or how to record data, and coders who misunderstand the coding structure, as well as non-response (United Nations 2008a).

The quality of the data obtained in censuses is often assessed through internal consistency checks using age and sex distributions and sex ratios, or where more than two censuses exist, quality of the data is assessed through evaluation of inter-censal growth rates, cohort survival ratios and post-enumeration surveys (PES) (Moultrie 2013, Blacker and Brass 1993). Additionally, assessment of the quality of the data has been achieved through the evaluation of the plausibility of the estimates derived from it. Researchers acknowledge that direct evaluation studies at national and subnational levels are needed to understand better the quality of census data (Mathers and Boerma 2010, Phillips 1997).

This research is one such direct study, which seeks to provide better insight into the quality of the data collected using direct and indirect demographic questions. In addition, the research seeks to assess the performance of less traditional questions, such as cause of death questions, maternal mortality questions and questions aimed at estimating the completeness of death registration in censuses. The study makes use of matched census and longitudinal Health and Demographic Surveillance System (HDSS) site data.

1.2 Aims and Objectives

The purpose of this research is to assess the accuracy of the data obtained through retrospective census questions used to measure child and adult mortality, and the extent of bias in the estimates produced by these questions in comparison to published estimates, in particular those using the AHDSS. Specifically, the research seeks to meet the following objectives:

1. to assess the population numbers by age and sex as captured through the census questions;
2. to assess the errors in the data gathered through censuses used to measure child and adult mortality using indirect demographic techniques, in particular the Children Ever Born/ Children Surviving and the Orphanhood methods;
3. to assess the errors in recent household deaths and the potential of asking additional questions aimed at determining cause of death and the extent of death registration;
4. to assess the child and adult mortality (both direct and indirect) estimates derived from the census data against published estimates; and
5. to investigate respondent characteristics that have a bearing on quality of responses to various demographic questions.

1.3 Structure of the dissertation

The next chapter will present a review of the literature relevant to this dissertation, starting with a review of the usefulness of the census questions, the methods for measuring childhood and adult mortality using single census data, the errors already identified to be in the census questions and sources of errors in censuses and/surveys. Chapter 2 also gives a brief description of the Agincourt Health and Demographic Surveillance System (AHDSS) site and the aims and objectives of the UCT survey. Chapter 3 details the methodology used to address the objectives of the research. Chapter 4 presents the results and the analysis of the results. Finally, Chapter 5 discusses the results, the limitations of the research, offers ideas for future research, and draws conclusions from the study.

2 LITERATURE REVIEW

This chapter examines the role that demographic questions asked in censuses play in the measurement of mortality in developing countries. The chapter reviews literature on sources of data errors in census data. It also reviews the indirect methods used to derive child and adult mortality estimates using single-census data and looks at some of the known issues with the mortality data collected through censuses. It concludes by looking at the background and objectives of the UCT survey and provides a brief background of the survey site and the role of the HDSS.

2.1 Usefulness of census questions

Periodic censuses are important for both developed and developing countries because they provide data that cannot be provided by other sources, such as the civil/vital registration system (Gil and Omaboe 1993, Stycos 1993, Hartmann 2009). Population censuses are considered a primary source of data on the population for the purposes of estimating birth and death rates together with age-sex specific fertility and mortality rates. Civil registration systems would provide numerators for the rates; however, in most developing countries, rates developed this way are often implausibly low and thus additional questions were introduced to get fertility and mortality information in censuses and surveys (United Nations 1989). Vital registration systems, censuses and surveys can be used as complementary sources of data (Mba 2006, United Nations 1989). Surveys, being relatively small compared to censuses, permit the collection of detailed and more accurate information. This is possible since enumerators are usually well trained and the resources for a small but nationally representative undertaking can reasonably and more easily be mobilised.

The majority of developing countries do not have vital registration systems that are complete, and are thus unusable for estimation of vital rates (Mathers and Boerma 2010, Garenne and Gakusi 2006). As of 2004, only 60 out of the 192 World Health Organization (WHO) member states reported death registration data that could be considered reasonably recent and essentially complete (coverage of at least 90% of the deaths) (Mathers and Boerma 2010). Mauritius, Seychelles, South Africa and Zimbabwe were the only countries out of 46 sub-Saharan countries that had usable vital registration data (Mathers, Fat, Inoue *et al.* 2005).

Whilst there have been little or no improvements in vital registration systems, demand for accurate demographic data has been on the increase (Mba 2006, United Nations 1989, Mathers and Boerma 2010). Researchers and policy makers have had to rely on censuses and nationally representative household sample surveys as principal methods of data collection (Mba 2006, Mathers and Boerma 2010, Coale, Kirk, Hauser *et al.* 1971, Cleland 1996, Blacker 2004, Timæus 1993). The coverage and accuracy of these data vary greatly over time and space (Hartmann 2009). In his paper, Hill (2006) notes that in the medium term, levels and trends in mortality for much of the world's population would have to be computed using indirect estimation techniques and data collected through the incorporation of relevant questions in censuses, supplemented by intensive, well designed demographic surveys. These surveys have taken the form of the World Fertility Survey (WFS), Demographic and Health Survey (DHS) and the Health and Demographic Surveillance System (HDSS), amongst others (Cleland 1996).

The development of indirect demographic estimation techniques based on responses to questions on the survival status of children, siblings and parents has led to an increase in the number of African countries adopting the so-called Brass questions in their censuses (Reniers, Masquelier and Gerland 2011). Table 2.1 shows that the majority of African countries rely on censuses or surveys as sources of mortality data and only a few rely on a vital registration system. In addition to the Brass-type questions, questions on household deaths are also increasingly becoming a source of direct mortality estimates and cause of death information.

Table 2.1 Number of African countries with data for estimating mortality by type of information and period (n=54)

Type of information collected	*1950–1959*	*1960–1969*	*1970–1979*	*1980–1989*	*1990–1999*	*2000 & later*
Infant and child mortality						
Children ever born/Children surviving	5	20	27	33	37	39
– Maternity histories	–	–	10	20	35	38
Adult mortality						
– Household deaths	5	14	20	22	20	20
–Maternal orphanhood[1]	–	5	13	26	18	13
–Paternal orphanhood[1]	–	5	13	20	15	13
– Survival of siblings	–	–	–	–	23	32
– Widowhood	–	–	6	5	–	1
– Vital registration	11	15	14	10	13	11

[1.] *These exclude censuses or surveys that collected data on maternal and paternal survival for youngsters under age of 18.*

Source: Reniers, Masquelier and Gerland (2011:152)

2.2 Mortality estimation methods using data from a single census

2.2.1 Child mortality

Child mortality can be measured from a single census using either indirect estimation through the CEB/CS method; or directly using deaths reported by households. Indirect estimation methods, as the name suggests, use indirect information about reported numbers of children ever born and children surviving or dead to estimate mortality.

2.2.1.1 CEB/CS mortality estimation

Brass (1968) devised an ingenious method which converts the proportion of children ever born who have died, reported by women in five-year age groups, into life-table mortality measures without asking mothers detailed questions but by asking them two simple questions about the number of children ever borne and number of children still alive (Hill 2013). In deriving this method, Brass assumed that fertility and mortality rates were constant prior to the census/survey.

To understand how the method was derived, we adopt the notation used by Hartmann (2009). If we assume that mothers have a uniform age-distribution by single age and five-year age groups, that between ages α and *y* the women have constant fertility with respect to time, $f(a)$, and their new-borns experience a time invariant risk of dying before age *y*, $q_s(y)$, then the proportion of deceased children to be reported by women aged *y* can be derived from the following formula:

$$D_s(y) = \frac{\int_{\alpha}^{y} f(a) q_s(y-a)\,da}{\int_{\alpha}^{y} f(a)\,da}$$

Further, Brass assumed that the probability of dying before age *y*, $q(y) = 1 - l(y)$, at different levels of mortality are proportional to each other, implying that the following relationship holds:

$$q(y) = k * q_s(y) \qquad (2.1a)$$

where $q(y)$ is an estimate of the observed child mortality and $q_s(y)$ is conveniently chosen standard mortality.

Using the mean value theorem of integrals, it can be shown that there exists some age *x* with $0 < x < y - \alpha$ such that:

$$D_s(x) = q_s(x) \qquad (2.1b)$$

Equation 2.1b implies that there is an age *y* of mothers for which the proportion of dead children is the same as the probability of $q_s(x)$ for new-borns dying before age *x*. Combining 2.1a and 2.1b results in the following equation:

$$q(x) = k * D_s(x)$$

Brass conveniently chose *x* to be 1, 2, 3, 5, 10, 15 or 20 and used model specifications of the mortality and fertility to derive the multipliers (constant of proportionality for each age group of the mothers) that convert the reported proportion of children dead ($D(i)$) by contiguous five-year age groups of the mothers into a probability of dying before exact age *x*, $q(x)$ through the following relationship:

$$q(x) = k(i) * D(i)$$

The $k(i)$ multipliers represent a scaling factor in the *i-th* age group to compensate for non-mortality factors that determine the value of $D(i)$ and $i = 1,2,...,7$ correspond to the 15-19, 20-24,…,45-49 age groups of mothers (United Nations 1983). Using these multipliers results, for example, in the Infant Mortality Rate (IMR), $q(1)$, being derived from information obtained from the mothers in the 15-19 age group.

By making the above conversion the method makes an implicit assumption that the risk of a child dying is a function of the child's age and no other factors, such as the parity or age of the mother, and that the mortality of the mother is independent of the mortality of the child(ren) (United Nations 1990, United Nations 1983). It has been observed that mortality of the children of very young mothers is significantly higher than average due to higher levels of prematurity and low birth weight. In addition, it has been noted that women giving birth at very young ages are usually from lower social classes (Collumbien and Sloggett 2001, United Nations 1983, Feeney 1980). It is for these reasons that the mortality of children of women in this age group (and sometimes the next age group) should be excluded from consideration.

Brass noted that the most significant determinant of the $k(i)$ multipliers was the age distribution of fertility and thus based the derivation of multipliers on the ratio of the parity of the youngest women, $P(1)/P(2)$. These multipliers were initially refined first by Sullivan (1972) then further by Trussell (1975). The Trussell multipliers, which are the most widely used, employ a third degree polynomial using data from the Coale-Trussell (1974) fertility schedules in a linear regression on the parity ratios $P(1)/P(2)$ and $P(2)/P(3)$. The Trussell multipliers have the form:

$$k(i) = a(i) + b(i).\frac{P(1)}{P(2)} + c(i).\frac{P(2)}{P(3)}$$

where $a(i)$, $b(i)$ and $c(i)$ are coefficients derived from linear regression of mortality and fertility from projected stable populations.

The assumption of constant recent mortality made by Brass presented problems in situations where mortality was changing, which has been made especially true by HIV/AIDS. In such situations, mortality experienced by one-year olds at census date is likely to be significantly different from that experienced by one-year olds a decade before the census (UN 1983) (United Nations 1983). Feeney (1976) proffered a solution where mortality rates could be assumed to be changing linearly over time by giving time references to the mortality estimates by making the assumption that the change in mortality had been linear and constant. This enabled mortality trends to be established by converting the $q(x)$ mortality measures into a common measure, say $q(5)$.

Following on the earlier work of Feeney (1976), Coale and Trussell (1977) developed a procedure that could be used to express time location in terms of different levels of model life tables. They made an assumption that the rate of mortality change was fairly constant over time and devised an indicator of the number of years $t(x)$ before the census at which the mortality estimates $q(x)$ roughly applied. However, this has the disadvantage of smoothing the child mortality estimates over time. Coale and Trussell used the following equation:

$$t(x) = a(i) + b(i).\frac{P(1)}{P(2)} + c(i).\frac{P(2)}{P(3)}$$

where $a(i), b(i)$ and $c(i)$ are coefficients for each of Coale-Demeny life tables obtained through linear regression on the parity ratios from simulated fertility schedules.

Although the child mortality estimates derived using the CEB/CS method have provided reasonable mortality indicators in developing countries where the method is frequently used, the method may still be prone to errors. The errors in the mortality estimates arise due to errors in the data and/or where the assumptions on which the method is based are invalid (Ewbank 1982, Feeney 1980). There are key assumptions used in the CEB/CS method that may be violated and thus produce estimates that are implausible and inconsistent. First, the method assumes that fertility is constant for the recent period before the census (Adetunji 1996, United Nations 1983, Brass and Coale 1968). If fertility has been declining for the period immediately before the census, then the cross-sectional parity ratios $P(1)/P(2)$ and $P(2)/P(3)$ will tend to be small and result in overestimation of mortality levels, and also locate the estimates to periods that are too close to the census date. Thus the method is affected by changes in the age pattern of fertility but not the level of fertility associated with no changes in the age pattern of fertility (Adetunji 1996, United Nations 1983). However, Feeney (1980) points out that the biases in the estimates due to changes in fertility are unnoticeable unless the errors from other sources are extremely small.

Second, the method assumes mortality has been constant for the immediate period prior to the census. As noted above, this assumption was relaxed by Coale and Trussell (1977) to one in which the rate of mortality change was linear, thus permitting the determination of dates to which the mortality estimates pertain. This assumption would be violated when the mortality change is non-linear. There has not been empirical evidence to support the claim of linear mortality change in Africa thus the method might produce biased estimates in the event that mortality change has indeed been non-linear (Adetunji 1996).

The third assumption is that the population under investigation follows a mortality schedule that is similar to that of the model life table from which the multipliers as well as the time location constants are derived. There are concerns that the Coale and Demeny (1966) regional life table models may not be appropriate for many of the countries in Africa. This is particularly true in light of the HIV/AIDS epidemic in sub-Saharan Africa, which has resulted in mortality schedules that cannot be described by the Coale Demeny regional tables (Mahy 2003, Preston, Heuveline and Guillot 2001).

The fourth assumption which is also important to the CEB/CS method is that of independence between the mortality of the mother and the children. With the generalised HIV/AIDS epidemic in sub-Saharan Africa, where prevalence levels are high, the assumption of independent mortality between mothers and their children has been violated due to mother-to-child transmission (Ward and Zaba 2008, Ng'weshemi, Urassa, Isingo *et al.* 2003). The assumption of independence between mother and child mortality in essence suggests that mortality reported by surviving parents is not different from that of the children of dead mothers, which is unlikely as there is likely to be high mortality experienced by the children of HIV positive mothers (Brass 1996, Preston, Heuveline and Guillot 2001). Ward and Zaba (2008) found that a prevalence of 5-10% results in an error, which is at least 5% in the estimates of ${}_5q_0$.

In addition to the violation of the assumptions discussed above, implausible and inconsistent child mortality estimates might result due to data errors. Data errors that could bias the indirect mortality estimates include omission of live births, inclusion of stillbirths, omission of dead children and maternal age misreporting (Brass and Coale 1968, Adetunji 1996, Preston, Heuveline and Guillot 2001). The data used for the children ever born and children surviving (CEB/CS) child mortality estimation procedure typically suffer from a general understatement of children ever born for two reasons: first, omission of dead children; and second, reporting of children might be done by respondents on behalf of others and the respondent might not be aware of the dead children. It is also likely that there might be omission of children whose father is not the current one or children who have been adopted (Feeney 1976, United Nations 1989). These data errors affect the average parity and proportion of children dead by mother's age. Adetunji (1996) gives an example in which there is high correlation between the proportion of children dead in a maternal age and the indirect estimate of infant mortality from that age group. He notes that in such a situation, the errors in the proportion dead have a direct effect on the indirect mortality estimate calculated.

Omission of live births results in overstated child mortality. To reduce omission errors in censuses, additional questions which seek to probe the recall of distant demographic events are now almost universally included in the questionnaire. Rather than asking just one question, “How many children have you ever borne alive” this has been replaced by asking separate questions on children who are alive and living at home, children alive but living away from home, and children who have died (United Nations 2004, Cleland 1996). However, asking such questions in censuses might not be feasible due to time and cost constraints.

A downward bias in the number of children reported results in a downward bias in proportion of dead children. It is more likely that surviving children are under-reported through omission of the children than it is surviving children are over-reported. Over-reporting would occur for children living with foster parents when both the foster parents and biological parents report on the same children. Non-response may occur in both children ever born and children surviving questions. The significance of non-response errors depends on the level of non-response and on the correlation of non-response between the two questions (Feeney 1976, United Nations 1989).

CEB/CS data may also be affected by age misreporting and age exaggeration resulting in mis-alignment of children to the age group of their actual mothers. Age misreporting and age exaggeration in censuses can vary markedly depending, *inter alia*, on the respondent’s age (Timæus 1991c). Some people in developing countries do not know their ages and in their culture it may be of little relevance. Typically, age exaggeration is a more serious problem with older respondents than for younger and more numerate respondents (Timæus 1991c, Timæus 1991a, United Nations 1989).

2.2.1.2 *Ward and Zaba method for adjusting biases in CEB/CS mortality estimates due to HIV/AIDS*

The Ward and Zaba (2008) method seeks to adjust for the bias in the indirect child mortality estimates derived using the CEB/CS technique due to the effect of HIV/AIDS. Hallett, Gregson, Kurwa *et al*. (2010) note that the bias due to HIV in the indirect child mortality estimates could be important in cross-country comparisons. This is because the bias may vary according to the level of non-AIDS-related background mortality. In addition the bias might be affected by the cross country variation in the magnitude and stage of the epidemic of HIV infection. Hallett, Gregson, Kurwa *et al*. (2010) further note that the bias could also confound trends analyses because the magnitude may change as the epidemic evolves. The change in the magnitude is due to changes in the age pattern of HIV infection, the number of women with advanced disease, and the uptake of services for the prevention of mother-to-child transmission and for the delivery of anti-retroviral therapy (ART).

The method proposed by Ward and Zaba (2008) uses stable population modelling and simulation to assess and adjust for the bias in the estimates derived from the application of the CEB/CS technique in populations with HIV. The corrected estimates are derived from information on seroprevalence in the population on the assumption of population and epidemic stability. The technique is detailed below.

The unadjusted estimates are computed following the standard CEB/CS procedure. The adjusted estimates are then obtained from the following equation:

$$q(z)^t = q(z)^e + n(z),$$

where the $q(z)^t$ is the adjusted mortality estimate, $q(z)^e$ is from the normal application of the CEB/CS technique and $n(z)$ is the adjustment for HIV/AIDS bias. The adjustment n(z) equation is given as:

$$n(z) = aPREV + b(PREV)^2 + cPREV15$$

where PREV and PREV15 are HIV prevalences in women aged 15-49 and 15-19 respectively; and *a*, *b* and *c* are regression coeffecients derived from regressing the difference in the adjusted and unadjusted mortality estimates ($n(z)$) with HIV prevalence figures. Ward and Zaba assumed that there is epidemic stability in HIV prevalence. However, there is a challenge in the application of the method in that prevalence figures are not as stable as the method assumes. To circumvent this challenge, Darikwa and Dorrington (2011) relaxed this assumption by using the HIV prevalence figures at the times to which the estimates pertain. In addition, they assumed that HIV does not materially affect the time location estimates.

2.2.1.3 Direct child mortality estimation

Given census data that contains information about births and deaths that occurred in the year before the census, it is possible to calculate the Infant Mortality Rate (IMR) from the proportion among children born during the last 12 months who died before the census. However, the estimate so derived may lead to values that are implausible and inconsistent with other information. The biases in the estimate are usually due to errors in the data obtained through the retrospective interviews. Blacker and Brass (2005) argue that questions on date of last birth are prone to non-response and some enumerators do not ask the question but rather rely on the age of the youngest child they see in the household. Compounding the bias in these questions is the issue of respondents giving the date of birth as “about a year ago”, resulting in heavy heaping on the date 12 months before the census. They further observe that the estimate of infant mortality derived is assumed to represent about two-thirds of the actual IMR. This could be explained by under-reporting of deaths as well as wrong-dating of births and deaths (Adetunji 1996, Blacker and Brass 2005).

2.2.1.4 Blacker and Brass method

The Blacker and Brass (2005) method seeks to correct the deficiency of the direct procedure described above, of heaping of births on the date 12 months before the census. The method is used where the most recent births and their survival are recorded. Blacker and Brass (2005) showed that a better estimate of IMR can be obtained by converting the proportion of children dead amongst those born 24 months before the census. They did this by multiplying the proportion dead by a factor of 1.09 and showed that the result was a more robust and current estimate of IMR. However, the estimates derived can be biased due to under-reporting of deaths and age misreporting (Blacker and Brass 2005). The estimate is also biased downwards when a couple decides to have another birth following the death of a child and the child's death is not reported in the census (Chowdhury, Khan and Chen 1976).

2.2.2 Adult mortality

2.2.2.1 Orphanhood method

Brass and Hill (1973) devised the orphanhood method of converting the proportions of children orphaned by age into adult conditional survival probabilities. The method was premised on the simplicity with which data can be obtained by asking about the survival of the parents of all household members in censuses. The rationale behind the original method is that a target person is known to be alive at birth or conception of the respondent and some information is available as to the age at the beginning of exposure as well as the length of exposure to the risk of dying (United Nations 1983).

Brass established an equation relating the female probability of surviving from age 25 to age $25+n$ to the proportions of respondents in two contiguous five-year age groups whose mother was still alive at the time of the interview. Timæus (1992) used the following variant of the equations to estimate maternal and paternal conditional survivorship respectively:

$$l(25+n)/l(25) = a_0(n) + b_0(n)M + c_0(n)S(n-5)$$

$$l(35+n)/l(35) = a(n) + b(n)M + c(n)S(n-5) + d(n)S(n)$$

where the $a_0(n)$, $b_0(n)$, $c_0(n)$, $a(n)$, $b(n)$, $c(n)$, and $d(n)$ are coefficients derived from linear regression using data from simulations of different fertility and mortality regimes, $S(n)$ is the proportion of respondents in five-year age groups with mothers alive and *M* is the mean age at child bearing.

The method assumes that mortality had been constant in the period prior to the survey, and the biases that can be introduced when this assumption is violated can be large (Palloni, Massagli and Marcotte 1984, United Nations 2002). Palloni, Massagli and Marcotte (1984), found that estimation of adult mortality using the orphanhood technique under the assumption of constant mortality when in fact mortality had been decreasing at a faster rate, resulted in values that had relative errors of between 5 and 39 per cent. The errors in the mortality values, as would be expected, are worse for the older age groups than they are for the younger age groups. The older age groups were exposed to progressively higher mortality in the past than the younger age groups because mortality had been decreasing over time. To circumvent this problem, time location estimates were derived on the assumption that the mortality change was linear (United Nations 1983, United Nations 2002).

An additional assumption is that there is no misreporting of orphanhood status by respondents (Palloni, Massagli and Marcotte 1984). The adoption effect – the reporting of adoptive parents as one's biological parents – can occur when the respondent fails to identify correctly the relationship between mother or father and her or his offspring (Palloni, Massagli and Marcotte 1984, Bah 1999). Palloni, Massagli and Marcotte (1984) note that failure to identify correctly the relationship between mother or father and her or his offspring might either be intentional or unintentional. Of the two, they single out unintentional concealment as the significant contributor to biases in the proportion of orphans. They further note that the quantum of the bias will depend on the nature of the social norms and practices that have an influence on adoption. Biases introduced by adoption are likely to be higher for young ages, because for older ages both the adoptive and biological parents will have been exposed to the risk of dying for longer and thus are more likely to be dead (United Nations 2002).

The mortality estimates computed using the orphanhood method only reflect the mortality of a select subgroup (United Nations 1983, United Nations 2002, Palloni, Massagli and Marcotte 1984). This is because orphanhood status is only reported on children who have survived to the interview date. Thus the population of dead children will not contribute information on orphanhood status. If the deaths amongst the children were independent of the mortality experience of their mothers or fathers, then the observed proportion of non-orphans would be no different from that observed in the absence of child mortality. However, in practice the mortality experience of parents (particularly mothers) is closely related to that of their children (Palloni, Massagli and Marcotte 1984).

Another weakness of the method derives from over representation of status of parents with large families (United Nations 2002, Blacker 1981). Additional questions that sought to identify the eldest surviving sibling were included in some censuses; however, analyses of the data revealed that too many siblings claimed to be the eldest sibling, resulting in overstatement of mortality (United Nations 2002, Blacker 1981).

The use of mortality models that do not accurately capture the mortality experience of a population under investigation can distort the mortality rates (Palloni, Massagli and Marcotte 1984, Bah 1999). This is particularly important in light of HIV/AIDS, which has led to mortality schedules that are different from the standard mortality schedules (Preston, Heuveline and Guillot 2001, Clark, Jasseh, Punpuing *et al.* 2008). Related to this are biases introduced by HIV/AIDS as discussed by Timæus and Nunn (1997). They identified selection bias in maternal mortality introduced by HIV induced reduced fertility in women and correlation of mortality between mothers and their children. For example, the presence of communicable diseases, such as diarrheal diseases due to lack of proper water and sanitation facilities in an area might cause the mortality of parents and their children to be correlated (Sartorius, Kahn, Vounatsou *et al.* 2010).

In addition to the violation of the assumptions of the method, errors in the values obtained from application of the orphanhood method may be attributable to the other errors in the data. These errors might be induced by non-response to the questions or respondents stating that they "don't know" (Bah 1999). Biases in these questions are due to under-reporting of deaths of parents who died when their children were quite young and brought up by other adults whom they report to be their biological parents (the adoption effect). Parents with no surviving children or those who never had children are completely excluded from the analysis (Gakidou, Hogan and Lopez 2004).

2.2.2.2 *Recent household deaths.*

Using census data, it is possible to calculate crude mortality rates as the ratio of total deaths occurring in a recent period before the census to the mid-period population. Age-sex specific crude death rates can also be computed in a similar manner. However, the rates produced are likely to be distorted by omission of deaths, multiple reporting of the same deaths and age misreporting (Adetunji 1996). Questions asking about recent household deaths typically suffer from under-reporting of events and reference period errors probably due, in part, to emotions associated with the death that invoke an unwillingness on the respondents to talk about the dead (Timæus 1991c, Gakidou, Hogan and Lopez 2004, Blacker 1977). In addition, some respondents may completely misunderstand the questions (Feeney 2001). Timæus (1991c) points out that not everyone is clearly attached to a single household and some individuals do not belong to any household, e.g. single people staying alone. He further argues that individuals who stay alone are amongst the more likely to die and thus their subsequent death more likely to go unreported. The illness and death of an adult in a household might also result in the dissolution of that household before the census is conducted and this is particularly noticeable at older ages (Dorrington, Moultrie and Timaeus 2004). There may also be over-reporting of deaths which might occur due to different households reporting on the same death of an individual who was considered to be part of these different households (Timæus 1991c, Feeney 2001).

2.3 Methods of assessing death registration and cause of death

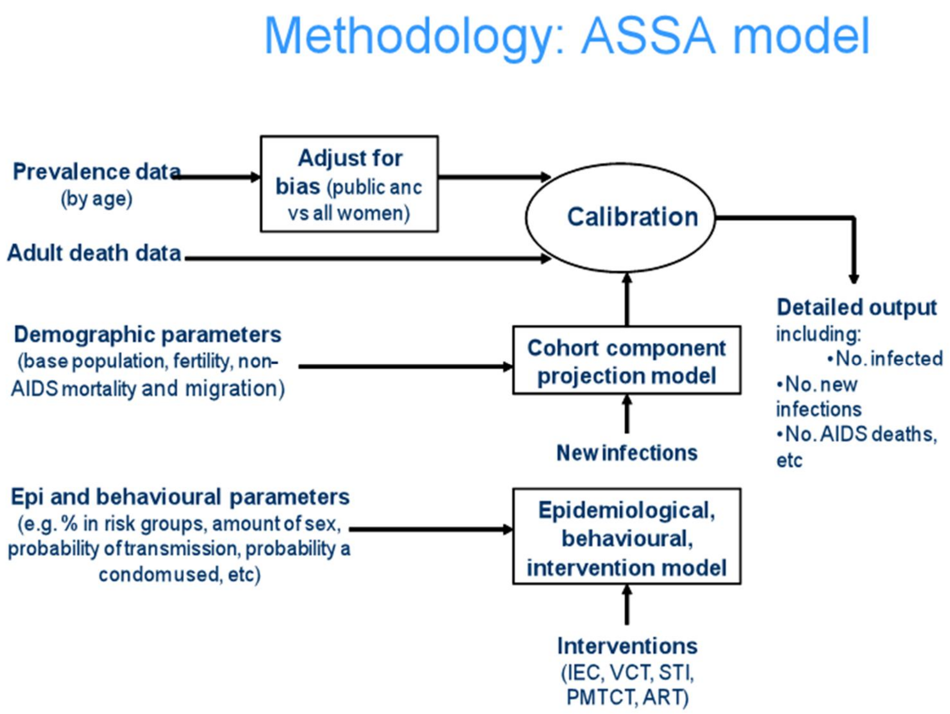
When one only has access to one census, the death distribution methods cannot be applied to assess the level of completeness of reporting of deaths in the census since it is very unlikely that the population is stable. Death distribution methods are demographic methods that make use of data on deaths and the population at risk by age and sex to estimate adult mortality (Timæus, Dorrington and Hill 2013). However, in a few developing countries, an estimate of the completeness of death can be obtained by comparison with the vital registration system after adjustments for under-reporting of those deaths. Inclusion of questions in censuses that ask if household deaths reported in censuses have been registered have the potential of giving an indication of completeness of registration.

In an effort to carry out cause of deaths analysis aimed at informing decisions on health care strategies and programmes, some censuses and surveys have included cause of death questions when recording household deaths. Related to this is the renewed interest in maternal mortality following the recommendations of the UN which have led to a sharp increase in their use in the 2010 census round (Dorrington and Bradshaw 2011, Queiroz 2011, Hill and Stanton 2011).

2.4 Other sources of mortality estimates

Models become handy in situations where there are no other sources of mortality estimates. In South Africa, the ASSA2008 model is the only currently available demographic model that provides estimates of mortality down to provincial level, which take into account the impact of the boundary changes that occurred in 2006. The ASSA2008 model is produced by the Actuarial Society of South Africa (ASSA). It provides national as well as provincial mortality indicators. The ASSA2008 model takes into account reported data for all recorded deaths up to 2008, recent antenatal surveys and household surveys, as well as recent data on the coverage of antiretroviral treatment and prevention of mother-to-child transmission programmes. The figure below shows a schematic representation of the ASSA model methodology.

Dorrrington, Johnson and Budlender (2010) describe the model as follows. The ASSA model is a cohort component projection model which allows a user to project a given population over a set period of time. The projections are normally done on a year by year basis. The output of the model is based on the effect of the HIV/AIDS on a given population per set period of time. The main outputs are: number of people infected with HIV, number of new infections, number of AIDS deaths and other demographic indicators. Furthermore, the model stratifies a given population by age and sex with the main catergories being young (up to age 13), adult (14-59) and old (60 and above). The model also allows the adult population to be stratified by percieved level of risk to contracting HIV through Heterosexual activity, into four risk.



Source: Department of Health (2011:23)

Various assumptions are taken into consideration in the model, including epidemiological and behavioural, and intervention assumptions. The table below shows a summary of main issues covered by each assumptions:

Table 2.2 Assumptions used in the ASSA2008 Model

Assumption Catergory	**Key elements covered under each assumption group**
Demographic	base population, fertility, non-AIDS mortality and migration
Epidemiological and behavioural	distribution of the adult population across the four risk groups, amount of sex, probability of HIV transmission, condom use
Interventions	Programmes such as voluntary counselling and testing (VCT), coverage and adopted guidelines on antiretroviral therapy (ART) and prevention of mother-to-child transmission (PMTCT)

Source: Dorrrington, Johnson and Budlender 2010

The number of new HIV infections is obtained by using epidemiological and behavioural, and interventions parameters. This number of new infections is then used in the component projection model resulting in a crude output before it is refined to match available adult death registration data (adjusted for under-reporting) and HIV prevalence data by age from anti-natal clinics (ANC) (Dorrrington, Johnson and Budlender 2010).

2.5 Agincourt site and mortality estimates

2.5.1 Agincourt site description

The Agincourt health and demographic surveillance site (AHDSS) is in a rural setting located near the border of South Africa and Mozambique, about 500km northeast of Johannesburg in the Agincourt sub-district of Bushbuckridge region in Mpumalanga province. Agincourt was part of Limpopo province before a boundary change in 2006 which resulted in it becoming part of Mpumalanga (MRC/Wits Rural Public Health and Heath Transitions Research Unit 2012). The AHDSS covered only 21 villages before the boundary changes, increasing to 26 after the boundary changes.

The AHDSS makes an in-depth coverage of demographic events of fertility, mortality, and migration occurring within the geographic confines of Agincourt possible, starting from the comprehensive register established during the baseline census in 1992. The baseline census has been updated annually by census update rounds by trained field workers. During each update, a knowledgeable person per household responds to questions and information about the household and its members is updated, as well as events recorded that occurred since the last census (Mee, Mgiba and Collinson 2006, Sankoh and Byass 2012).

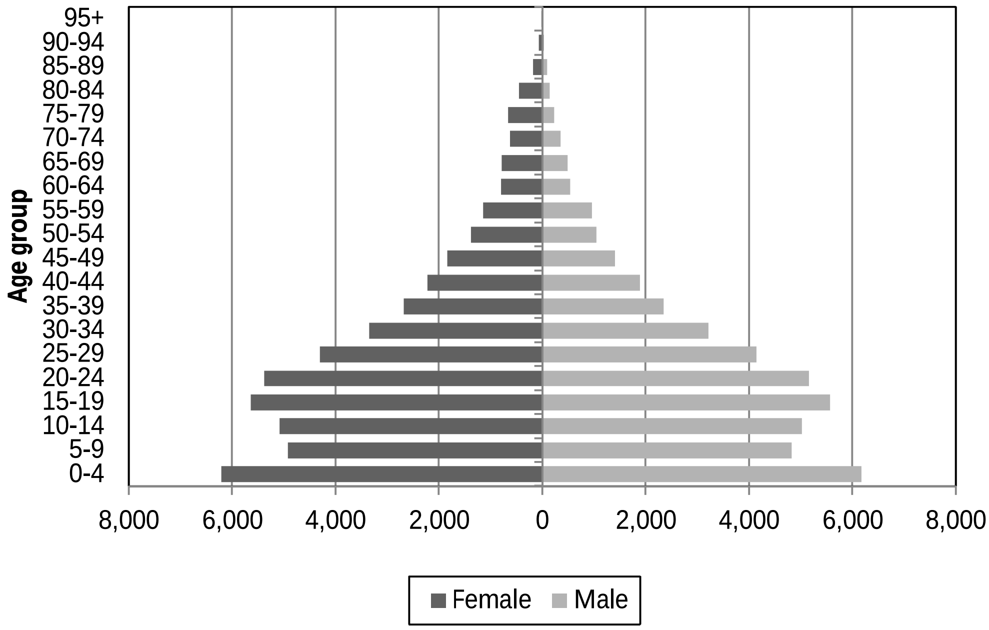
The population enumerated in the initial census consisted of 57 509 individuals residing within 8 896 households. By 2010, the *de jure* population had increased to around 84 000, consisting of 25 villages (Williams, Schatz, Clark *et al.* 2010). The primary objective of the ADHSS is to inform health and social policy by providing a research infrastructure and longitudinal database for a range of community based studies relating to burden of disease, health systems' interventions and social/household/community dynamics (Mee, Mgiba and Collinson 2006).

More than half of all men in the Agincourt site aged 25–59 are labour migrants due to scarce employment opportunities in the area, with those getting local employment likely to be in the public sector. There is one health centre linked to four satellite clinics serving the area, with the district hospital some 40 km away (Tollman, Kahn, Garenne *et al.* 1999).

2.5.2 Agincourt age sex profile

Figure 2.1 shows that females outnumber males at almost all age groups which is corroborated by the overall sex ratio of 92 males per 100 females. Age-specific sex ratios are very close to 100 for age groups below age 20 and they become significantly lower than 100 for higher ages. This is likely due to migrant labour, which is more common amongst men than it is amongst women at Agincourt (INDEPTH Network 2011).

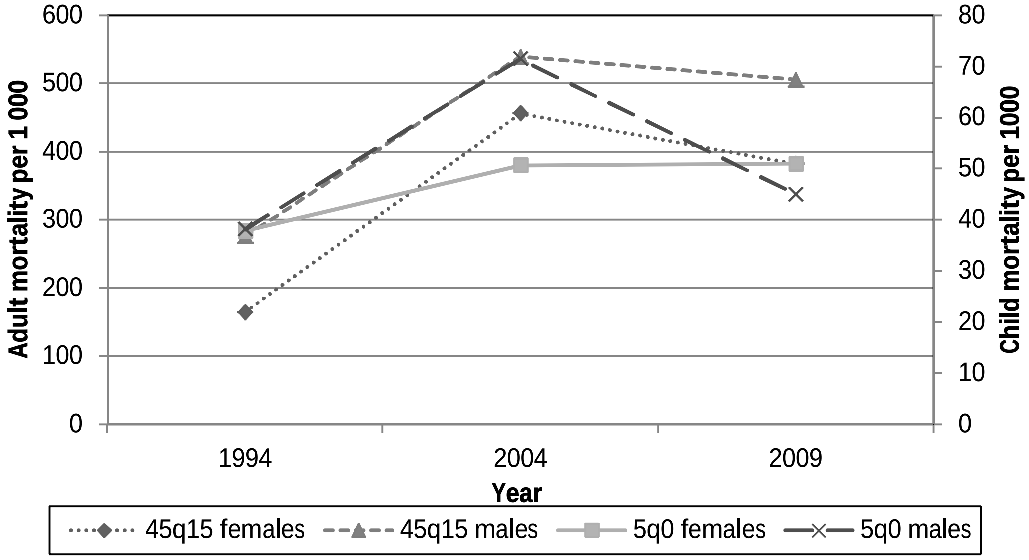
Figure 2.1 Agincourt HDSS population pyramid 2009



2.5.3 Mortality estimates

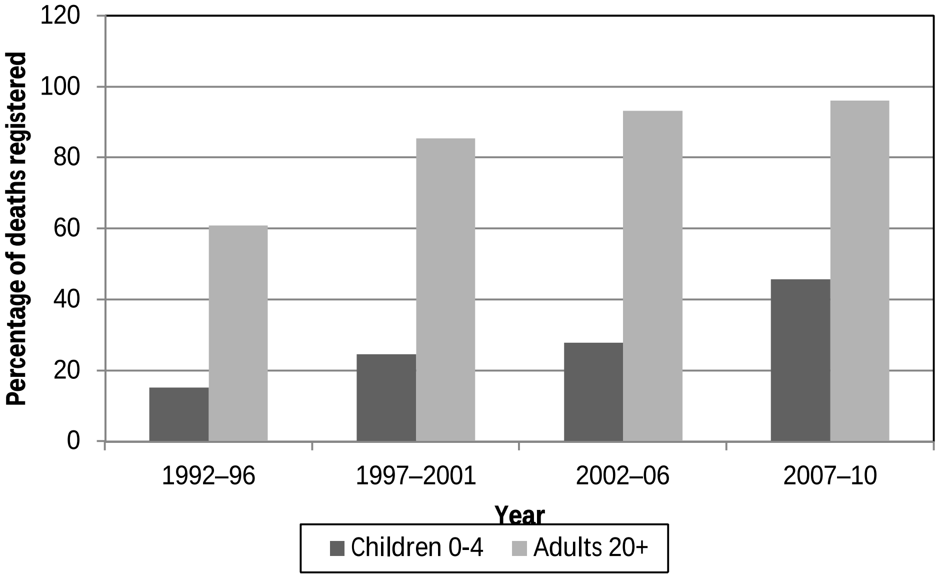
Figure 2.2 presents child $\left({}_5q_0\right)$ and adult $\left({}_{45}q_{15}\right)$ mortality measures at Agincourt for the period 1994 to 2009. The life-table probabilities, ${}_nq_x$ represent the probability that an individual age *x* will die before reaching age *x+n*. Over the period since the inception of Agincourt in 1994 to 2009, both child and adult mortality have increased over time peaking around the mid to early 2000s before declining again (Kahn, Collinson, Gómez-Olivé *et al.* 2012). The reversal in child mortality trends is attributable to health interventions such, as the prevention-of-mother-to-child transmission programmes that have been implemented at AHDSS. Life expectancy at birth decreased from 72.7 to 64.4 years and 68.2 to 55.7 years for females and males, respectively over the period. Adult female mortality almost trebled between 1994 and 2004, whilst adult male mortality doubled over the same period (Kahn, Collinson, Gómez-Olivé *et al.* 2012).

Figure 2.2 Child and adult mortality at Agincourt, 1994-2009



There has been an increase in death registration for both children under five and adults (20+) from 15.2% to 45.8% and 60.9% to 96.1% (see Figure 2.3), respectively, between 1992-6 and 2007-10 (Kahn, Collinson, Gómez-Olivé *et al.* 2012).

Figure 2.3 Death registration by age group and time period



2.6 UCT survey objectives

The objectives of the UCT survey are threefold. First, the survey aims to identify and understand the errors and mistakes that often arise in censuses. This is to be done by comparing the data collected using census instruments with those collected on the same population of Agincourt for the past fifteen years, including an analysis of which households/respondents provided the most accurate data on such questions. Second, the survey results will be used to evaluate the demographic methods used to estimate mortality, fertility and migration in Africa and similar settings. Third, it is hoped that the survey results can be used to motivate for the development of better census data collection methods and demographic estimation methods.

2.7 UCT survey background

As alluded to in section 2.1, the majority of developing countries collect demographic data through the process of question and answer in the form of censuses or large surveys (Cleland 1996). As huge resources are committed in these censuses and surveys it becomes imperative that the data should be as accurate as possible for the topics the census covers, or at least we should know accurately by how much they vary from the truth. To this end, the UCT survey is one of the scientific undertakings to investigate and assess the extent of errors and mistakes that occur in censuses. It is meant to address three main goals for Sub-Saharan countries that rely on census-type questions as tools for demographic data collection due to lack of complete vital registration systems. First, the study seeks to evaluate and assess the errors in the responses to the census-type questions used to collect data used with indirect demographic estimation methods to produce demographic indicators. Second, another objective of the study is to assess the performance of the indirect techniques against direct estimates. The third goal of the study, which is related to the second, is to find ways of improving the indirect methods and estimates. The Agincourt HDSS provides a rich longitudinal data source spanning over 15 years to permit the attainment of the UCT survey objectives.

The UCT census validation exercise is collaborative work between the University of Cape Town, Agincourt HDSS and the University of Washington scientific staff, and was mainly funded by the INDEPTH network. The UCT survey covered a *de facto* population of the original 21 villages of AHDSS excluding five villages that became part of the AHDSS following boundary changes that occurred in 2007.[1]

2.8 UCT survey methodology

Field workers were selected from within the Agincourt location and were not part of AHDSS regular staff. Locals were recruited to encourage buy-in and participation of the respondents through identifying with the enumerators as well as to minimise barriers to communication that would likely arise if enumerators were from other areas that had different languages (United Nations 2008a). Also in terms of logistics it was cheaper and easier to deal with locals.[1]

In any census or survey, training of staff is essential to ensure that they can carry out the necessary tasks. Training presents an opportunity to demonstrate to staff the importance of

[1] Personal communication with Prof. Tom Moultrie, 30 April 2012

quality in all aspects of the census/survey process and helps the entire census team to have a shared responsibility for maintaining and improving quality (United Nations 2008a). The UCT enumerators were trained for a full day and practised their field work using simulated data.

The UN emphasises the need for adequate supervision of the whole enumeration process. The UCT survey had the maximum enumerator-to-supervisor ratio (6:1) recommended by the United Nations (2008a).

According to the United Nations (2008b), it is desirable to keep the enumeration period short in order to avoid double counting and omissions, which can occur in spite of a single reference date. On the other hand, the shorter the enumeration period, the greater the number of field staff that have to be recruited, trained and supervised. This increases costs and may compromise the quality of the data (U.S. Bureau of the Census 1985). In the UCT census, a third enumeration team was added midway through the survey period to ensure that the survey was carried out within a reasonably short time to avoid the errors mentioned above.

The timing of the survey has an important bearing on the results and any choice should be made to take into account the season of the year as well as accessibility of the area at that particular time. Any date chosen should be such that the likelihood of finding people at their usual place of residence is highest and thus increases the likelihood of them being enumerated (United Nations 2008a). In the UCT survey, where an enumerator's first visit failed to find a suitable respondent, two subsequent visits were made. These visits were made on different days and at different times (but not after dark) to increase the likelihood of finding a suitable respondent.

Regardless of a thorough and accurate survey enumeration, the usefulness, quality and timeliness of the survey information will be compromised unless the data collected are properly processed (United Nations 2008b). Data processing/capture is defined by the United Nations (2008b) as the process of converting the information obtained in the census to a format that can be interpreted by a computer. There are various methods that can be employed, which include keyboard data entry, optical mark reading, optical character reading and imaging processing techniques (United Nations 2008b). The UCT survey used keyboard data entry to capture the responses from the paper questionnaires. In the process, errors are likely to be introduced by data typists failing to read illegible handwriting of the enumerators as well as fatigue from repetitive work or carelessness. Therefore in order to minimise these errors at data capture, two teams independently captured the survey data, and then compared the captured data sets and corrected any inconsistencies.

The United Nations (2008a) recommends that in order to enrich data analysis and minimise costs it may be desirable to link the population census and other surveys/censuses on the same area/population that are held within a short period of each other. In the case of the UCT census, the initial phase of linking the UCT survey with the AHDSS data was done in the implementation phase by making use of the Enumerator Area (EA) maps as well as Household Identities (HHIDs) that were available from the most recent AHDSS census update. The important and practical assumption being that there were no material changes in the EAs over the intervening period between the AHDSS census round and the UCT census. As for the HHIDs, the AHDSS provided a basis upon which the UCT survey could be undertaken.

The questions the UCT survey asked included child mortality (CEB/CS), adult mortality (orphanhood) and deaths in the household. It also sought to test additional questions that could be used in mortality and fertility analysis but have not been included in censuses. Additional questions related to mortality analysis included the following: “Was your father or mother alive on your 18th birthday?”; “Was your father or mother alive when your first child was born?”; “Was your father or mother alive at the time of the first elections of 1994?” These questions can be used to enhance the recall of distant events in the past as some of the events have a deep and vivid meaning to the respondents. For example, respondents are likely to recall if their father or mother was alive when they gave birth to their first born, particularly in African culture where birth of the first born child is of significance. In the case of the question which asks about the survival of the parent in relation to the first election of 1994, respondents are likely to recall whether their parents were alive or dead during the first election of 1994 because the event is of great historical meaning to most South Africans. Answers to these questions can permit the estimation of levels and trends in adult mortality, for example, using Timæus' (1991b) variant of the orphanhood method, which gives estimates that are free from the effects of adoption.

Although in South Africa the vital registration systems can be considered to be fairly complete, there have been marginal increases in the level of completeness of registration of deaths over time but the exact level of completeness is unknown (Dorrington and Bradshaw 2011). The UCT survey included a question that asked whether deaths that occurred in the 12 months before the census were registered with the Department of Home Affairs or not. With suitable adjustments for under-reporting of deaths in the census, this question can be used to obtain an estimate of the completeness of death registration.

The UCT survey also included questions that sought to establish whether the death was accidental or non-accidental, and for females aged 15-49 whether the death was pregnancy-related. These questions have recently become of interest in light of measuring progress towards the millennium development goals (MDGs) (Dorrington and Bradshaw 2011).

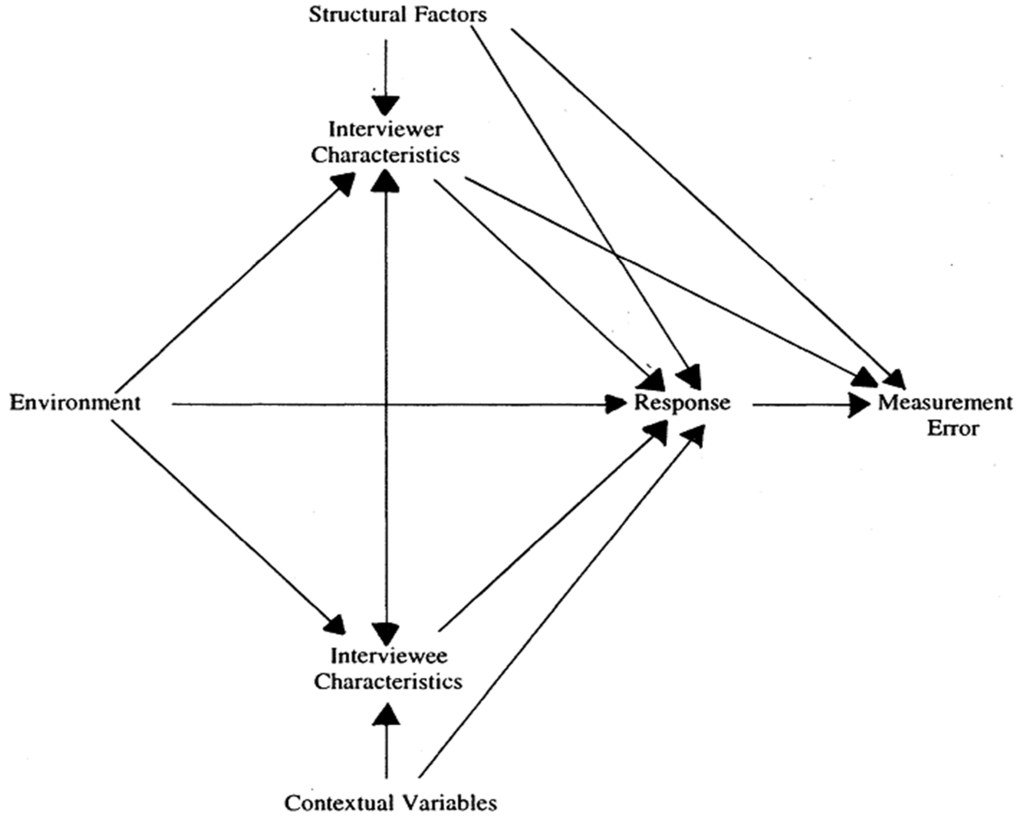
2.9 Sources of measurement errors in censuses/surveys

There are numerous errors that can arise in the data collected through censuses/surveys. The UCT survey is no exception. This section discusses some of the potential sources of errors that could be found in censuses and surveys, including the UCT survey.

Figure 2.4 shows a conceptual framework that can be used for analysing measurement error in data collection (Vemuri 1994). Response or measurement errors occur for individuals that are properly included in the census or survey but for whom improper characteristics/information are/is recorded (United Nations 2004, U.S. Bureau of the Census 1985). Vemuri (1994) notes that in addition to interviewer and interviewee characteristics, structural factors, contextual variables and environment in which the data are collected affect the quality of the response.

Vemuri (1994) citing work by others notes that interviewer preconceptions were correlated with the answers received by the census office. He cites work by others in Canada in which they found that interviewers had certain inclinations on questions related to ethnic origin, mother tongue and bilingualism.

Figure 2.4 A conceptual framework for analysing measurement errors in data collection



Source: Vemuri (1994)

There are a number of ways in which content errors may arise, namely, poorly designed questions or poor sequencing of the questions, or by poor communication between respondent and enumerator, as well as by mistakes in coding and data entry, errors in manual and computer editing, and erroneous tabulations of results (United Nations 2004). In addition, enumerators may contribute to content errors when they misread the questions to the respondents or they incorrectly record the responses from the respondents. Content errors caused by enumerators can, however, be minimised when the enumerators receive adequate training and are made to understand the importance of the role that they play in the whole exercise (United Nations 2004).

Census/survey questionnaires should be written in local language or at least an enumerator copy written in a respondent's language should be available to avoid the enumerator carrying out on-the-spot translations. Allowing enumerators to carry out ad hoc translations will result in non-uniformity and inconsistency in the way questions are translated into local languages by different enumerators thus introducing errors in the data (Kiregyera 1982).

Another source of content error is the respondent, who might misunderstand the questions being asked, or deliberately or through ignorance misreports on events, particularly when he/she e reports on behalf of others (United Nations 2004, Hansen and Waksberg 1970).

Other errors might be introduced as a result of miscoding of responses and at the data capturing stage. If optical character recognition (OCR) devices are used at data capturing this can be another source of errors, particularly for characters that have a similar structure.

3 METHODOLOGY

This chapter presents the methodology employed to assess the errors in the data collected through the survey, and the biases that are in the direct and indirect mortality estimates derived from these data.

3.1 Data matching

To enable the assessment of the errors in the data collected, a clerical matching exercise was done for the households and the individuals enumerated in the UCT survey with that of the AHDSS 2009 census update. First, the matching was done for the households in the 21 villages covered by the UCT survey using the unique household I.D., common between the UCT survey and AHDSS census. The matching of households provided a basis for finding matches for the individuals and also permitted analysis of household dynamics to be investigated.

Second, once the households were matched, individuals in those households were then matched by comparing the characteristics, such as name, surname and date of birth (DOB) or age. This study used this approach to match individuals between the UCT survey and the AHDSS census as the UCT survey did not collect individual identity numbers (IDs). (It did not collect IDs because only one person responded in each household, and it was assumed that the respondent might not know the ID numbers of others in the household.) Third, an attempt was then made to match the remaining unmatched individuals by comparing the characteristics of the individuals between the two data sets without regard to their household IDs.

3.2 Accuracy and completeness of the information gathered by the survey

The overall completeness of the UCT survey was determined by dividing the UCT survey total population by the total AHDSS census population of the 21 villages covered by the UCT survey. The AHDSS population for this purpose, and for further analyses, excluded household members who were deemed to be migrants, as indicated by the residence status variable. Only non-migrant residents were included as these were the ones likely to have been present to be enumerated by the de-facto UCT survey. The residence status was determined from the AHDSS census update "Resident Status table", which keeps a record of the time spent in the year preceding the update and assigns resident status codes of 'migrants' (mainly migrant labourers) or 'permanent residents.' Residents who spent more than six months outside AHDSS were regarded as migrants; otherwise they were regarded by the AHDSS as permanent residents.

The overall household coverage was determined as the ratio of the total number of the households enumerated by the UCT survey to the total number of households in the 21 villages in AHDSS covered by the UCT survey. Households in which all adult members were migrants were excluded from the AHDSS data set as enumerators would have been unlikely to enumerate these households.

3.2.1 Accuracy of population numbers by age and sex

The measurement of completeness of the survey was disaggregated by sex to investigate if there was differential coverage by sex. Further, the accuracy of the population estimates as captured by the UCT survey was assessed by tabulating the population numbers by sex and five-year age groups, and compared to the corresponding tabulation of the population numbers from the 21 villages in the AHDSS census that were covered by the UCT survey. The age used for this preliminary assessment was the age derived from dates of birth where these were provided, or if not available, it was based on recorded age. Individuals with missing age were distributed, assuming they had the same distribution as those with specified age or dates of birth. Plots of UCT survey population numbers by sex and five-year age groups and the corresponding benchmark[2] population numbers by sex and five-year age groups were inspected for differential omission by age and sex.

Overall consistency between reported/recorded age at last birthday and dates of birth in the UCT survey was assessed by plotting the distribution of the difference between the age reported/recorded and the age at last birthday derived from dates of birth. The ages at last birthday were derived from dates of birth by taking the integer value that results from subtracting the date of birth (measured in days) from household visit date (measured in days) and dividing the result by 365.25 days.

The accuracy of reported/recorded age was assessed by plotting the distribution of the difference between the true age (as captured by the AHDSS) and the reported/recorded age for the individuals that were successfully matched between the UCT survey and the AHDSS census. A negative difference implies age exaggeration in the survey; a difference of zero implies accurate reporting/recording; whilst a positive difference implies age under-statement in the survey. Similarly, the accuracy of the reported dates of birth is assessed by plotting the distribution of the difference between the true age (as captured by the AHDSS) and the age derived from reported/recorded dates of birth.

For individuals in the UCT survey with no dates of birth or age recorded but which are matched, the age (as captured by the AHDSS) by sex was plotted to investigate differential non-reporting/recording by age and sex.

[2] The AHDSS census data are, for practical purposes, the most accurate data set that are available and are used as the benchmark data for the assessment of the data obtained through the UCT survey.

3.2.2 Accuracy of household numbers and composition

A tabulation of the household size in the UCT survey alongside the households in the 21 villages in AHDSS census covered by the UCT survey reveals if all the individuals in the households have been captured accurately. For the AHDSS, migrant household members are again excluded.

3.2.3 Accuracy of answers to each of the specific questions

3.2.3.1 CEB/CS data assessment

In order to assess the errors in the CEB/CS data, equivalent data for the women with matches from AHDSS were extracted. For each woman the number of children she ever gave birth to was derived from the mother's ID, which is recorded for all children born to each women resident at Agincourt. The numbers of these children who have died were determined from the status field in the AHDSS census. Using these extracted data, the number of children ever born and then the number of children surviving for each woman in each survey was compared to the benchmark by tabulating the UCT survey numbers versus the AHDSS census. For the women who were matched, the ratio of the children reported in the UCT survey to the actual number of children was then plotted by sex of the children and age of their mothers.

Births in the previous 12 months collected by the UCT survey were used to compute an estimate of the mean age at birth, *M*. The true mean age was computed from AHDSS using the maternity history data of the matched permanent resident women who had given birth within the year of the UCT survey. The UCT survey estimate was then compared to the true estimate.

3.2.3.2 Orphanhood data assessment

For the AHDSS census, maternal orphanhood data were derived from the survival status of the mothers indicated in the AHDSS census data set, taking into account the date of death of the mother and the date of household visit. Unfortunately, the AHDSS does not collect data on the survival status of the fathers hence a similar analysis could not be done. The derived maternal orphanhood data were then compared to the UCT survey data for each of the matched individuals. Tabulation of orphanhood status by sex and age for the UCT survey alongside the AHDSS census were used to investigate orphanhood misreporting in the survey. Plots of the proportion orphaned by age and sex were produced to identify the presence of the unrealistically high proportion of young respondents with surviving parents, which could be used to indicate adoption effects.

3.2.3.3 Household death data assessment

For the AHDSS, household deaths in the year preceding the UCT survey date were derived from the date of death and the date when the household was interviewed in the UCT survey. These household data were then compared with those from the UCT survey. For each household, the objective was to determine the difference between reported deaths and actual deaths that occurred in the household. Analysis of accuracy of age at death was done in a similar manner to that of the general population.

An investigation into whether deaths without specified age have a different age distribution from those with stated ages at death was done using the deaths that were matched between the two surveys. This was accomplished by computing the ratio of the number of deaths without age at death to those with age at death by true age. If the distribution of those with no stated age at death was the same as that of deaths with stated age, then the ratio was expected to be roughly constant with age.

To investigate the accuracy of reporting of categorisation of deaths by whether the death was due to violent or maternal causes, AHDSS cause of death information was derived from the International Classification of Disease and Health Related Conditions 10 (ICD10) classification variable to match the categories used in the UCT survey. The two main categories as captured by the UCT survey were considered; whether the death was accidental or non-accidental, and for females aged 15-49, whether the death was pregnancy related. Cross tabulations of the UCT survey versus the AHDSS of each category were used to investigate the accuracy of the recording of these in the survey.

The overall extent of death registration was assessed first by sex then by age and sex. The proportion registered by age and sex were then plotted on the same set of axes to investigate differential registration by age and sex. The age categories under-1, 1-4, 5-14, 15-49 and 50+ were considered for differential registration by age. The accuracy of the UCT survey estimates of the proportions of deaths that were registered are compared to those of AHDSS through cross-tabulations. In addition, the registration estimates were compared to other published estimates.

3.2.4 Investigation of respondent characteristics that affect accuracy of responses

In this study, a multinomial logistic model (MNLM) was used to evaluate the effect of age, sex, household size, residence status and whether the respondent was reporting on himself/herself, on the accuracy of the responses to each of the questions. Ordinary logistic regression was considered but was dropped as it was not as detailed as the MNLM. In particular, ordinary logistic regression would have only allowed assessment of the output variable using two broad categories – "accurate" and "inaccurate". On the other hand, the MNLM permitted a further breakdown of the "inaccurate" dependent category making the output variable a three-level nominal variable. Specifically, the inaccurate category (in ordinary logistic regression) was split into "understated" and "overstated" subcategories allowing for the evaluation of the direction of the bias. The respondent characteristics were regarded as the independent variables, while the dependent variables were the “response” variables, which were recoded into the three-level nominal variable taking the categories: “understated”, “accurate” and “overstated”, with their intuitive meaning. In essence, two logistic models were run with each logistic model to compare with the other two levels of outcomes (“understated” and “overstated”) to the “accurate” dependent category level, i.e. the “accurate” level was used as the base category. These logistic models were run first on respondents reporting on themselves and secondly on respondents reporting on other household members.

The dependent variables were derived by subtracting the UCT survey value from the AHDSS value. For the purposes of the univariate analysis all the variables were treated as categorical variables. Age was split into the four categories 15-24, 25-44, 45-59 and 60+, following the third level age classification recommended standards for demographic, social and related economic data used by the United Nations (United Nations 1982). The household size variable was recoded into 3 levels: “small” (1-2 individuals), “average” (3-5 individuals) and “large” (6+individuals). Household categories were chosen taking into account the average household size of 3.8 for Limpopo and Mpumalanga (Statistics South Africa 2012b). However, for the logistic models both age and household size variables were entered as discrete variables with a unit step. This was done to investigate the effect of a unit increase in the respondent’s age or household size on the accuracy of responses.

A Chi-square test of association between all the selected independent variables and the dependent “response” variable was performed to determine if there was a significant association amongst the covariates as well as between the covariates and the dependent variable. A test for multi-collinearity was used to assess if two or more of the covariates were highly correlated, which might result in unreliable regression coefficients. A stepwise forward logistic regression of the “response” variable was used for each of the questions considered here, starting with main effects and then including interaction terms. The logistic model was used to assess the effects of the respondent characteristics on the errors in the questions on children ever born (CEB), children ever born now dead (CD), and reporting of age and dates of birth. From the above model, the characteristics of respondents who gave the most accurate responses were determined from the relative risk ratios computed from the final model for each question. For each final model generated, model diagnostics were used to check that the models were appropriate and that there were no unusual observations that were exerting undue influence on the model. The Hosmer-Lemeshow test was used to assess the overall goodness of fit whilst the Pregibon Delta beta statistic was used to assess if the fitted models are excessively influenced by one or a small number of covariate patterns (Obermeyer, Rajaratnam, Park *et al.* 2010, Chen, Ender, Mitchell *et al.* 2008, Menard 2002, DeMaris, Teachman and Morgan 1990).

The multinomial logistic model (MNLM) used in effect estimates the following binary models:

$$\ln\left[\frac{P(Understate|\bar{x})}{P(Accurate|\bar{x})}\right]=\beta_{0,U|A}+\sum\beta_{k,U|A}x_k+\sum_{j\neq l}\beta_{jl,U|A}x_j*x_l$$

$$\ln\left[\frac{P(Overstate|\bar{x})}{P(Accurate|\bar{x})}\right]=\beta_{0,O|A}+\sum\beta_{k,O|A}x_k+\sum_{j\neq l}\beta_{jl,O|A}x_j*x_l$$

where $O|A$ and $U|A$ signify that the logistic was comparing the "overstated" and the "understated" to the base category "accurate", respectively.

3.3 Accuracy of the estimates of mortality produced from data gathered by the survey.

This section describes how the direct and indirect mortality estimates were derived and how the accuracy of these estimates was assessed.

3.3.1 Direct mortality estimates

Crude death rate for the UCT survey were computed as the ratio of the deaths reported in the survey to the total UCT survey population backward projected by half a year. The backward projected population (${}_nP_x^r$) was used to ensure correspondence between the deaths and the population exposed to the risk of death.. Migrant deaths and individuals reported to be migrants in the UCT survey were excluded. The growth rate used for backward projection was obtained using Agincourt population estimates for two different periods published by Clark, Collinson, Kahn *et al.* (2007) and Kahn, Collinson, Gómez-Olivé *et al.* (2012). This crude rate was then compared to that from published literature and estimates from the ASSA2008[3] model population projection for Limpopo province[4].

Age-sex specific mortality (${}_nM_x^r$) rates were computed using the deaths and population estimated above after both the deaths. The age-specific rates were scaled up so that the all-age rate was the same as that from AHDSS. Those without age were distributed in proportion to those with age and assuming they were all adults aged 19 years and above, although indications are that this assumption is not well supported by the data. The age-sex specific mortality rates were computed using the formula:

[3] A detailed discription of the ASSA2008 model can be found in Dorrrington, R., Johnson, L. and Budlender, D. 2010. *ASSA2008 AIDS and Demographic models user guide*. Beta Version. Cape Town: AIDS Committee of the Actuarial Society of South Africa..

[4] The Limpopo province is used for two reasons. First, between the two adjacent provinces, Limpopo province has a mortality level that is lower than that of Mpumalanga and the Limpopo mortality level is more consistent with that of Agincourt. Second, since some of the indirect mortality estimates we seek to assess accuracy of are for periods that are before 2007 when AHDSS was still part of Limpopo, this choice seems reasonable.

$$ {}_nM_x^r = \frac{{}_nD_x^r}{{}_nP_x^r}, $$

where r represents the sex group (males or females) and ${}_nP_x^r$ is the survey population aged between exact age x and $x+n$-1 last birth day at the time of the survey backward, projected by half year and ${}_nD_x^r$ are the deaths aged between exact age x and $x+n$-1 last birth day at the time of death.

The age-sex specific central mortality rates for the UCT survey and those obtained from ASSA2008 model population projection for Limpopo province or from published literature were then plotted on the same axes by age and sex to reveal any biases in the UCT survey estimates.

3.3.2 Direct mortality estimates by categorisation of death.

Crude death rates by categorisation of death and sex were computed as the ratio of the number of deaths in a particular category divided by the population exposed to the risk of dying. This was done after distributing the deaths without age at death and those with missing age for the population at risk by assuming that they had a distribution similar to that of those with reported/recorded age. Age-categorisation-specific death rates were not computed due to the low numbers of expected deaths.

3.3.3 Indirect estimates of child mortality using the CEB/CS method

In order to use the CEB/CS method, the number of children ever born and children dead were first adjusted for overall under-reporting determined from the matched women. Indirect estimates of child mortality measures $q(1)$, $q(2)$, $q(3)$, $q(5)$, $q(10)$, $q(15)$ and $q(20)$ were then computed using the CEB/CS method. The appropriate family of life tables to use from the Princeton life tables were determined by comparing the quotient of child mortality (${}_4q_1$) to infant mortality rate (${}_1q_0$) of recent published estimates for Agincourt to those of Princeton life table families (United Nations 1983). The life table family with a quotient closest to that of the Agincourt one is chosen as the most appropriate. Trussell multipliers from the chosen family were then used to convert the proportion of children dead among children ever born to women in successive reproductive age groups 15-19, 20-24,…, 45-49 to yield estimates of $q(1)$, $q(2)$, $q(3)$, $q(5)$, $q(10)$, $q(15)$ and $q(20)$ respectively. The Trussell multipliers were also used to determine the time periods to which the estimates refer. These mortality estimates were then converted to a common measure $q(5)$ through interpolation using the chosen family of life tables.

The estimates obtained as described above will underestimate mortality due to the effects of HIV, as discussed in section 2.2.1.1, and thus require some adjustment. This adjustment was done using the method suggested by Ward and Zaba (2008), with relaxation of the assumptions of epidemic stability by using the prevalence at the times to which the estimates refer in addition to assuming that HIV does not materially affect the computed time location estimates (Darikwa and Dorrington 2011). HIV prevalence rates for Agincourt were obtained from published literature and from population projections for the Limpopo province using the ASSA2008 model.

Accuracy of the child mortality estimates derived using the CEB/CS method were then assessed by comparing them with those of other researchers and estimates from the Limpopo provincial ASSA2008 model.

3.3.4 Indirect estimates of adult mortality using orphanhood method

Timæus and Nunn (1997) revised regression coefficients for the estimation of female adult mortality from data on maternal orphanhood using the Brass general standard to compute the conditional survival life table measures ${}_{n+25}p_{25}$, convert them into a common measure, ${}_{45}q_{20}$ and determine the time periods to which they pertain. Timæus and Nunn (1997) revised regression coefficients were used to estimate female adult mortality because they adjust for HIV-related selection bias. Male mortality was computed using the regression coefficients by Timæus (1992). The estimates derived were then compared to estimates published by Kahn, Garenne, Collinson *et al.* (2007) and from the ASSA2008 model.

4 RESULTS

4.1 UCT survey population numbers

The UCT survey enumerated 44 334 individuals in 9 005 households from 21 villages of the AHDSS. Of these individuals, 3 261 (7.2%) were missing both age and date of birth, 157 (0.4%) did not have sex specified and 72 (0.2%) did not have both age/date of birth and sex specified. Of the 3 190 with no age or date of birth specified but with sex specified, 1 540 were males and 1 649 were females.

4.1.1 Household and population coverage

The UCT survey sought to cover all the 12 153 households in the original 21 villages of Agincourt. Of these, 1 212 were not visited at all, 1 630 households were unsuccessfully enumerated and a further 306 questionnaires were duplicates. This leaves the number of households successfully visited at 9 005. Table 4.1 shows the distribution of the household visits against their respective outcomes for the households successfully surveyed.

Table 4.1 Household visits

Outcome	*Visit 1*	*Visit 2*	*Visit 3*	*No visit number*	*Total*
Visit Successful; Completed	8 250	631	59	65	9 005
Visit Successful; Not Completed	32	5	0	0	37
No-one Present	586	51	0	0	637
No-Adult Available to Answer	82	7	0	0	89
Refusal	5	0	0	0	5

To facilitate the comparison of the de-facto UCT survey population figures with the benchmark standard (AHDSS), individuals from the 9 005 households surveyed by the UCT survey are identified from the 2009 AHDSS census update. Only permanent residents were included for the comparison as these are the ones most likely to have been present to be enumerated by the de-facto UCT survey.

In the 21 villages that were covered by the UCT survey, the AHDSS 2009 census update shows that there were 54 913 permanent residents in 12 153 households, of whom 31 101 were females and 23 812 males, giving an average household size of 4.52 permanent residents. Relative to the whole 'permanent' population in the 21 villages, the UCT survey covered 80.7% of the individuals and 74.1% of households. A total of 3 148 households with 10 328 individuals were either not visited at all or were unsuccessfully visited due to either no one being present or no adult being available to answer or refusal to participate. The average household size of the households not covered by the UCT survey was 3.28, which is significantly smaller than the households covered by the UCT survey, which suggests that small households were more likely not to be covered by the UCT survey.

The 9 005 households in the 2009 AHDSS census updates had a population of 44 351 ‘permanent’ residents. Correspondingly the household size works out to be 4.92 for the UCT survey and 4.95 for the 2009 AHDSS census update. Thus, for the successfully surveyed households, the relative coverage of the UCT survey to the 2009 AHDSS census updates is 99.4% of ‘permanent’ residents. In the UCT survey 96.8% were permanent residents and 3.2% were migrants/visitors whilst in the AHDSS 83.2% were permanent residents and 16.8% were migrants.

A check for individuals reported in two or more different households shows that there were 476 individuals reported as belonging to two different households, which represents about 1% of the enumerated population.

4.1.2 Population distribution by age

Figure 4.1 and Figure 4.2 show the distribution by age and sex of the population in the 9 005 households successfully enumerated by the UCT survey, compared with the AHDSS 2009 population in the households covered by the UCT survey. The UCT survey figures include individuals without age and sex specified in the survey redistributed proportionally. The AHDSS 2009 population excludes household members who were deemed to be migrants during the year 2009.

The numbers of the males and females by age for the UCT survey are very similar to those of the 2009 AHDSS census update at all ages except for the 0-4 age group. There is slight under-reporting of children in the 0-4 age group and this is more pronounced for female children than it is for male children. The shapes of the male and female population distribution by age in the missed households are similar to both the UCT survey and AHDSS, except for females above the age of 25, which show relatively more females in missed households. A check on household headship by sex of these small households shows that they are more likely to be female-headed than male-headed. The two graphs combined confirm the observed underlying male dominated migration trends.

Figure 4.1 Male population distribution

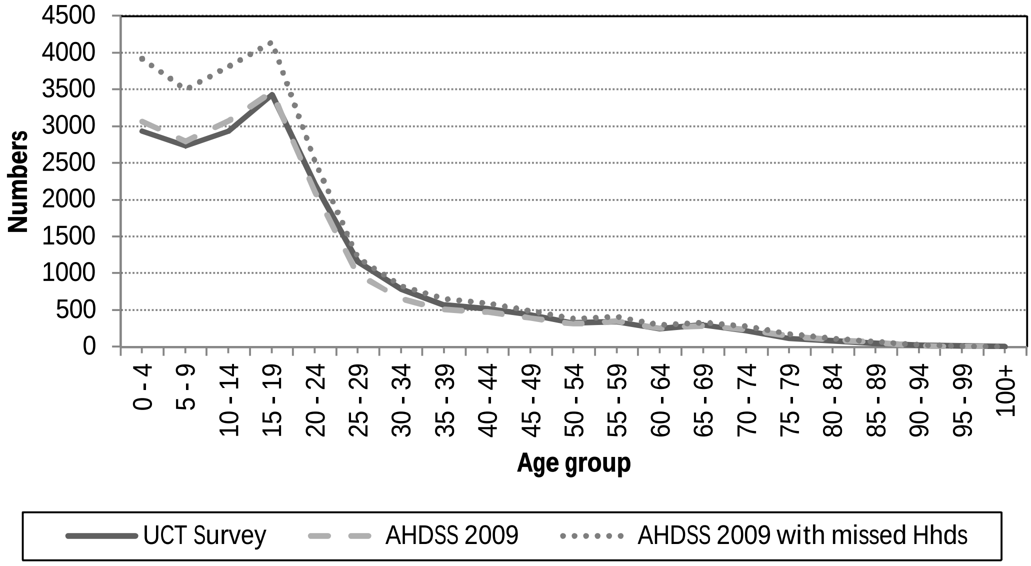


Figure 4.2 Female population distribution

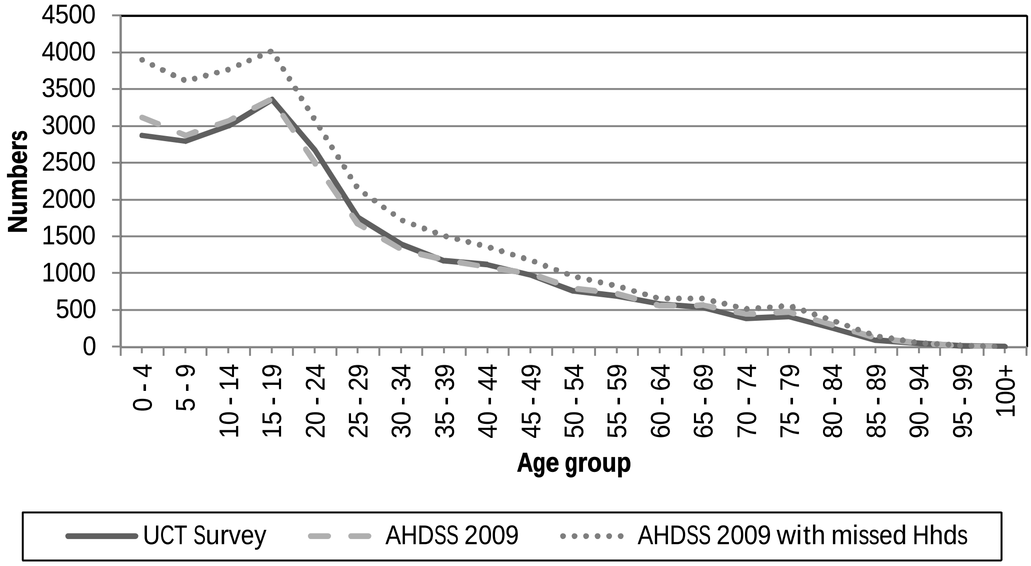
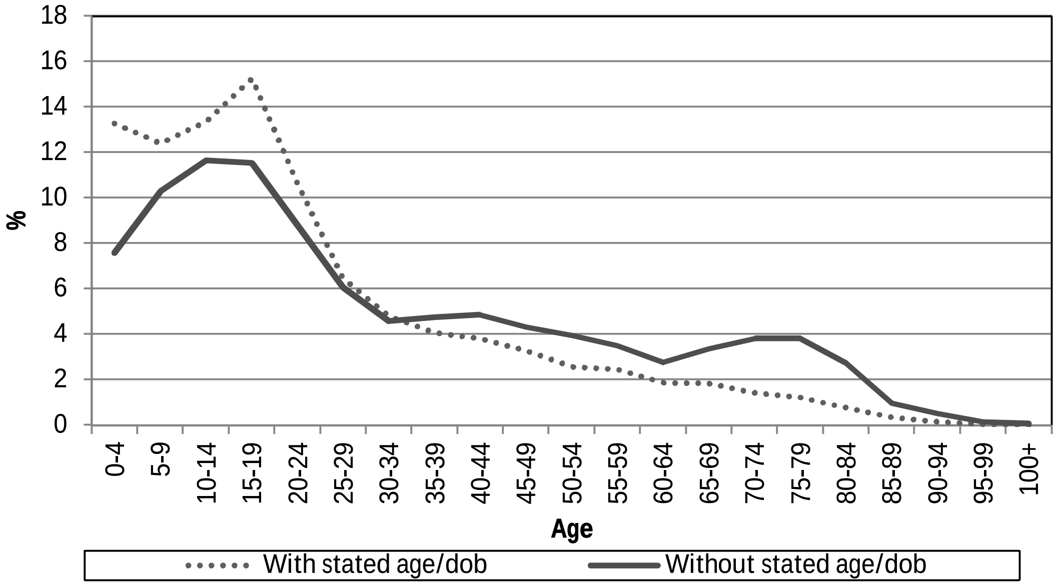


Figure 4.3 compares the age distribution of those for whom age was recorded to those for whom age was not recorded for individuals with known age (matched individuals). Demographers sometimes make the assumption that individuals with unspecified ages in censuses are adults, as children are likely to have their ages reported for them. Figure 4.3 shows that the proportion of those for whom age is not stated is greater than the proportion of those for whom age was stated, for those older than 19 years. This observation supports the assumption that individuals with unspecified ages in censuses are more likely to be adults as children are more likely to have ages reported for them.

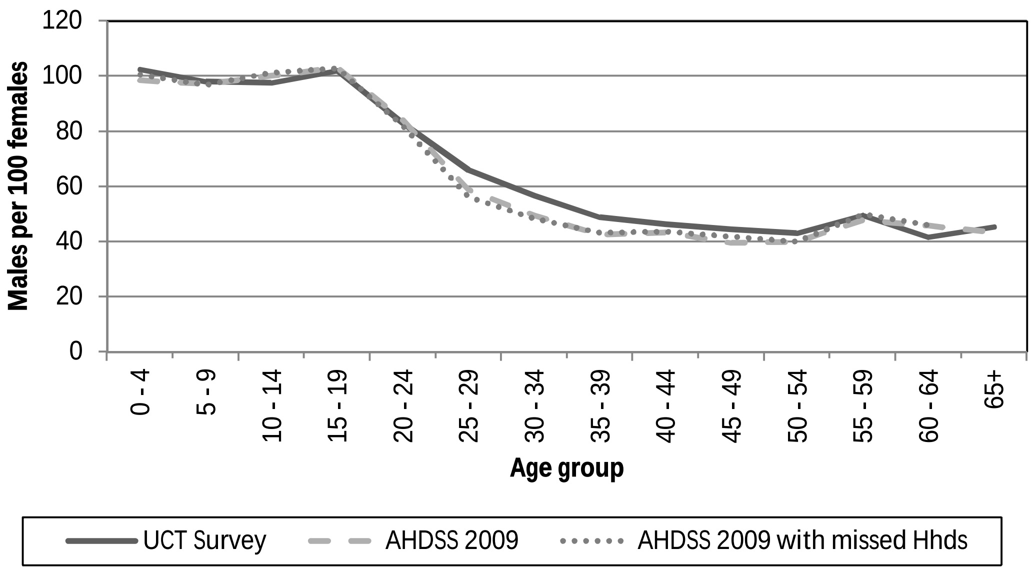
Figure 4.3 Population distribution by true age, specified vs. unspecified age



4.1.3 Sex ratios

The overall sex ratios of the population enumerated by the UCT survey, the AHDSS 2009 census update of the households enumerated by the UCT survey, and the AHDSS 2009 census update of all the households that should have been enumerated by the UCT survey, are reasonably similar, with an overall sex ratio of 78.05, 76.46 and 76.56 males per 100 females respectively. Figure 4.4 shows the sex ratios by age group. The UCT survey had higher sex ratios than the AHDSS for the working age groups (25-54) because it managed to interview some of the migrant labourers who were present during the UCT survey. The sex ratio of migrants in this age group in the AHDSS census update was 150 males to 100 females.

Figure 4.4 Sex ratios



4.1.4 Accuracy of reporting of sex in the UCT survey

Accuracy of reporting of sex is assessed by comparing the reported sex in the UCT survey with the actual sex as captured in the AHDSS. Of the 40 120 individuals in the UCT survey identified in the AHDSS, 39 324 (98.1%) had their sex accurately reported, 650 (1.6%) had their sex either recorded as males when they are females (304) or recorded as females when they are males (346) and for 122 (0.3%) sex was missing. A plot by age and sex for those with missing sex (not shown) reveals that misreporting of sex is independent of both the age and sex of the individual. The accuracy of reporting in the UCT survey is relatively high as sex is a binary variable and scope for reporting/recording error is limited.

4.1.5 Accuracy of reporting of age in the UCT survey

Using the matched individuals, the accuracy of age reporting and reported dates of birth in the UCT survey is assessed by comparing the age or computed age at last birthday at the UCT survey date to the true age at last birthday computed from true date of birth in AHDSS. Henceforth, whenever age is mentioned, it refers to age at last birthday.

Of the 44 351 individuals enumerated in the UCT survey 40 120 have been positively identified in the AHDSS census update. Out of these, 32 414 (80.8%) individuals had date of birth specified, 38 145 (95.1%) had age at last birthday as at the survey date recorded and 30 079 (75.0%) individuals had both age and date of birth recorded in the UCT survey.

Figure 4.5 shows the distribution of the difference between the age at last birthday computed from the stated date of birth and the stated age for those for which both were recorded. A value of zero signifies consistent reporting of date of birth and age. A negative or positive difference, respectively, shows rounding up and rounding down of completed years as at the survey date. Figure 4.5 shows that for about 30% of the population with both date of birth and age at last birthday recorded, the recorded age was age at next birthday compared to the date of birth recorded in the UCT survey.

Figure 4.5 Distribution of the difference between age computed from stated date of birth, and stated age

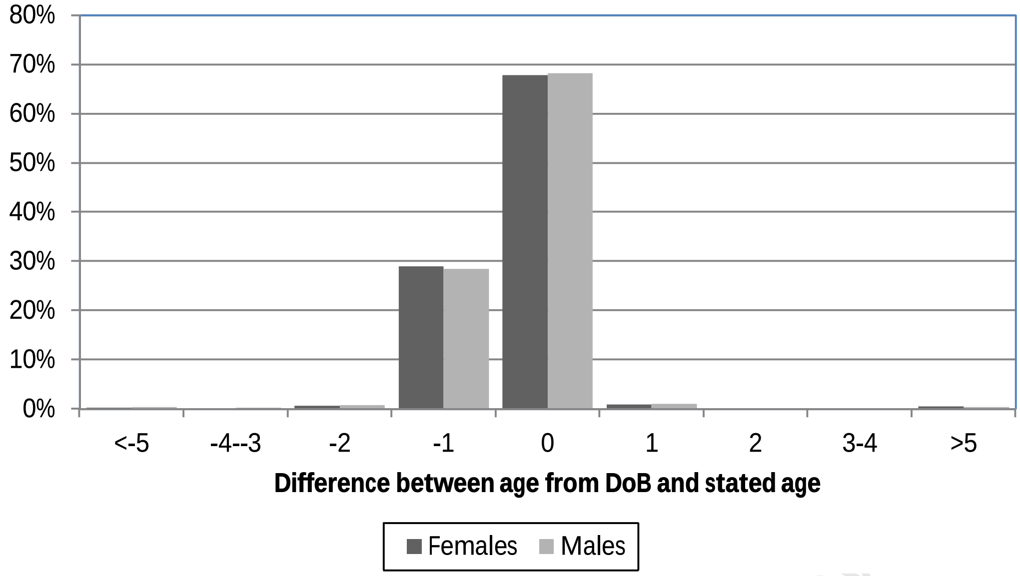


Figure 4.6 shows the distribution of the difference between the true age and the age computed using dates of birth recorded in the UCT survey for the 32 414 individuals with recorded date of birth. The negative numbers on the horizontal axis indicate age exaggeration by the respondents whilst the positive numbers indicate age understatement. Amongst those with specified date of birth in the UCT survey, the majority of individuals had their date of birth reported more accurately than their age at last birthday. The relative accuracy of reporting of dates of birth is confirmed by the symmetrical distribution (about zero) of the difference between true age and age computed from specified dates of birth, with the majority (83.9%) having dates of birth accurately reported.

Figure 4.6 Distribution of the difference between true age and age computed from given date of birth

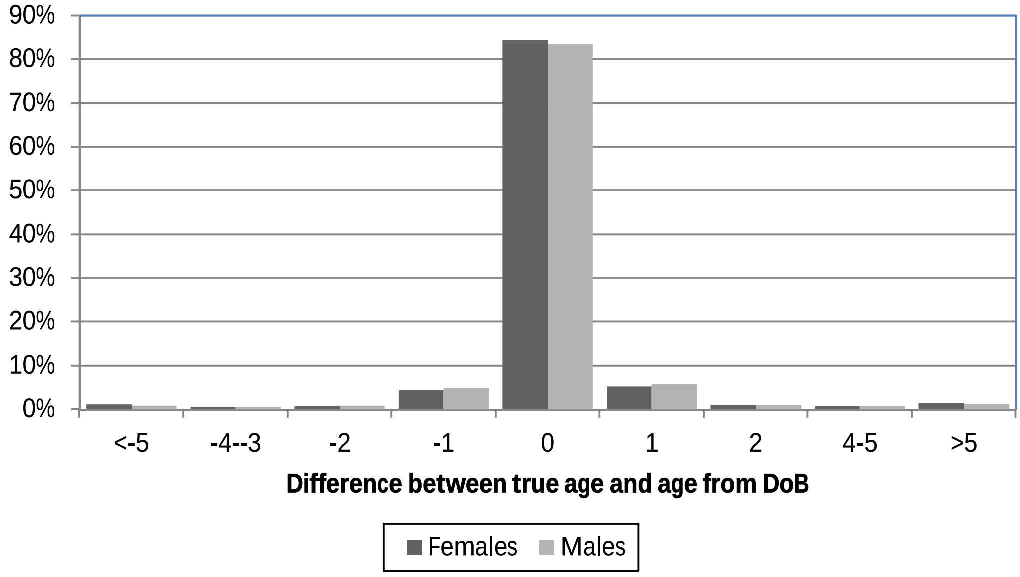
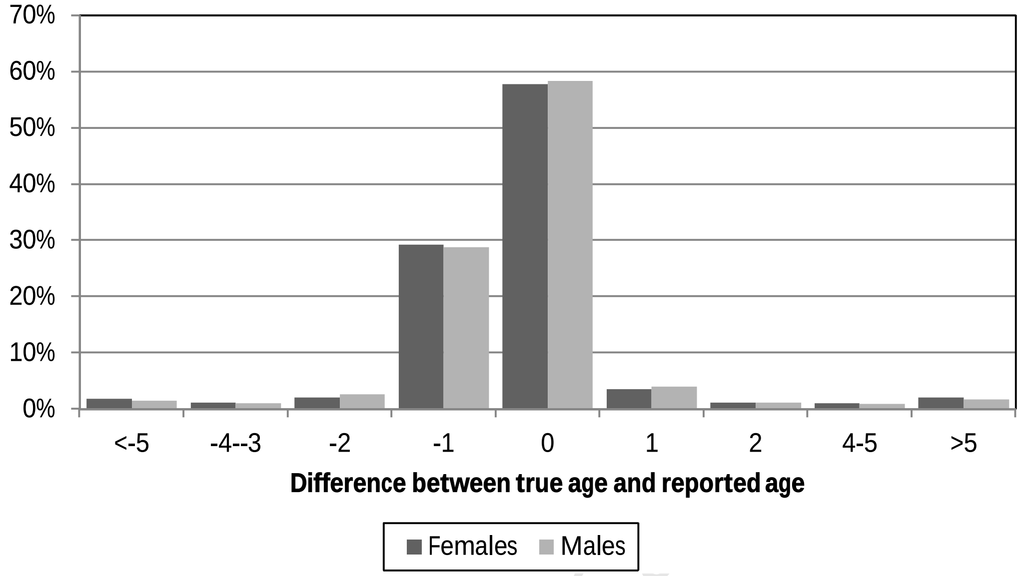


Figure 4.7 shows the distribution of the difference between the true age and the stated age of those with age reported. The graph shows that there is a tendency for respondents/enumerators to round up the ages of those they are reporting on to the age next birthday.

Figure 4.7 Distribution of the difference between true age and stated age



Figures 4.5 to 4.7 taken together show that the reporting of dates of birth is more accurate than reporting of age last birthday and that the accuracy of reporting of ages and dates of birth are similar between males and females.

4.2 Accuracy of children ever born and children dead data

There were 15 087 women in the 15-49 age groups in all the households in the 21 villages of AHDSS that the UCT survey sought to enumerate. However, the UCT survey enumerated 12 458 females aged 15-49 who were eligible to have their children ever born and children who had died reported on. However, only 11 593 had their summary birth history reported and of these only 9 825 were successfully matched.

Figure 4.8 shows the ratio of the number of children ever born (CEB) reported in the UCT survey to the true number as captured by the AHDSS by age group of the mother. The figure shows that the UCT survey failed to record about 5% of the boys and 10% of the girls born to women over 24, nearly 20% of children born to women aged 20-24 and more than 40% of children born to women aged 15-19. It is possible that there is a tendency for this question to miss some of the most recent births.

Figure 4.8 Ratio of reported to actual CEB

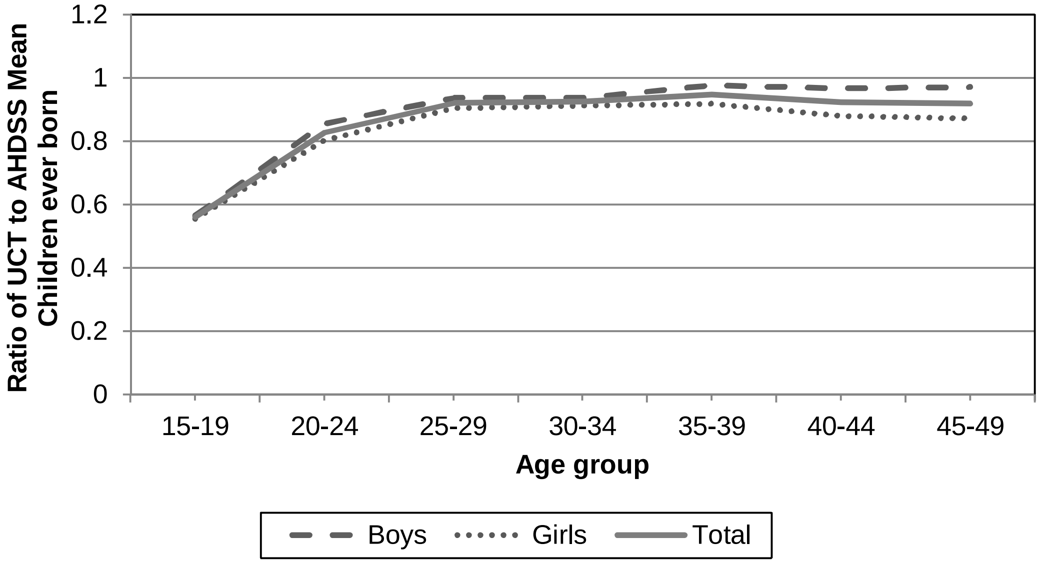


Figure 4.9 shows the mean parity curve by age. As might be expected, this shows increasing parity with increasing age of the mother.

Figure 4.9 Sex specific mean children ever born by age

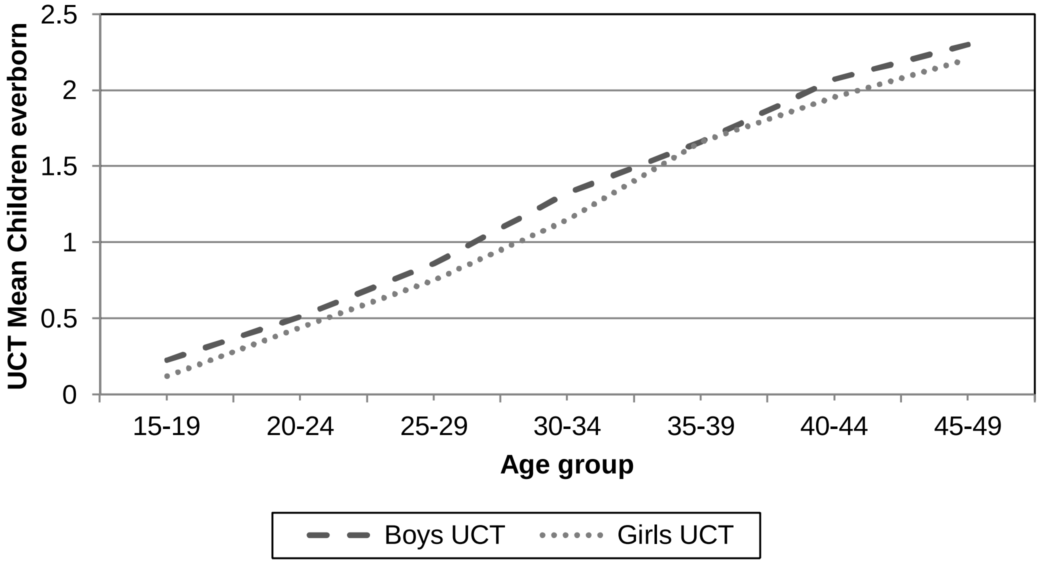
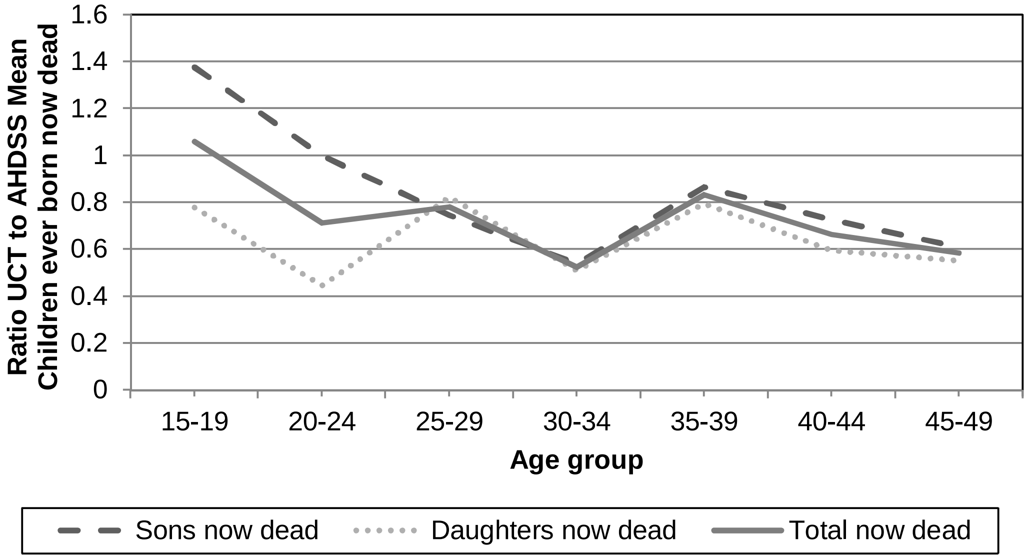


Figure 4.10 shows the ratio of the number of children ever borne now dead (CD) reported in the UCT survey to the true number, as captured by the AHDSS by age group of the mother. The figure shows that between 20% and 40% of the deaths of children of women 20 and older were not captured by the UCT survey. Figure 4.8 and 4.10 taken together show that under-reporting of CD is greater than CEB with the consequence that estimates of mortality rates derived from these data will understate mortality. The data show consistency in terms of the boys dead compared to girls dead, at least for mothers 25 and older.

Figure 4.10 Ratio of reported to actual CD



4.3 Accuracy of orphanhood data

In the UCT survey, the question on maternal orphanhood status was answered for 44 196 people and of these, 39 517 could be matched with AHDSS 2009. For these individuals with IDs, 37 661 (95.3%) had their maternal orphanhood status accurately reported/recorded (7 991 maternal orphans, 29 670 non-orphans) and the maternal orphanhood status was misreported for 1 771 (4.5%). Amongst those with misreported maternal orphanhood status, 1 435 (3.6% of those with IDs) reported their mothers to be alive when actually they were dead, and 336 (0.8% of those with IDs) reported their mother dead whilst they were alive.

Figure 4.11 Proportion of individuals reported not to be orphans but were maternal orphans, by age group

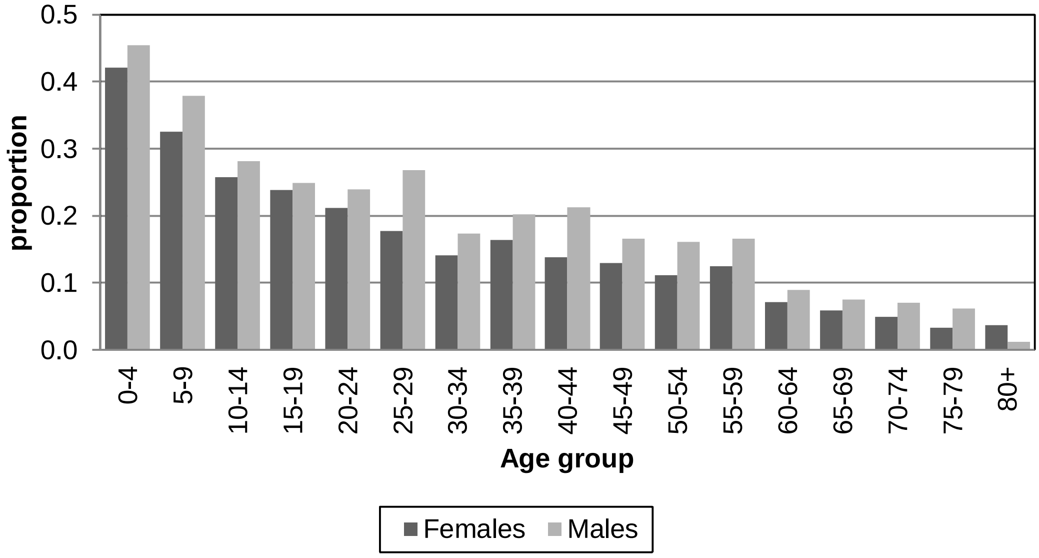
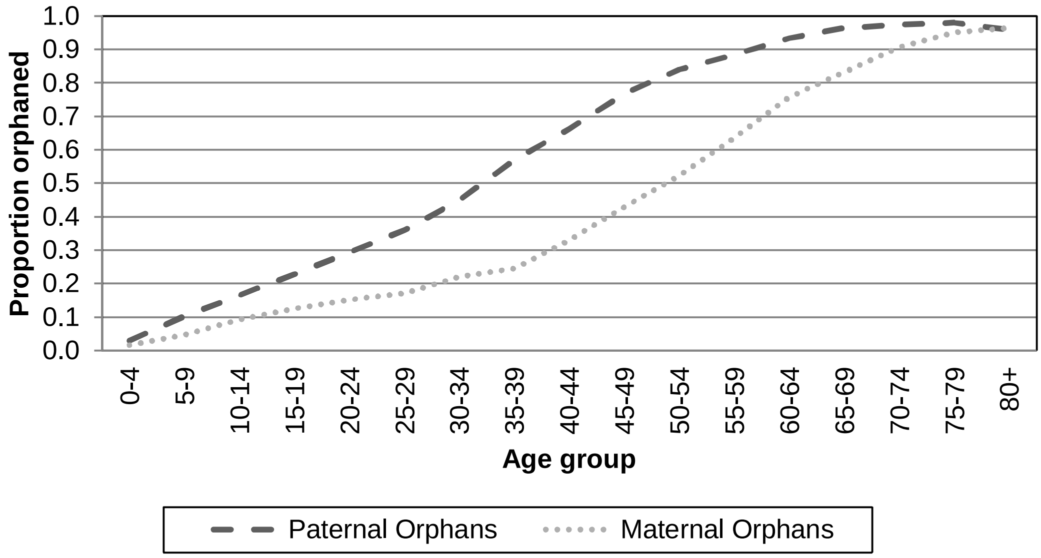


Figure 4.11 shows the proportion by age and sex of individuals reported to have their mothers alive when the mothers were dead. Unsurprisingly the "adoption effect" is apparent from this figure. The figure also shows that the quality of the responses on maternal orphanhood is marginally better for females than males.

For paternal orphanhood status, 44 018 individuals had their paternal orphanhood status reported on and for 333 individuals the respondents did not provide their orphanhood status. Amongst the individuals with specified orphanhood status, 15 129 (34.4%) were classified as paternal orphans, 27 999 (63.6%) paternal non-orphans and for 890 (2%) individuals the respondents didn't know the survival status of their fathers. Unfortunately, the accuracy of these figures on paternal orphanhood cannot be verified using the AHDSS census as it does not collect data on paternal orphanhood.

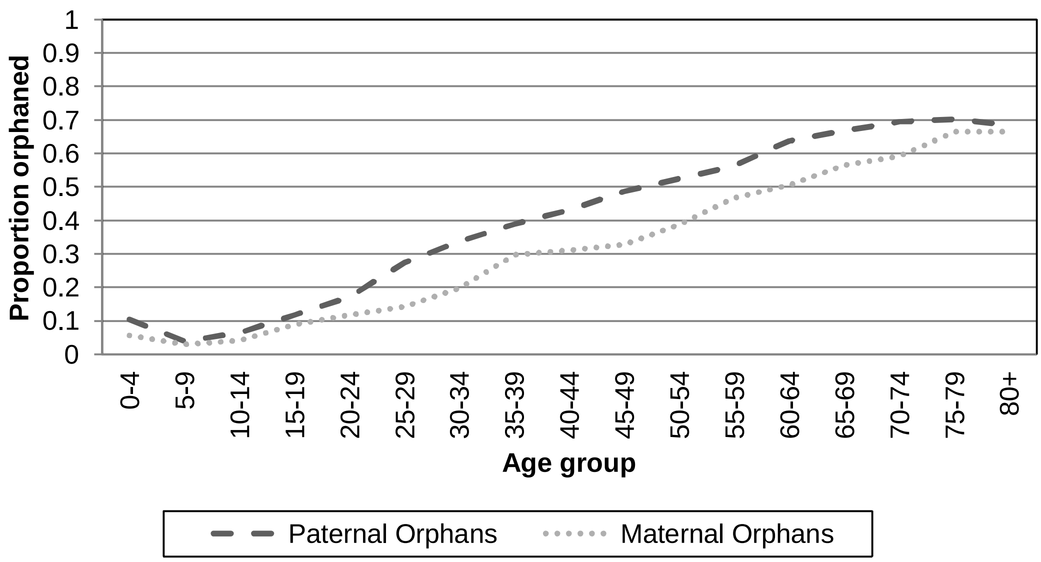
Figure 4.12 shows the proportion of maternal and paternal orphans by age. The maternal and paternal curves both have the expected sigmoidal shape, with a higher proportion of paternal orphans than maternal orphans at all ages except at 0-4 and the open age interval. Although the proportion of paternal orphans is expected to be higher than maternal orphans due to higher male mortality than female mortality, the difference between the two curves is too large to be explained by mortality alone. In addition to mortality, misreporting alive fathers as being dead could also add to the difference, especially in cases where fathers are absent in the upbringing of their children.

Figure 4.12 Proportion orphaned by age, UCT survey, 2009



In addition to the maternal and paternal orphanhood status data as at the survey date, the UCT survey also enquired about the orphanhood status at the time of the 1994 election. A total of 9 216 (20.8%) responded to the maternal orphanhood question and 14 958 (33.7%) answered the paternal orphanhood status question. Figure 4.13 shows the proportion of maternal and paternal orphans by their age at survey date. The graph reveals that for the individuals less than 15 years old, and possibly for older individuals, the respondents might have misinterpreted the reference date as it is impossible for a parent to have died before their child was conceived and thus the results don't appear to be very reliable.

Figure 4.13 Proportion orphaned by age, UCT survey, 1994



4.4 Accuracy of household death data

There were 622 deaths reported by households in the UCT survey. Table 4.2 shows a breakdown by sex of the deaths and whether age was reported. The table shows that 44% of the deaths did not have age stated, which is a relatively high proportion compared to the proportion reporting age for those who were enumerated in the UCT survey.

Table 4.2 Captured deaths in the UCT survey

	Sex			
	Females	***Males***	***Unspecified***	***Total***
With specified age at death	179	169	0	348 (56.0%)
With no specified age at death	116	156	2	274 (44.0%)
Total	295 (47.4%)	325 (52.3%)	2 (0.3%)	622

In order to assess the accuracy of deaths reported by households, the characteristics of the deceased individuals as captured by the UCT survey are compared to the true characteristics as captured by the AHDSS. Of the 622 deaths in the UCT survey, 468 (75%) deaths were identified in the AHDSS while 154 (25%) were not. The unidentified deaths were not matched because they had missing information such as age and/or date of birth and/or date of death. Thus assessment of the accuracy of the reporting of information on deaths is based on these 468 deaths. A check of the matched deaths reveals that 15 deaths were reported in two different households, representing about 3.3%.

The deaths which were not identified in the AHDSS are distributed equally between males and females. Of the unidentified deaths with age, the highest proportion (36.7%) was in the 0-4 age group.

Table 4.3 Deaths not identified in AHDSS

	Sex	
Age at deaths	***Females***	***Males***
Stated	44	35
Not stated	33	42
Total	77	77

Figure 4.14 shows the distribution of the 622 deaths by age and sex. The figure reveals, as might be expected, if a significant number of deaths are due to AIDS, the peak in the number of deaths in young adults for females occurs before that for males. However, it is curious that reports of deaths of women peak at the 25-29 age-group which appears too young for women instead of 30-34 age-group as observed by Bradshaw, Laubscher, Dorrington *et al.* (2004).

Figure 4.14 Population distribution by age at death (of those for whom age was reported)

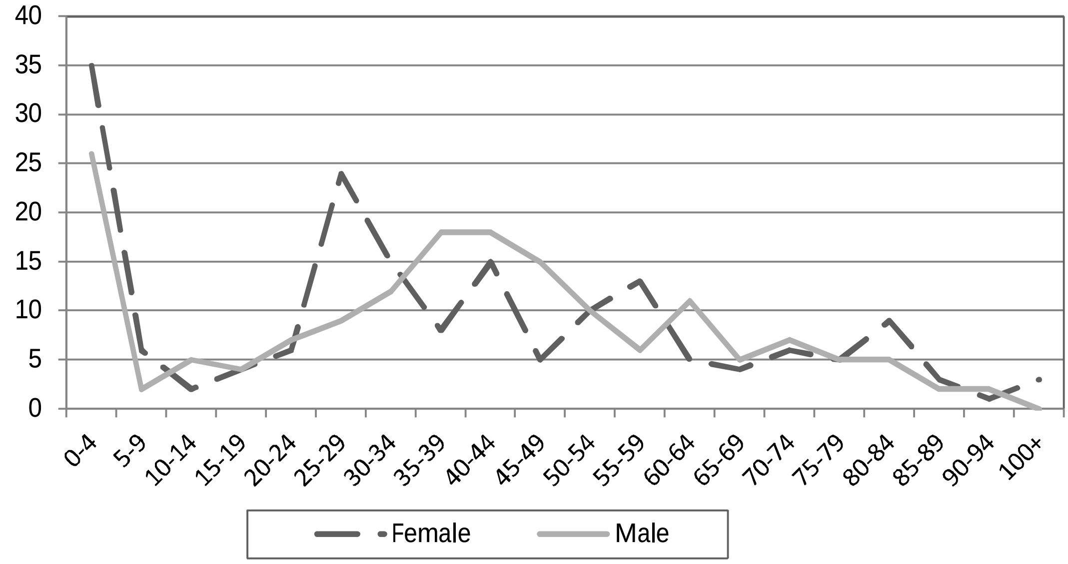


Figure 4.15 shows the distribution of the difference between the true age at death and the stated age at death for the 268 dead individuals which matched and had age at death stated. The figure reveals that although there is some age understatement and exaggeration, of those with misstated age there is an overall tendency to report/record age at death as the age next birthday.

A comparison of the reported dates of death amongst those that match shows that the dates of death (month and year) were accurately reported for all deaths with dates of death specified.

Figure 4.15 Difference between true age at death and stated age at death

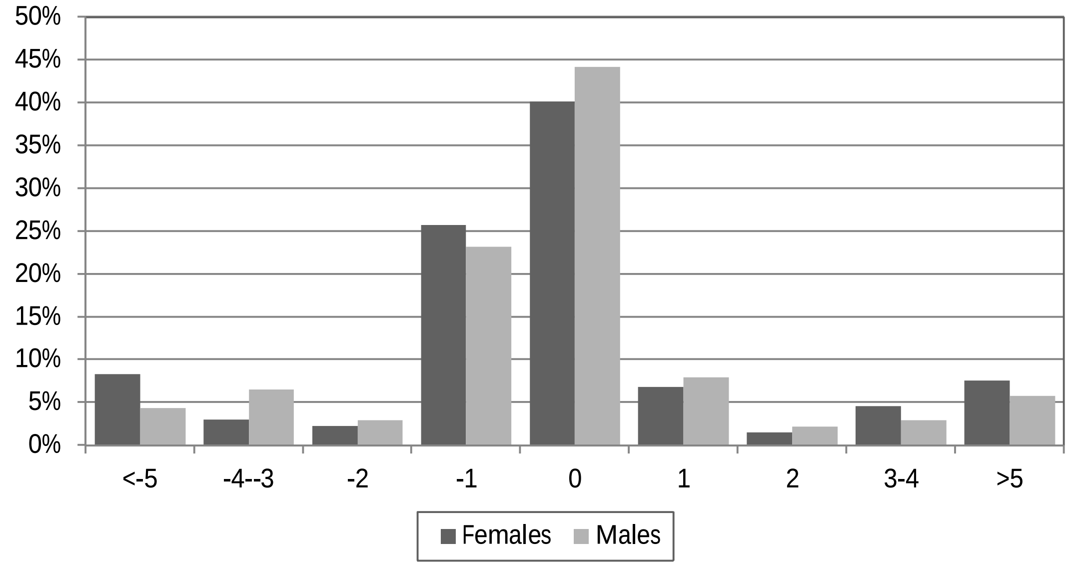


Figure 4.16 shows the ratio of the number of deaths with no stated/recorded ages to the number of deaths with stated/recorded ages in the UCT survey by their true age bands. This figure is based on the deaths that were successfully matched and therefore their true age could be

computed. The figure reveals that deaths of older individuals are more likely to have their ages at death not specified than are deaths of younger individuals. If all deaths matched had their ages specified then the ratio would have been expected to be zero. The ratio of the number of deaths to those with age, by the true age (as captured by the AHDSS) is about 0.5 for deaths under the age of 30. For deaths between age 30 and 64 the ratio is between 0.5 and 1, whilst for deaths at age 65 and above the ratio averages 1.5.

Figure 4.16 Ratio of number of deaths with no stated age to those with stated age by true age

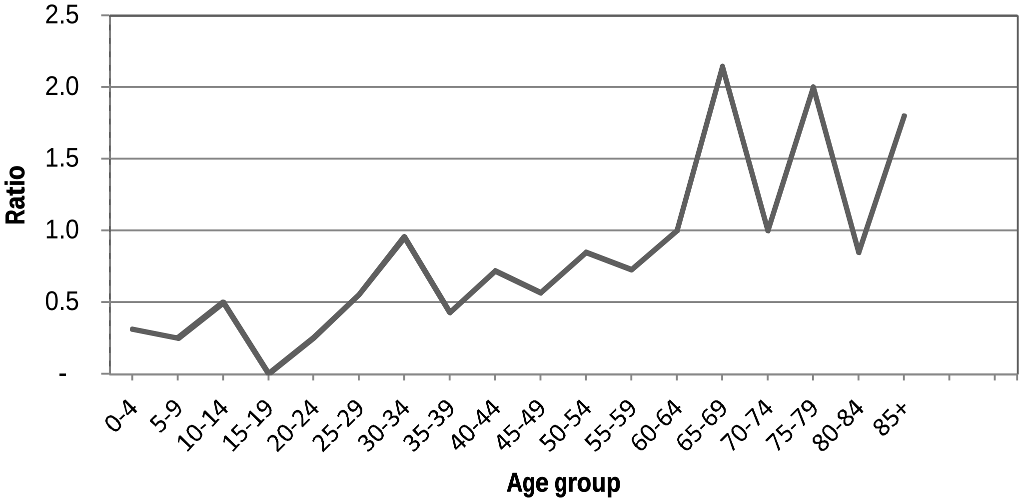


Table 4.4 shows the reporting of sex amongst the reported household deaths. For the 466 deaths which matched and had sex stated, 2 had their sex recorded as male instead of female and 4 had their sex recorded as females instead of males.

Table 4.4 Accuracy of reporting of sex in the UCT Survey

		UCT sex			
		Unspecified	***Female***	***Male***	***Total***
True sex (AHDSS)	Not matched	0	77	77	154
	Female	2	214	2	218
	Male	0	4	246	250
Total		2	295	325	622

4.4.1 Accuracy of death registration reporting in the UCT survey

Of the 622 deaths reported in the UCT survey, 532 deaths (85.5%) were reported to have been registered, 73 (11.7%) were reported not to have been registered, and for 17 (2.7%) the respondents didn't know or didn't report whether the deaths were registered or not. Registration of male deaths (88.1%) was reported to be better than female deaths (83.4%). Table 4.5 shows the breakdown of the responses to the death registration question by sex and whether the age at death was reported.

Table 4.5 Death registration as captured by the UCT Survey

Age at death	*Sex*	*Death registration status*			*Total*
		Don't know	*No*	*Yes*	
Specified	Males	1	20	148	169 (27.2%)
	Females	4	17	158	179 (27.8%)
Unspecified	Males	8	25	123	156 (25.1%)
	Females	3	11	102	116 (18.6%)
Unspecified	Unspecified	1	0	1	2 (0.3%)
Total		17 (2.7%)	73 (11.8%)	532 (85.5%)	622

Death registration was reported to have been incomplete for deaths under age 1, 1-4, 20-24, 30-34 (males only), 35-39 (females only), 55-59 and 80+ (males only) age groups as shown in Figure 4.17, for all other ages registration was reported as complete.

Figure 4.17 Death registration by age group and sex

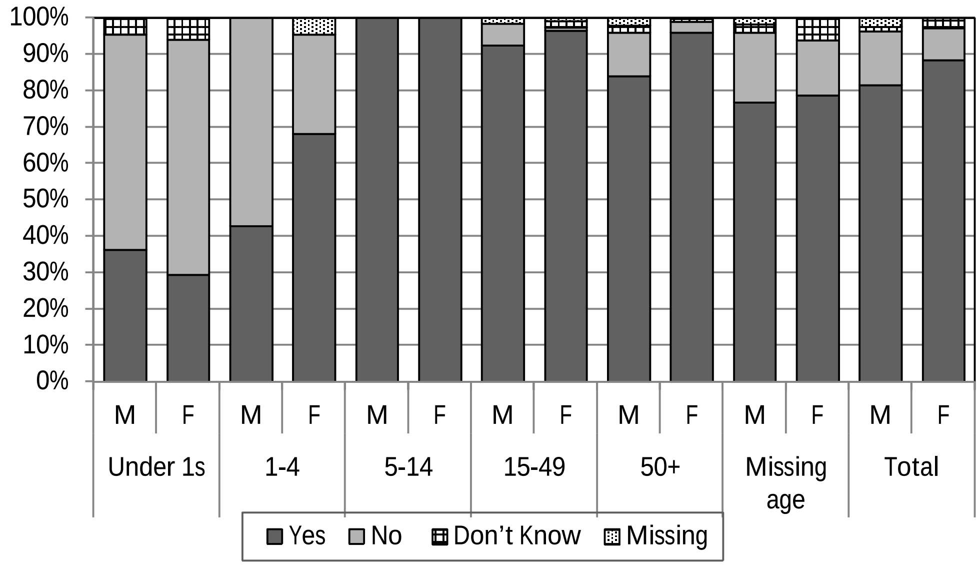


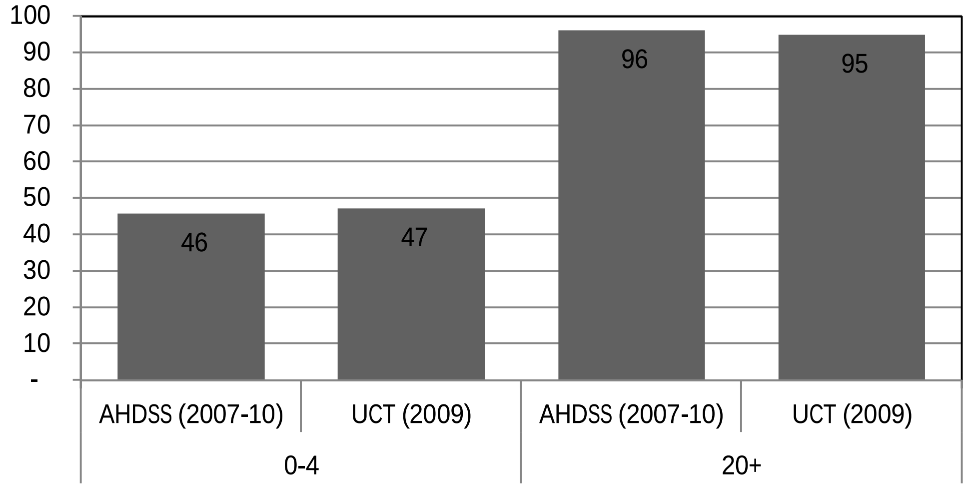
Table 4.6 shows the accuracy of reporting of death registration in the UCT survey compared to AHDSS. The table shows that although only 88.7% of the 468 deaths have the death registration status reported accurately, there is overall consistency in the proportion of deaths reported as having been registered between AHDSS and the UCT survey at 88.0% and 89.1% respectively. However, about 8% of the deaths confirmed as not registered and registered by both the UCT survey and AHDSS show an inconsistency suggesting that there are potentially some problems with the UCT survey deaths data.

Table 4.6 Accuracy of reporting of death registration in the UCT survey

	UCT Survey				
AHDSS	***Not recorded***	***Don't Know***	***No***	***Yes***	***Total***
Not Known	1	1	3	9	14 (3.0%)
No	0	2	21	19	42 (9.0%)
Yes	2	4	17	389	412 (88.0)
Total	3 (0.6%)	7 (1.5%)	41 (8.8%)	417 (89.1%)	468

Figure 4.18 compares the registration of deaths estimated from the UCT survey with estimates by Kahn, Collinson, Gómez-Olivé *et al.* (2012). Assuming that the three-year average computed from AHDSS reflects the actual death registration in 2009, the figure shows that there is an overall concurrence of the UCT estimates to those by Kahn, Collinson, Gómez-Olivé *et al.* (2012).

Figure 4.18 Death registration by wide age groups



4.4.2 Accuracy of death categorisation in the UCT survey

Table 4.7 shows the breakdown of whether deaths were reported as accidental or not in the UCT survey. For females 89.15% of the deaths were non-accidental, 8.81% were accidental whilst the remaining 2.04% of the deaths did not have the cause of death classified as accidental or not. For males, 81.5% were non-accidental and 17.2% were accidental deaths with the remaining 1.3% not stated. Overall, 13.1% of the deaths were due to accidents.

Table 4.7 Reporting of accidental and non-accidental deaths

Sex	*Accidental death*			*Total*
	Not stated	*No*	*Yes*	
Female	6 (2%)	263 (89%)	26 (9%)	295
Male	4 (1%)	265 (82%)	56 (17%)	325
Unspecified	0 (0%)	2 (100%)	0 (0%)	2
Total	10 (2%)	530 (85%)	82 (13%)	622

Table 4.8 shows the accuracy of the categorisation of deaths as accidental or not. The respondents (enumerators) mis-reported (mis-recorded) that 48 deaths were accidental when they were not, and 2 were non-accidental when they were in fact accidental deaths. 16% of the deaths with matches were migrant deaths.

Table 4.8 Accuracy of cause of death categorization: accidental or non-accidental

		UCT accidental death question response			
		Not stated	*No*	*Yes*	*Total*
AHDSS accidental death	Not matched	7	124	23	154
	No	3	404	48	455
	Yes	0	2	11	13
Total		10	530	82	622

Table 4.9 shows the distribution of the responses to the question on whether women were pregnant or had given birth within 40 days of their death as reported (recorded) by the household respondents (enumerators).

Table 4.9 Reporting of pregnancy in women at time of death

Pregnancy at death	*All women*	*Women aged 15-49(with stated age)*
Not stated	112	7
Don't Know	4	2
No	174	64
Yes	5	4
Total	295	77

Table 4.10 shows the accuracy of the reporting of maternal deaths in the UCT survey amongst the 138 female deaths with matches. Of these, the UCT survey has 2 deaths reported as maternal deaths, while the AHDSS has 3 deaths recorded as maternal deaths. A closer look shows that the deaths reported in the UCT as maternal are not. Though the deaths recorded are too few to draw any conclusions about the accuracy with regards maternal mortality, they indicate that the maternal reports in the UCT survey cannot be relied on.

Table 4.10 Accuracy of categorization of deaths: pregnancy related deaths in the UCT Survey

		UCT			
		Don't	***No***	***Yes***	***Total***
AHDSS	No	3	130	3	136
	Yes	0	2	0	2
Total		3	132	3	138

4.5 Effects of respondent characteristics on accuracy of responses

4.5.1 Univariate analysis

As would be expected the majority of the respondents (91%) were permanent residents of Agincourt whilst the remaining 9% were considered to be migrants. Unsurprisingly, the majority (75.6%) of the respondents were females, whilst the remaining 24.4% were males. Table 4.11 shows the age and household size distribution of the respondents. More than half (59.5%) of the respondents were household heads.

Table 4.11 Age and household size distribution of the respondents

Age distribution		***Household size***	
Age group	**Per cent**	**Category**	**Per cent**
15-24	20.9	*Small*	20.8
25-44	39.8	*Medium*	43.7
45-59	19.3	*Large*	35.5
60+	19.7	*Total*	*100.0*
Missing	0.3		
Tota	*100.0*		

4.5.1 Bivariate analysis and logistic regression models

The dependent variables for age, CEB and CD accuracy, for each respondent can each take one of three outcomes, "understated", "accurate" and "overstated", as explained in section 3.2.4. However, for the bivariate analysis presented below the two categories "understated" and "overstated" are collapsed into one category, namely "inaccurate". In addition, the "inaccurate" category of the dependent age variable includes all responses for which respondents did not provide an age or date of birth variable in the UCT survey. The analysis excludes those individuals for whom matching with the AHDSS was not possible due to incorrect or lack of information, such as age or dates of birth in the UCT survey or because they were visitors at the time of the survey.

4.5.1.1 Effects of respondent characteristics on accuracy of reporting of age

Cross tabulations of the accuracy of reporting of age, CEB and CD; and the recoded variables reveal that the accuracy of these vary with the level of each recoded variable as shown in Tables 4.12-4.14.

As would be expected, respondents give more accurate reports of their own ages than when reporting for other household members (76.2% vs. 71.9%). This trend generally persists for all the categories of the respondent variables as shown in Table 4.12**.**

In terms of age the overall picture is that middle aged (25-44) respondents give more accurate reports of their ages and ages of other household members than do younger and older respondents. Respondents above the age of 60 are the worst respondents. Respondents in the 45-59 age group are odd in that in this age group, respondents give better reports on other household members' ages than their own ages.

Compared to males, female respondents report more accurately the ages of members of their households. The proportion of accurate age reports from female respondents (for themselves and other household members) is 8.3 percentage points higher than those from males. There are notable differences between male and female respondents when reporting age for themselves or other household members. There is a 9.8 and 3.2 percentage point difference for males and females, respectively, between the percentage of respondents reporting age accurately for themselves and on behalf of others.

Contrary to expectations, respondents from small households report the least accurately on the age of members of their households compared to respondents from medium to large size households (3 or more individuals).

As might be expected, permanent respondents report more accurately the ages of their household members compared to migrant respondents. Both migrant and permanent respondents report more accurately their ages than they do for other household members.

Household heads report ages with almost the same accuracy when reporting for themselves or on behalf of other household members. Other respondents who are not household heads report more accurately their ages than they do ages of other household members.

Table 4.12 Percentage accurately reporting age by:

	a. Age		
	Self	*Other*	*Total*
15-24	81.5	70.3	72.5
25-44	82.8	77.4	78.5
45-59	69.6	72.5	71.9
60+	64.6	58.3	60.0
Total	*76.2*	*71.9*	*72.9*

	b. Sex		
	Self	*Other*	*Total*
Males	73.5	63.7	66.2
Females	77.1	73.9	74.5
Total	*76.2*	71.9	72.8

	c. Household Size		
	Self	*Other*	*Total*
Small	69.3	56.5	65.0
Medium	78.1	72.3	73.8
Large	77.9	72.4	73.1
Total	*76.2*	*71.9*	*72.8*

	d. Residence		
	Self	**Other**	**Total**
Migrant	72.4	68.3	69.2
Permanent	76.6	72.2	73.1
Total	*76.2*	*71.9*	*72.8*

	e. Household head		
	Self	**Other**	**Total**
Head	72.8	72.6	72.6
Other	81.5	71.3	73.0
Total	76.2	71.9	72.8

4.5.1.2 *Effects of respondent characteristics on accuracy of reporting of CEB*

There are marginal differences between respondents reporting birth history for themselves and for other household members. Whilst, overall, respondents report slightly more accurately on other people's birth histories than their own (54.3% vs. 53.9%), this is not the case when one looks at the categories for each variable as shown in Table 4.13**.**

In terms of age the overall picture is that young to middle-aged (15-44) respondents report more accurately their birth histories and for other household members than older respondents. Also, respondents in this age group give better accounts of their own birth histories than they do for other household members. This trend is reversed for respondents who are older than 45 years.

The sex variable can only be analysed for respondents reporting birth histories for other household members as the question is only applicable to females. Compared to males, female respondents report more accurately the birth histories for members of their households. The proportion of accurate age reports from female respondents is 3.4 percentage points higher than that of males.

Contrary to expectations, respondents from small households report the least accurately on the age of members of their households compared to respondents from medium to large size households (3 or more individuals). However, respondents from small and medium households give more accurate reports of their birth histories than when reporting for other household members.

Again, as might be expected, permanent respondents report more accurately their birth histories and those of other household members compared to migrant respondents. However, comparing permanent residents' own birth history reports with reports for other household members shows that they report marginally more accurately birth histories for other household members than their own. On the other hand, as might be expected, migrants report more accurately their birth histories than when reporting for other household members.

Household heads report birth histories with slightly lower accuracy when reporting for either themselves or on behalf of other household members. Household heads give more accurate birth histories for other household members than their own. Other respondents who are not household heads report more accurately their birth histories than they do ages of other household members.

Table 4.13 Percentage accurately reporting CEB by

a. Age	*Self*	*Other*	*Total*
15-24	62.5	52.7	56.5
25-44	58.6	51.6	56.1
45-59	37.0	58.0	49.9
60+	3.4	57.2	47.9
Total	*53.9*	*54.3*	*54.1*

b. Sex	*Self*	*Other*	*Total*
Males	N/A	51.9	51.9
Females	53.9	55.3	54.5
Total	*53.9*	*54.3*	*54.1*

c. Household Size	*Self*	*Other*	*Total*
Small	46.4	45.2	46.2
Medium	55.1	51.6	53.7
Large	54.1	55.7	55.1
Total	*53.9*	*54.3*	*54.1*

d. Residence	*Self*	*Other*	*Total*
Migrant	52.6	48.7	50.4
Permanent	54.0	54.9	54.5
Total	*53.9*	*54.3*	*54.1*

e. Household head or not	*Self*	*Other*	*Total*
Head	51.2	55.8	53.5
Other	56.9	52.9	54.8
Total	*53.9*	*54.3*	*54.1*

4.5.1.3 *Effects of respondent characteristics on accuracy of reporting of CD:*

Contrary to expectations, respondents give less accurate reports of their own dead children (CD) compared to when they report for other household members (88.7% vs. 91.7%). This trend generally persists for all the categories of the respondent variables as shown in Table 4.14.

In terms of age the overall picture is that the accuracy of reported CD (reports for both themselves and other household members) decreases as the age of the respondent increases. With the exception of those in the 15-24 age-group, respondents give more accurate reports of other household members' CD than their own. Respondents above the age of 44 report noticeably less accurately their CD than when reporting for other household members.

Female respondents report more accurately on CD for other household members than their male counterparts. As noted above, the effet of the sex independent variable on the quality of birth history reports can only be assesed for other household members as the question is only applicable to females.

Again, contrary to expectations, respondents from small households report the least accurately on CD for both themselves and other household members. The accuracy of the CD reports increases with increasing household size. Across all household sizes respondents report more accurately for other household members than when reporting on their own CD.

As might be expected, permanent respondents report more accurately their CDs and those of other household members compared to migrant respondents. Both migrants and permanent respondents report more accurately for other household members than when reporting on their own CD.

Household heads report ages with almost the same accuracy when reporting for themselves as reporting on behalf of other household members. Other respondents who are not household heads report more accurately their ages than they do ages of other household members.

Table 4.14 Percentage accurately reporting CD by:

a. Age	*Self*	*Other*	*Total*	*b. Sex*	*Self*	*Other*	*Total*	*c. Household Size*	*Self*	*Other*	*Total*
15-24	97.1	88.8	91.9	Males	*N/A*	*89.7*	*89.7*	*Small*	84.9	87.7	85.4
25-44	90.6	92.0	91.1	Female	*88.7*	*92.5*	*90.3*	*Medium*	88.9	89.7	89.2
45-59	78.4	94.2	88.2	*Total*	*88.7*	*91.7*	*90.3*	*Large*	89.2	92.7	91.4
60+	59.5	92.4	86.7					*Total*	*88.7*	*91.7*	*90.3*
Total	*88.7*	*91.7*	*90.3*								

d. Reside	*Self*	*Other*	*Total*	*e. Household head or not*	*Self*	*Other*	*Total*
Migrant	88.9	89.3	89.1	Head	*87.0*	*93.2*	*90.1*
Permanent	88.7	91.9	90.4	Other	*90.6*	*90.3*	*90.4*
Total	*88.7*	*91.7*	*90.3*	*Total*	*88.7*	*91.7*	*90.3*

4.5.2 Multinomial logistic regression models (MNLM)

This section presents the results from the multinomial logistic models used to determine the respondent characteristics that have a bearing on accuracy of responses. In particular, it focuses on three variables: age, children ever borne (CEB) and children who are now dead (CD). An important point that should be highlighted in this section is that the interpretation of accuracy of CD is somewhat different from that presented in section 4.2. Here the focus was on whether the respondents gave a correct response with respect to answering the question about deaths of children to women of reproductive ages in the household. Thus if, for example, all women of reproductive ages in a specific household had not experienced the death of their child(ren) and the respondent stated the CD as zero for all these women, this would be regarded as accurate for this section. Given the relatively small number of women who have experienced death of their child(ren) in the ages considered here, this would result in a high percentage being accurate for CD. On the other hand, a fair share of women in the reproductive ages had given birth(s). Thus there was a higher scope for reporting error for CEB than CD. It should, however, be pointed out that since most children don't die in childhood this measure of accuracy of CD significantly misrepresents the accuracy of reporting of deaths of children who have died. As pointed out previously, these deaths were signficantly under-reported.

Tables 4.15 and 4.16 show the results from the MNLM used to assess the respondent characteristics that have a bearing on the accuracy of reporting of age, CEB and CD. The two tables separately present the results from respondents reporting on themselves and reporting on other household members, respectively. The tables show the relative risk ratios (RRR) of respondents' tendency to understate or overstate age, CEB and CD relative to reporting accurately on them with respondent characteristics as predictors for the model. The multinomial logistic model estimates two models, where each equation is relative to the referent category, here taken be the "accurate" category. The two models, which are run separately with age, CEB and CD as the outcome variables, respectively, estimate the log-odds of a respondent to "understate" or "overstate" a response relative to giving an "accurate" response.

When the model is written in an exponentiated form where the predictor of interest is evaluated at x and at $x+\Delta x$ for outcome j (j= "understate" or "overstate") relative to referent group ("accurate"), where Δx is the change in the predictor we are interested in (Δx is traditionally set to one) while the other variables in the model are held constant. If the predictor of interest is a categorical variable, this can be evaluated at one level of the predictor relative to a chosen referent category. For example, the categorical variable sex can be evaluated for males with the referent category being females. If we then take their ratio, the ratio would reduce to the ratio of two probabilities, the relative risk ratio (RRR). In this sense, the exponentiated multinomial logistic coefficient provides an estimate of relative risk. The RRR of a coefficient indicates how the risk of the outcome falling in the comparison group ("overstated" or "understated" compared to the risk of the outcome falling in the referent group ("accurate") changes with the variable in question. A relative risk ratio greater than unity indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group increases as the variable increases. In other words, the comparison outcome is more likely. On the other hand, a RRR less than unity indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group decreases as the variable increases (Chen, Ender, Mitchell *et al.* 2008).

Whilst the bivariate analysis used age and household size variables recoded to grouped categories, these were treated as discrete variables with unit step in the MNLM. Table 4.15 and 4.16, respectively, summarise the logistic output for the three questions, separately based on whether respondents were reporting for themselves or on behalf of other household members.

Table 4.15 Multinomial logistic regression results for respondents reporting for themselves

	Age			*CEB*			*CD*		
	RRR	***Std***	***P>|z|***	***RRR***	***Std Er.***	***P>|z|***	***RRR***	***Std Er.***	***P>|z|***
Age									
Understated									
Increase by 1year	1.021	0.002	0.000**	1.021	0.005	0.000**	1.076	0.010	0.000**
Overstated									
Increase by 1year	1.008	0.002	0.000**	1.063	0.007	0.000**	1.073	0.012	0.000**
Household size									
Understated									
Increase by 1year	NS	NS	NS	1.028	0.015	0.068*	NS	NS	NS
Overstated									
Increase by 1year	NS	NS	NS	0.969	0.021	0.151	NS	NS	NS
Sex									
Understated									
Males	1.249	0.103	0.007**	N/A	N/A	N/A	N/A	N/A	N/A
Females	BC	BC	BC	BC	BC	BC	BC	BC	BC
Overstated									
Males	1.408	0.103	0.000**	N/A	N/A	N/A	N/A	N/A	N/A
Females	BC	BC	BC	BC	BC	BC	BC	BC	BC
Residence Status									
Understated									
Migrant	1.781	0.207	0.000**	1.074	0.155	0.621	0.870	0.268	0.652
Permanent	BC	BC	BC	BC	BC	BC	BC	BC	BC
Overstated									
Migrant	1.246	0.140	0.050	1.595	0.289	0.010**	2.104	0.573	0.006**
Permanent	BC	BC	BC	BC	BC	BC	BC	BC	BC

Note
** means significant at both 1% and 5%, * means significant at 5% only

Table 4.16 Multinomial logistic regression results for respondents reporting for other household members

	Age			*CEB*			*CD*		
	RRR	***Std***	***P>|z|***	***RRR***	***Std Er.***	***P>|z|***	***RRR***	***Std***	***P>|z|***
Age									
Understated									
Increase by	1.013	0.001	0.000**	NS	NS	NS	NS	NS	NS
Overstated									
Increase by	1.004	0.001	0.000**	NS	NS	NS	NS	NS	NS
Household									
Understated									
Increase by	NS	NS	NS	NS	NS	NS	NS	NS	NS
Overstated									
Increase by	NS	NS	NS	NS	NS	NS	NS	NS	NS
Sex									
Understated									
Males	1.481	0.077	0.000**	1.054	0.080	0.489	NS	NS	NS
Females	BC	BC	BC	BC	BC	BC	BC	BC	BC
Overstated									
Males	1.578	0.062	0.000**	1.593	0.223	0.001**	NS	NS	NS
Females	BC	BC	BC	BC	BC	BC	BC	BC	BC
Residence									
Understated									
Migrant	1.249	0.095	0.003**	NS	NS	NS	NS	NS	NS
Permanent	BC	BC	BC	BC	BC	BC	BC	BC	BC
Overstated									
Migrant	1.204	0.070	0.001**	NS	NS	NS	NS	NS	NS
Permanent	BC	BC	BC	BC	BC	BC	BC	BC	BC

Note

** means significant at both 1% and 5%, * means significant at 5% only

4.5.2.1 *MNLR model investigating the effects of respondent characteristics on accuracy of reporting of age*

Overall; age, sex and residence status of the respondent were statistically significant predictors of the accuracy of the reported ages. Household size and whether a respondent was the household head or not, were both statistically insignificant at both the 5% and 10% level of significance.

The relative risk ratio (RRR) for a one year increase in age predicting a respondent to fall in the "understated" relative to the "accurate" category for age reporting would be expected to increase by a factor of 1.021 when reporting for themselves or 1.013 when reporting for other household members, given the other variables in the model are held constant. On the other hand, the RRRs of predicting a respondent falling in the "overstated" relative to the "accurate" category would be expected to increase by a factor of 1.008 when reporting for themselves or 1.004 when reporting for other household members, controlling for other variables in the model. More generally, we can say that the older the respondent's age, the more accurate they would be expected to report an age. Respondents are more likely to understate (for both reporting on themselves and other household members) than to overstate ages.

Although the RRRs appear to be very small and close to unity, given the wide range of respondents' ages, they are likely to have huge cumulative effects. For example, huge differences in accuracy of reporting of ages are likely to be observed when comparing respondents who are on the extremes of the age spectrum.

Compared to females, males were 1.249 (reporting on themselves) and 1.481 (reporting on other household members) times more likely to report understated than accurate ages, given that the other variables in the model are held constant. Male respondents were also more likely to report overstated ages compared to female respondents. In addition, for both sexes respondents were more likely to report more accurately their ages than when reporting for other household members.

Migrants are 1.781 and 1.578 times more likely to understated ages when reporting on their ages and when reporting on other household members' ages, respectively. On the other hand, respondents are 1.204 times more likely to overstate age when reporting for others. However, there are no tendencies for respondents to overstate their own ages.

4.5.2.2 *MNLR model investigating the effects of respondent characteristics on accuracy of reporting of children ever born (CEB)*

Accuracy of reports on birth histories seem to be associated with the four predictor variables: age, household size, residence status and sex. The first three predictors are only significant for respondents reporting their own birth history and not when reporting for other household members. The sex predictor variable is statistically significant though only applicable to respondents reporting on other household members' birth histories as explained in section 4.5.1.2.

The relative risk ratio (RRR) for a one year increase in age predicting a respondent to fall in the "understated" relative to the "accurate" category would be expected to increase by a factor of 1.021 when reporting their own birth histories, given the other variables in the model are held constant. On the other hand, the RRRs to predicting a respondent to fall into "overstated" relative to the "accurate" category would be expected to increase by a factor of 1.063 when reporting their birth histories, given the other variables in the model are held constant. Similar to accuracy of reporting of age, we can say that the older the respondent's age, we would expect the accuracy of CEB reports to get worse. Respondents are more likely to understate than to overstate the number of children they have given birth to.

Compared to permanent residents, migrants are 1.595 times more likely to overstate the number of children to which they have given birth. There are no statistical differences between respondents' ability to give accurate reports and understate their birth histories.

Compared to females, males (reporting on other household members) are 1.593 times more likely to overstate the number of births for the women in their households, controlling for other variables in the model. Male respondents are also more likely to overstate the number of births compared to female respondents.

4.5.2.3 MNLR model investigating the effects of respondent characteristics on accuracy of reporting of children ever born now dead (CD)

Age and residence status of the respondents are the only two predictors that have a bearing on the accuracy of reports on dead children. The predictor variables are only significant for respondents' reports of their own dead children.

The relative risk ratio (RRR) for a one-year increase in age predicting a respondent to fall in the "understated" relative to the "accurate" category would be expected to increase by a factor of 1.076 reporting on their number of dead children, given the other variables in the model are held constant. On the other hand, the RRRs to predicting a respondent to fall into "overstated" relative to the "accurate" category would be expected to increase by a factor of 1.073 when reporting the number of their dead children, controlling for the other variables in the model. Similar to accuracy of reporting of age, we can say that the older the respondent's age, the worse we would expect the accuracy of CEB reports to be. Respondents are more likely to understate than to overstate the number of their dead children.

Compared to permanent respondents, migrant respondents are 1.595 times more likely to overstate the number of their dead children. There are no significant statistical differences between respondents' tendency to accurately report their dead children and to understate their dead children.

4.6 Direct mortality estimates

The crude death rate computed from the reported deaths and the *de facto* population as estimated by the UCT survey is 14 per 1000 for both migrant and permanent residents. Excluding migrant residents and deaths gives an estimate of 12 per 1000, which is reasonably close to the figure of 11 per 1000 published by the INDEPTH Network (2010) for the Agincourt HDSS. The INDEPTH Network (2010) does not specify whether it includes migrants and migrant deaths in its computation.

Agincourt was part of Limpopo before the boundary changes in 2006, which resulted in it becoming part of Mpumalanga (Wits Rural Public Health and Health Transitions Research Unit 2011). Here we base the comparison of the UCT survey mortality estimates with those of Limpopo for two reasons. First, between the two adjacent provinces, Limpopo province has a mortality level that is lower than that of Mpumalanga and the Limpopo mortality level is more consistent with that of Agincourt. Second, since some of the indirect mortality estimates we seek to assess accuracy are for periods before 2007 when AHDSS was still part of Limpopo, this choice seems reasonable.

The ASSA2008 model is chosen for comparison because it has been calibrated both to the antenatal survey data and to produce mortality rates that are consistent with empirical evidence on mortality. Although comparisons are made using the Limpopo provincial ASSA output, it should be born in mind that differences between the ASSA and UCT survey might be due to the ASSA model being inappropriate and incorrect rather than the empirical estimates of underlying mortality being wrong. However, the model provides a reasonable basis for comparison, particularly for periods where direct estimates are not available.

Figures 4.19 and 4.20 compare the age-specific mortality rates from the UCT survey to the Limpopo provincial estimates obtained from the ASSA2008 model for females and males respectively. The figures show that the shapes of the mortality schedule are similar, with the levels of the UCT estimates higher, particularly for males (where the observed rates are nearly ten times higher), than the provincial average from ASSA2008 model. The direct mortality estimates from the UCT survey are higher than the ASSA2008 estimates, particularly for males, partly due to different households reporting on the same death (3.3% of matched deaths) in the survey, which could be due to the deceased being considered a member of more than one household or to inaccurate matching of the deaths.

For both males and females, adult mortality rates appear not to increase exponentially with age at the older ages. The rates are not smooth at older ages, particularly for women, due to errors in reporting/recording of age at death and the small sample size.

Figure 4.19 Female age specific mortality rates, Agincourt and Limpopo province

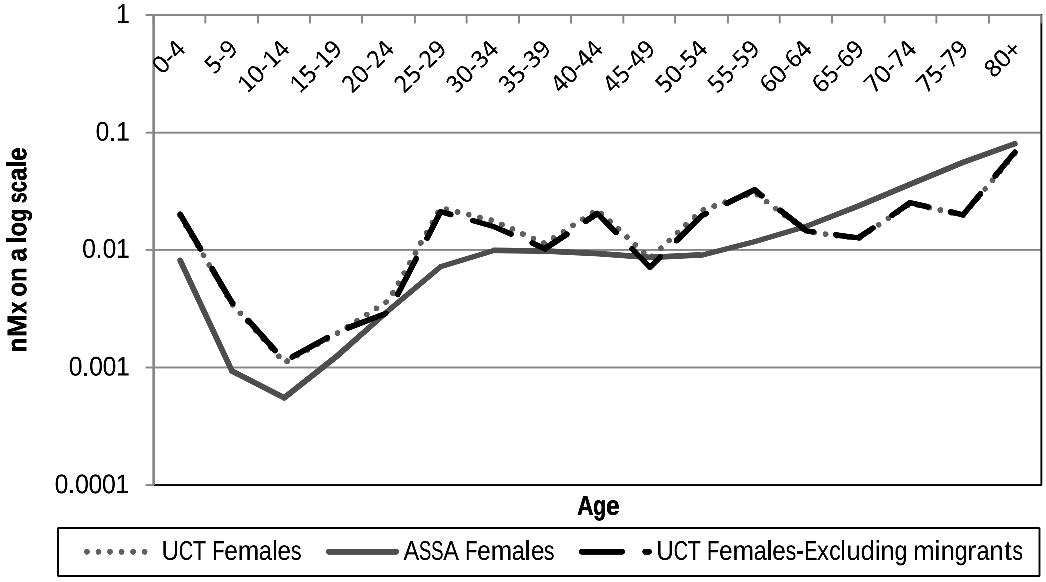


Figure 4.20 Male age specific mortality rates, Agincourt and Limpopo

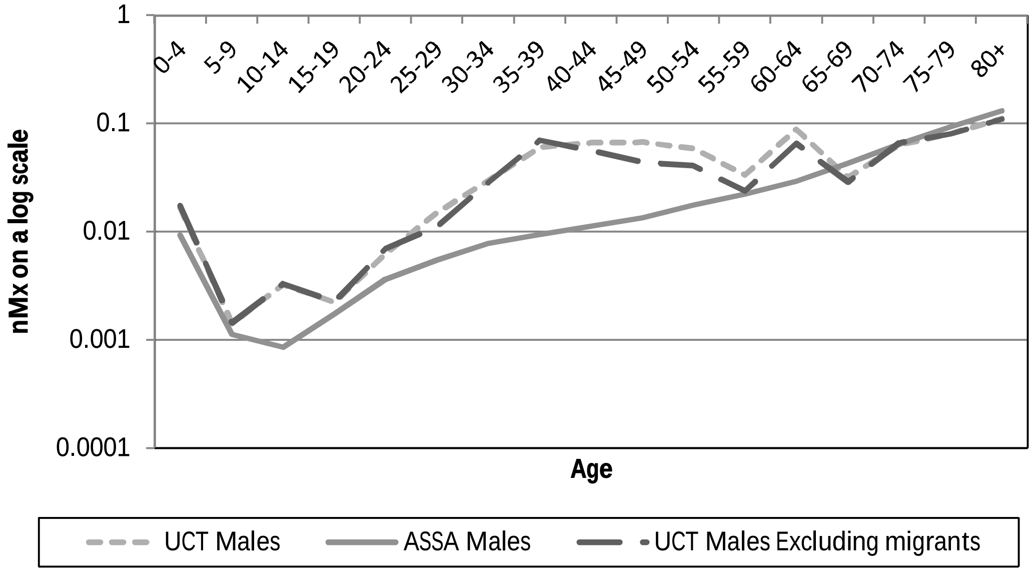


Figure 4.20 confirms the findings by Clark, Collinson, Kahn *et al.* (2007) that migrant male residents experience higher mortality than permanent male residents.

4.7 Indirect mortality estimates

4.7.1 Child mortality estimates from CEB/CS method

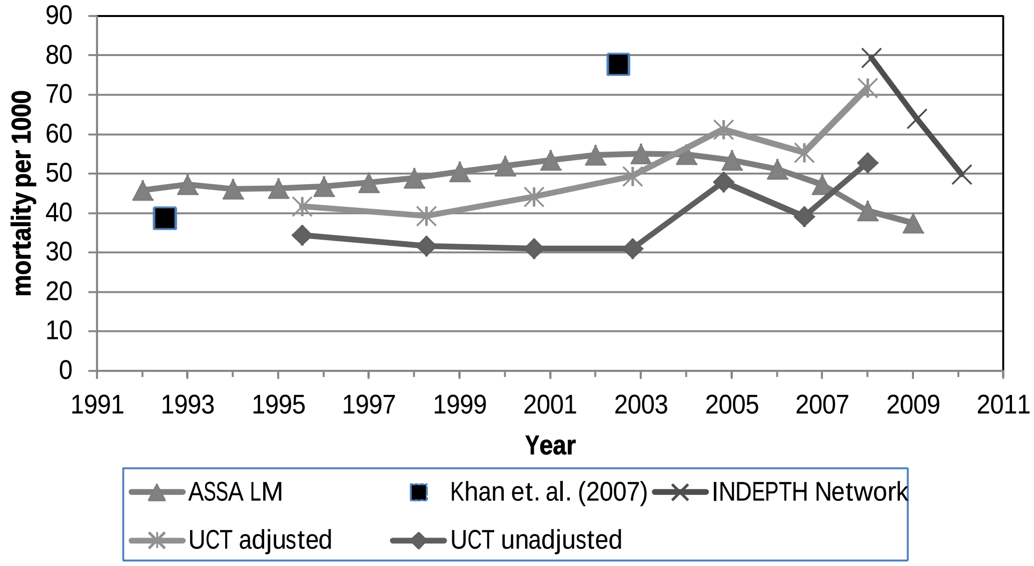
Figure 4.21 shows the trends in the probability of a new born surviving to exact age 5 for females and males combined, ${}_5q_0$, derived using the children ever born children surviving method compared to the Limpopo provincial trends obtained from the ASSA2008 model. The figure also compares the indirect estimates from the UCT survey to AHDSS direct estimates produced by Kahn, Garenne, Collinson *et al.* (2007) for the early 1990s but not for the mid-2000s.

Compared to ASSA, both the level and trends in ${}_5q_0$ from CEB/CS using the UCT survey data are different. Generally the estimates derived from CEB/CS are lower than those from ASSA2008 in certain instances, being almost half those from ASSA2008 even after adjusting for overall under-reporting of children ever born and dead children ever born. Adjustment for the effects of HIV/AIDS shifts the mortality trends to almost in line with those from both ASSA2008 model and by Kahn, Garenne, Collinson *et al.* (2007).

The most recent estimate of ${}_5q_0$ derived from CEB/CD reports on women in the 15-19 age group is above that estimated from ASSA2008. Mortality estimates from this age group have been shown to be typically overstated (Hill and Figueroa 1999). This is due to the effect of higher mortality among children born to younger mothers, who are also often from lower socioeconomic class and tend to start childbearing at younger ages (Hill and Figueroa 1999). In addition, the sample size of births and deaths from this age group is relatively small and hence random errors are usually greater (Hill and Figueroa 1999, Statistics South Africa 2012b). It is for these reasons that estimates from this age group are normally disregarded.

Compared to the direct estimates by Kahn, Garenne, Collinson *et al.* (2007), the estimates derived for the early 1990s are plausible. However, taking into account their direct estimates a decade later suggest that the CEB/CS method produces mortality estimates that are lower than direct estimates. The discrepancy is likely due to violation of some of the assumptions underlying the method (see section 4.9) and/or deficiencies in the data as discussed in section 4.2.

Figure 4.21 Comparison of ${}_5q_0$ from ASSA2008 and UCT survey, males and females combined



4.7.2 Female adult mortality using the maternal orphanhood method

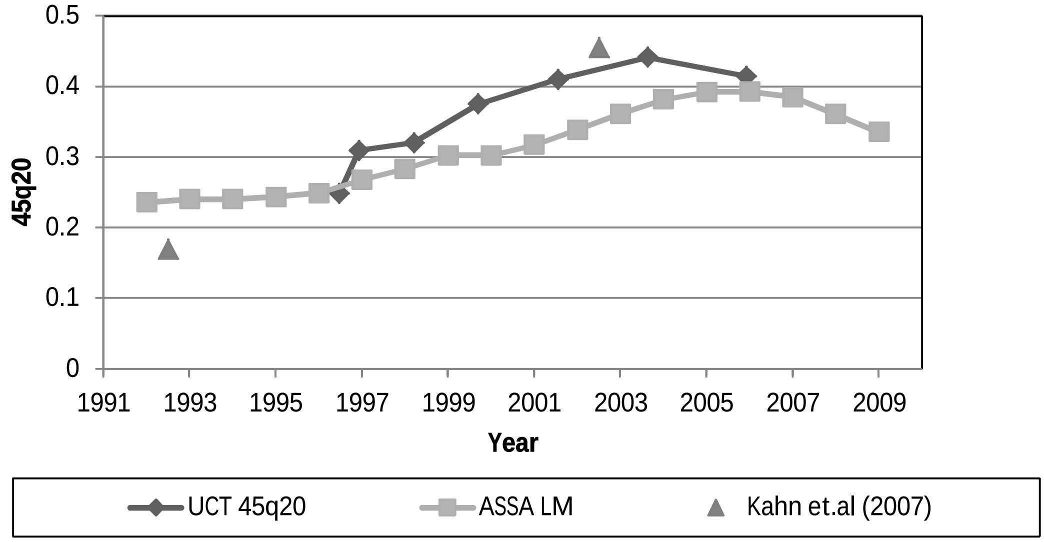
Figure 4.22 compares indirect estimates of the probability of dying for females between exact ages 20 and 65 years $({}_{45}q_{20})$ from the UCT survey with direct estimates of ${}_{45}q_{20}$ from the study by Kahn, Garenne, Collinson *et al.* (2007) and model estimates of ${}_{45}q_{20}$ for Limpopo province derived from ASSA2008. Timaeus' (1992) revised regression coefficients with the Brass general standard were used to compute the conditional life table measures, followed by conversion of these conditional probabilities into a common measure, ${}_{45}q_{20}$ before determining the time periods to which they pertain.

The trend in female mortality estimates from the orphanhood method, direct estimates by Kahn, Garenne, Collinson *et al.* (2007) and those from ASSA2008 are consistent. For the period for which direct estimates and orphanhood mortality estimates are simultaneously available, the estimates are close. However, the orphanhood estimates appear higher compared to estimates from the ASSA2008 model and the orphanhood estimates fall off before those from the ASSA2008 model.

The average age of the mothers at the time they gave birth to their children of 26.7 is different to that computed using birth histories from AHDSS of 24.3 years. The use of the overstated average age of the mothers at the time of their births in the maternal orphanhood method understates the level of mortality and distorts time location estimates slightly but does not affect the mortality trend (see section 4.10).

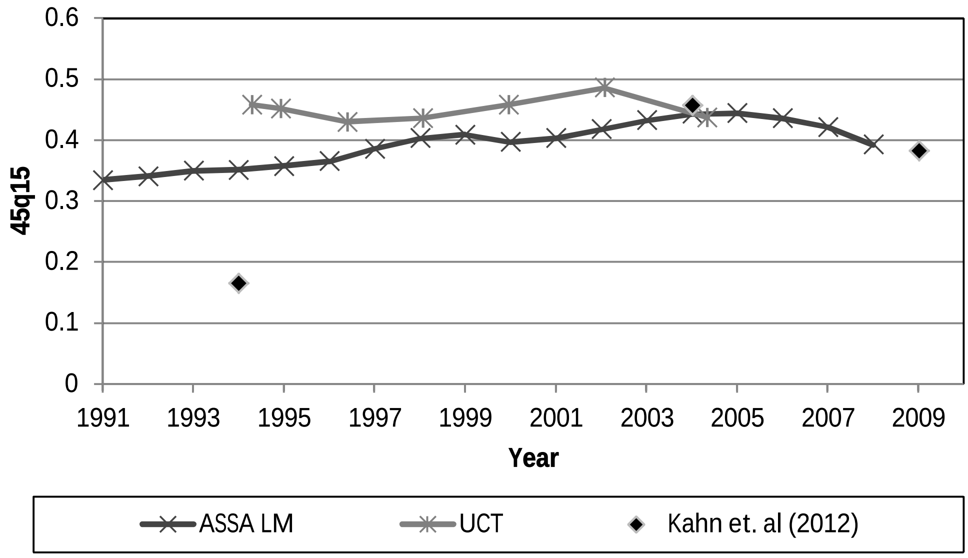
Figure 4.23 compares indirect estimates of the probability of dying for males between exact ages 15 and 60 years $({}_{45}q_{15})$ from the UCT survey with direct estimates of ${}_{45}q_{15}$ from the study by Kahn, Collinson, Gómez-Olivé *et al.* (2012) and model estimates of ${}_{45}q_{15}$ for Limpopo province derived from ASSA2008.

Figure 4.22 The probability of a female dying between exact age of 20 and 65 $({}_{45}q_{20})$ by year, Agincourt and Limpopo



The most recent estimates, although slightly higher, are reasonably close to those from both from ASSA2008 and those by Kahn, Collinson, Gómez-Olivé *et al.* (2012). However, for the early 1990s the UCT survey estimates are higher than either the ASSA2008 and the direct estimates by Kahn, Collinson, Gómez-Olivé *et al.* (2012). These differences could be suspected to be due to errors in the paternal orphanhood data as discussed in section 4.3.

Figure 4.23 The probability of a male dying between exact age 15 and 60 $({}_{45}q_{15})$ by year, Agincourt and Limpopo

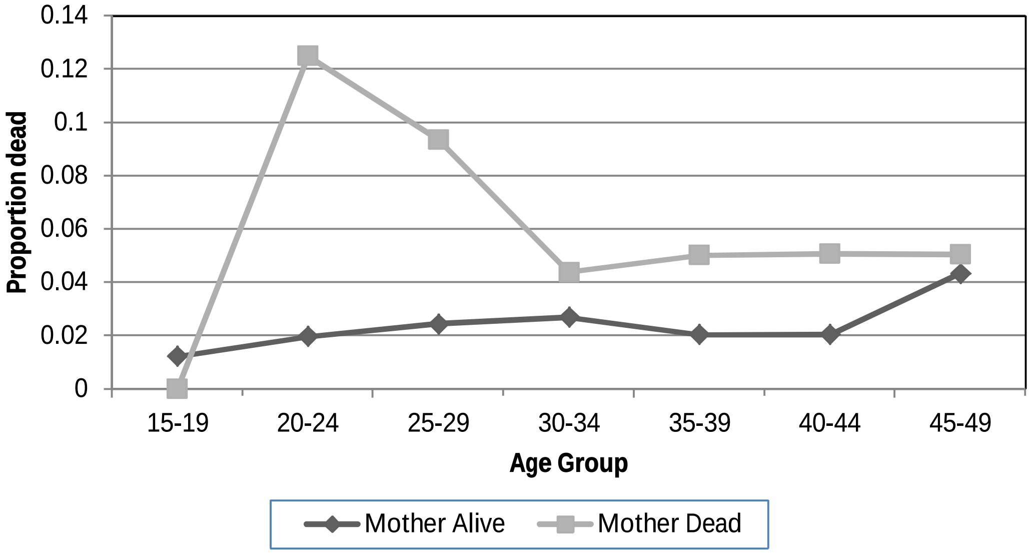


4.8 Investigating the validity of some of the assumptions underlying the CEB/CS method

4.8.1 Independence of mortality of the child and survival status of the mother

Figure 4.24 shows the proportion of children dead by age and status of the mother using AHDSS data on matched women. The proportion of dead children of dead mothers is higher than the proportion of dead children of live mothers for all ages, particular so for high-fertility ages 20-29. A two-sample t-test as shown in Table 4.15, confirms that the survival of the mother has a bearing on the survival of the child(ren), with dead mothers having a higher proportion of their children ever born dead than living mothers. Overall, all ages combined, the p-value of 0.000 shows that there is statistical significance between the survival status of children and that of their mothers.

Figure 4.24 Proportion of children dead by survival status and age of the mother



Given the relatively small numbers of children with dead mothers, the results from this analysis should not be over interpreted. Despite the small numbers involved, these findings are similar to other findings (e.g. Zaba, Whitworth, Marston *et al.* (2005), Chitiyo (2011)) and suggest that the mortality estimates derived might be biased due to the violation of the assumption. Chitiyo (2011), using HDSS data from Manicaland in Zimbabwe, found that the selection bias due to non-survival of mothers results in biases of between 3-10% in the mortality estimates. Thus if survival of children depends on the survival of the mother then mortality estimates derived from CEB/CS using reports from surviving mothers will be biased. However, despite the statistical tests being significant, the practical implications of the small numbers of children with dead mothers involved might result in negligible (if any) effect on the child mortality estimates derived without regard to the survival status of the mother.

Table 4.17 Two sample t-test for the proportion of children dead by survival status and age of the mother

Age group	***Status of Mother***		***P value (two sided)***
	Alive	Dead	
15-19	0.012	0.000	0.770
20-24	0.020	0.125	0.000**
25-29	0.024	0.094	0.000**
30-34	0.027	0.044	0.324
35-39	0.020	0.050	0.017*
40-44	0.020	0.051	0.325
45-49	0.043	0.050	0.696
15-49	0.025	0.064	0.000**

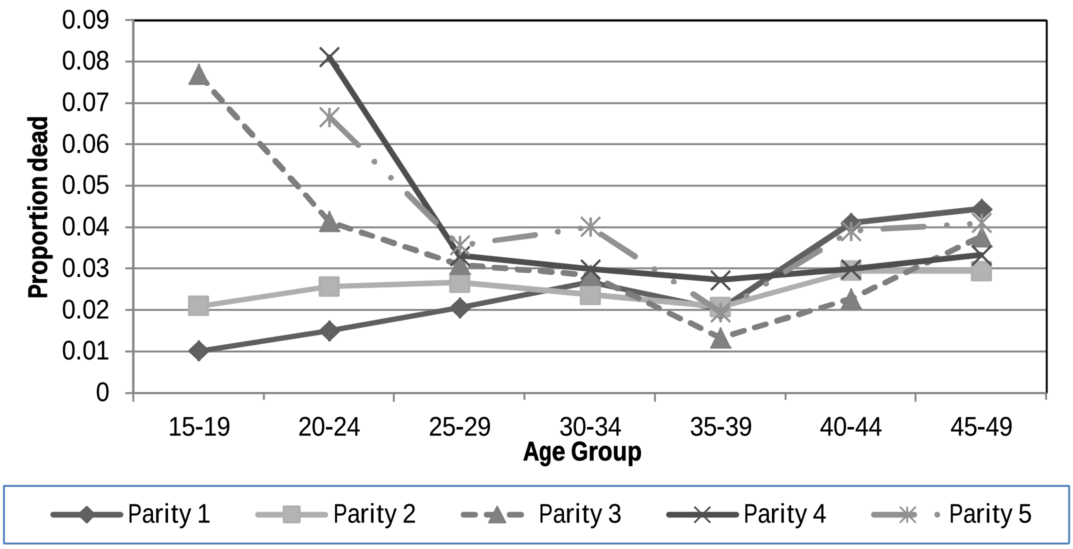
Notes

**means significant at both 1% and 5%, * means significant at 5% only

4.8.2 Independence of child mortality and parity and age of the mother

Figure 4.25 shows the proportion of children ever born who are now dead by parity and age of the mother. If the assumption of independence of mortality of the child and age and parity of the mother is not violated, one could expect the curves to be horizontal lines that coincide for all the parities. The curves shown in Figure 4.25 appear to be broadly horizontal and almost coinciding. The higher proportion of children ever born and now dead among younger women with high parity (parity 3, 4 and 5) does not lie on the horizontal line; showing that the mortality of children is not independent of the parity and age of their mothers. This can be explained by the fact that younger women who have a high parity are likely to have short spaced births and less likely to give better maternal and health care to their children and hence their children experience elevated mortality risks (Girma and Berhane 2011, Bhuyan 2000, Aguirre 1995). However, it is also possible that mortality of a child amongst these younger women could lead to short spaced births because of the biological effects of infant death and the associated child-replacement motivational response to child death resulting in higher parity (Chowdhury, Khan and Chen 1976). However, the results from the young women with high parity should not be over interpreted as the numbers involved are relatively small.

Figure 4.25 Proportion children ever born now dead by age and parity of the mother



4.8.3 Time reference for CEB/CS estimates

Table 4.18 shows the time references derived using the CEB/CS method compared to those derived using AHDSS data on matched women. The empirical time references are calculated as the average time at which the children died for each age group of the mothers. The difference between the actual reference period and that derived using the

CEB/CS are almost insignificant, averaging 1 year and ranging from -0.5 to 1.9 years. It would suggest that the time references are relatively robust to biases that might be in the data set and to violation of the two assumptions discussed above.

Table 4.18 Comparison of actual average to CEB/CS time references

Age group	*x*	*CEB/CS Time ref*	*Actual average*	*Diff. in years*
15-19	1	2008.0	2009.5	1.5
20-24	2	2006.6	2007.7	1.1
25-29	3	2004.8	2006.3	1.5
30-34	5	2002.8	2004.7	1.9
35-39	10	2000.6	2001.3	0.7
40-44	15	1998.3	1997.8	-0.5
45-49	20	1995.5	1996.0	0.5
			Average	1.1

4.9 The mean age at birth

The true mean age at birth computed using birth histories from AHDSS is 25.3 years for the matched (to UCT survey) women and 26.7 years for all women in the 21 villages. The figure of 26.7 years corresponds exactly, to that computed using the mothers' age at birth from births in the year before UCT survey. Figure 4.26 shows the effect of using the lower estimate on the derived mortality estimates.

Figure 4.26 Sensitivity of mortality estimates and time periods to mean age at birth

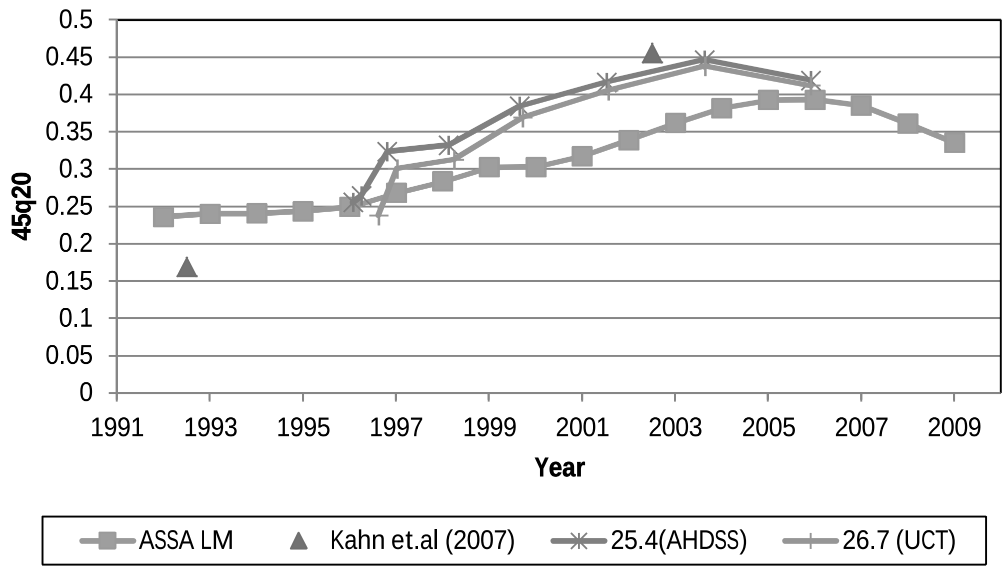


Figure 4.26 reveals that the mortality trend is not much affected by using an understated/overstated mean age at birth. However, as might be expected, the mortality levels are affected, with minimal effect on mortality estimates based on the survival of

children of women of younger ages and increasing for mortality estimates derived from women of older ages.

5 DISCUSSION AND CONCLUSIONS

This research sought to assess the accuracy of data obtained through retrospective census/survey questions typically used in population settings without complete vital registration systems, in order to derive child and adult mortality measures by indirect demographic techniques. In addition, the research sought to evaluate the extent of bias in the mortality estimates produced by these data and the validity of some of the assumptions underlying the indirect methods. The research also investigated respondent characteristics that might have an effect on the quality of the responses to the survey questions. This chapter examines and reflects on the extent to which the set objectives were met and the limitations of the research, and identifies areas for possible future research.

5.1 Count of the population

The UCT survey was meant to be a *de facto* census of the original 21 villages of Agincourt and should have captured all 54 194 permanent residents. However, there was an undercount of 19%. Whilst the figure of 19% is higher than the South African census national undercounts of 10.7% in 1996 ,17.6% in 2001 and 14.6% in 2011 (Masiteng and Kekovole 2010, Statistics South Africa 2012a), the difference could partly be attributable to the difference in the manner in which the undercount was determined in addition to actual omission of individuals from the survey. For the two national censuses, post enumeration surveys (PES) were used to determine the extent of the undercount. The UCT survey undercount was determined by comparing the enumerated population to the population of permanent residents that should have been enumerated (as captured by the AHDSS 2009). This approach presents challenges in that the UCT survey only captured those temporary migrants who had not slept at home the night before the survey but were returning the following day and excluded those temporary migrants not returning the following day. Since there was no temporary migration module conducted by the AHDSS in 2009, it was impossible to ascertain the number of such temporary migrants and to exclude them from the AHDSS population used to determine the undercount. These temporary migrations have been shown to be prevalent, roughly around 60% of males and 20% of females aged 20-60 years (Collinson 2010, Clark, Collinson, Kahn *et al.* 2007) and thus they are likely to have an effect on the magnitude of the undercount, as determined above. In addition, the census population adjusted for undercount using the PES is considered by demographers still to underestimate the population. The 1996 South Africa PES adjusted estimate of the population is estimated to have understated the population by between 2% to 5% (Dorrington 1999).

The household undercount, at 26%, was higher than the individual undercount. The undercount was split equally between households that were never visited and households that were visited but either no one eligible to answer the questionnaire was present with no subsequent follow-up visits (the majority of cases) or refusals to participate in the survey (very few cases). These missed households were on average small households (about 3 persons per household) with more than half of the residents being under the eligible age to respond (18 years) thus making it more likely that these households were not enumerated, either due to temporal migration of adults or to the absence of an adult at the time of the visit. This is similar to the finding by Saloojee, Kahn, Clark *et al.* (2005) who, when analysing Verbal Autopsy (VA) surveys from 1999 to 2004, found that failure to identify a suitable respondent, usually because of migration, accounted for the majority of unsuccessful or incomplete responses. The household undercount was higher than the national South African census undercounts of 6.6% , 20.4% and 14.3% in 1996 , 2001 and 2011 respectively (Masiteng and Kekovole 2010, Statistics South Africa 2012a). Although the UCT survey tried to emulate a typical census, it did not conduct repeat visits outside normal hours to enumerate households missed during the day, as is the case with national censuses, which probably contributed to a large extent to the higher undercount, compared to the South African national censuses. Given that the proportion of households missed by the UCT survey exceeded that of the 2011 census by more than the proportion of people missed, it would appear that the UCT survey was relatively better at enumerating people in the households than the national census.

5.2 Age and sex distribution of the population

A relatively high proportion of individuals enumerated (8%) did not have dates of birth and age recorded. Distributing those with missing dates of birth and ages, assuming they were adults (aged 20+) with the same age distribution as those with recorded age or dates of birth, appeared to be a reasonable assumption. For those with both age and date of birth recorded, the inconsistency between the reported age and the dates of birth for about 30% of the individuals suggests a tendency to record age as the age at next birthday instead of age at last birthday. This is probably because the enumerators attempted to compute the ages from the given dates of birth, as the survey asked, contrary to the usual order of these questions in censuses, the date of birth before asking the age at last birthday. The reporting of dates of birth was more accurate relative to reporting of age at last birthday though there were more individuals without reported date of birth than without reported age at last birthday.

Reporting of sex was relatively accurate given that the sex variable only takes either of two values and is something the respondent is likely to know. About 0.3% of individuals had no sex recorded, and for 1.5% it was recorded incorrectly, probably because of careless completion of the questionnaire.

Other than the undercounts pointed out above, comparing the UCT survey and the permanent AHDSS population revealed a similar age-sex profile. Thus, although the UCT survey failed to enumerate the whole population, the enumerated population was representative of Agincourt with respect to age and sex in the sense that a simple scaling up for undercount would have produced a reasonable approximation of the total population by age and sex.

5.3 Aspects of the data on mortality questions

The extent to which the matching of the UCT survey to the AHDSS census was achieved was dependent on the accuracy and completeness of the information on the enumerated individuals collected in the former. The exercise managed to match a high proportion (95%) of the enumerated individuals, and thus reasonably little data were lost for assessment of their accuracy. However, where possible, consistency and plausibility checks of the data for the unmatched individuals were carried out. Also it was reassuring to note that the distribution by age and sex of the unmatched individuals was similar to that of the matched individuals.

O'Reilly, Rosato and Connolly (2008), when attempting to link census data and longitudinal death record data, found that the non-linkage of census and death records was due to a combination of non-enumeration at census and deficient information about the deceased recorded at the time of death and/or at census. They also noted that the unmatched individuals might have been demographically different from those that matched. In this study, non-enumeration of some households in the UCT survey resulted in those households subsequently not being matched with AHDSS, and thus no analysis of the quality of the responses that could have been obtained from those households was done. Whilst there appear to be no significant demographic differences in the aggregate enumerated and non-enumerated population, there are notable differences in the size of the households. The enumerated population had a larger average household size than the non-enumerated population.

The reporting of children ever born was more accurate for women above the age of 20, with an overall omission of about 5% of boys and 10% of girls. However, the same cannot be said about the reporting of children ever born who are now dead. Whilst the general observation from other literature has been that reporting of child survivorship data gets worse with age of the mother, this research showed that this is not the case. As shown in section 4.2, the reported proportion fluctuates with age between 45% and 85% of the actual number, except at the youngest ages where there were more recorded dead children than the actual number of dead children. However, the sample size of women who have experienced death of their child(ren) was small, just under 5% of all matched women of reproductive ages. The data on children ever born (CEB) and children ever born now dead (CD) from the youngest age group (15-19) appear more problematic and appear to have more errors than any other age group. This is in part due to the small number of women who have given birth in this age group. Overall, due to the retrospective nature of the birth history data, for all women, it is possible that some children ever born, especially those that do not stay with their mothers and those children who have died, were omitted (Hill 1991). This research confirmed that relatively more dead children than children ever born were omitted, with the consequence that the proportion of children ever born who are now dead are distorted, which results in underestimation of mortality.

Overall, the reporting/recording of maternal orphanhood status at the time of the survey was accurate for 95.3% of the enumerated population. The inaccurate 4.5% was split into two groups: 3.6% with their mothers recorded alive when they were in fact dead, and the remaining 0.8% who specified that their mothers were dead when they were in fact alive. This resulted in a cancelling effect at roughly all age groups for 0.8%, leaving an aggregate of 2.8% inaccurate. The adoption effects might explain those individuals for whom their mother is reported as alive in the UCT survey when she was in fact dead, as is often, or assumed to be, the case for African settings (Timæus 1991a, Timæus 1991b, Brass 1996, Palloni, Massagli and Marcotte 1984).

Whilst the survey seems to have recorded all household deaths in the year before the survey, more than half of the deaths had no age at death. About 3.3% of the matched deaths were reported in two different households contributing to higher direct mortality being derived from the survey than from ASSA2008. There was overall consistency in the proportion of deaths that are reported registered between the UCT survey and the AHDSS, suggesting death registration can be estimated by including the question on death registration in surveys.

Supplementary questions on maternal orphanhood status at the time of the 1994 elections had a low response rate. It appears the question might not have been understood by the respondents. To be able to estimate mortality using these data requires that the majority of events are reported (Timæus 1991a), thus, given the low response rate to the question on maternal orphanhood status at the time of the 1994 elections, the data were rendered unusable for mortality estimation.

5.4 The characteristics of the best respondents

Whilst there are many respondent characteristics that could affect the quality of the responses, this research only looked at a few because these were the only characteristics captured by the UCT survey. An alternative would have been to use respondent characteristics as captured by the AHDSS, although this has the disadvantage of excluding from the analysis respondents who did not match. This study examined the effects of the few respondent characteristics on the accuracy of the responses to questions. In particular, it considered the accuracy of answers to questions on age, children ever born and children ever born who are now dead.

Overall, age was the most important characteristic for all the three questions, with the younger respondents (15-29) being the better respondents and the responses getting worse with increasing age. This could be explained by the fact that respondents in this age group are likely to have a closer relationship with household members than older respondents (they are either reporting on their children, spouse, siblings or parents) and thus more likely to know about them.

Age was followed by residence status, with permanent residents being better respondents than migrant respondents, as might be expected. This is plausible because permanent respondents spend more time with the other household members than migrant respondents and would thus have more accurate knowledge about the other household members.

On the accuracy to responses to children ever born, women seemed to give more accurate responses than men. This could be partly because the women were reporting on their own childbirth history as well as due to the fact that women respondents are more likely than male respondents to be permanent residents due to the sex differentials in migration patterns.

The household size of the respondents affected the accuracy of the age reported. Somewhat surprisingly the accuracy of age reporting increases with increasing household size. Respondents from large households were more likely to give their correct ages and correct ages for other household members than small households. This might be due to large households having close-knit families than smaller households. It could also potentially be that small households are made up of people who are cohabiting or not necessarily related and thus likely to know little about each other. Predictor variables appear only to affect the accuracy of respondents' own responses/measurements and not the accuracy of their responses on other household members. The only notable exception occurs for reporting of age.

5.5 Biases in the indirect mortality estimates

Adjustments to the child mortality estimates using the modified Ward and Zaba (2008) method brings the level closer to those from the ASSA2008 model. However, given the understatement in the recorded children ever born now dead, the trend still remains somewhat out of line with the ASSA2008 model for the province of Limpopo. Also, it should be noted that the estimates of HIV prevalence of women aged 15-49 figures used for the Ward and Zaba adjustment were taken from ASSA2008. The ASSA2008 prevalence figures for 2011 appear to be lower than those from an empirical study by Gómez-Olivé, Clark, Houle *et al.* (2012), suggesting that they might well have been too low for the earlier years. Thus adjustments using the Ward and Zaba method based on these prevalence estimates ought to be higher than the ones used in this study although the effect would have been small. However, time references seem unaffected by the errors in the data and the violation of some of the assumptions. Time reference estimates for the child mortality estimates derived from the CEB/CS technique are reasonably accurate with an absolute difference that ranges between 0.5 and 1.9 years, and averaging 1.1 years.

The trends in the indirect adult mortality estimates derived from maternal orphanhood data are reasonably similar to the direct estimates, with the one point of the direct estimates, which occurs in the same time period as indirect estimates, suggesting that the level is plausible although a bit low. Comparison to the ASSA2008 model for Limpopo suggests the trend is plausible with the level being higher than ASSA2008. This seems to suggest that ASSA2008 might not be an appropriate benchmark for Agincourt. However, in the absence of alternative independent estimates on Agincourt, the model provides a reasonable basis for the comparison.

Crude mortality rates computed using household death data are reasonably close to those published, and confirms that mortality of migrants is higher than that of permanent residents (Clark, Collinson, Kahn *et al.* 2007). The age-specific mortality rates produced from the survey data are not smooth, showing some kinks at some ages. This suggests that the age distribution of individuals with recorded age at death might have been significantly different from those without recorded ages. This could not be verified using matched deaths as the majority of unmatched deaths did not have age data.

5.6 Limitations of the study

The conclusions of the study have to be viewed in terms of a caveat. Whilst the HDSS data are incredibly rich and can be considered relatively accurate, care needs to be exercised in generalising the results obtained from using these data to the national population due to the unrepresentative geographic location and size of the study site, which is relatively homogeneous with limited community variation. In addition, the presence of the HDSS system for a long time in the community could "contaminate" the site due the effect of the continued research on the behaviour of study participants where they become "educated" participants and tend to give a "trained" response, the so-called Hawthorne effect (Kahn 2006).

Another limitation is that the size of the study population was not big enough to allow all the investigations to be done. For example, investigation of the accuracy of reporting of maternal deaths could not be undertaken due to there being too few maternal deaths and the data were too few to allow the estimation of child mortality by sex.

The study could not look at all the aspects that affect the quality of data and hence the derived mortality estimates. According to Vemuri's (1994) conceptual framework on the sources of error in demographic data, these aspects include, but are not limited to, structural factors, interviewer characteristics, interviewee characteristics and contextual variables. This study only looked at the interviewee characteristics and some of the structural factors. Interviewee characteristics have been seen to have an effect on the quality of the responses, in some instances interviewer perceptions which influence the recorded responses (Vemuri 1994).

There were no benchmark mortality estimates that could be used to assess the estimates derived more accurately. The proposed INDEPTHStats could be of great help in this regard. The INDEPTHStats is an online tool that aims to compile indicators of population, fertility, migration, cause of death and mortality at each of the INDEPTH

Network member sites (INDEPTH Network 2012). The study assumed that the initial census of the AHDSS was accurate and if there were errors, these were rectified in the subsequent census updates.

While efforts were made to simulate the conditions of the census, the training and implementation of the UCT survey were not exactly the same as those of the South African censuses. This could introduce differences between the collected UCT survey and what might have been collected on Agincourt in a national census.

5.7 Ideas for future research

Whilst this research did make some adjustments for the effects of violation of the underlying assumptions in the indirect techniques, explicit evaluations of each of the assumptions in the CEB/CS and orphanhood method should be done. This will serve to enlighten researchers if the methods still work and possibly provide a basis for improving the current methods. This is particularly important in light of the prevalence of HIV/AIDS in developing countries which are usually under-resourced and rely on indirect methods for deriving mortality estimates. A study similar to this one should be undertaken using the recent South African census and combined data sets from the two health and demographic surveillance sites in South Africa. The combined data sets will allow the evaluation of the quality of the South African 2011 census data. Alternatively, another separate census enumeration at the Africa Centre can be done with the view of carrying out similar research to this study. Working with combined HDSS will improve the sample sizes and reduce variability and thus improve the confidence in the results that will be obtained. The census data will also permit the assessment of cause-specific mortality rates by combining the Verbal Autopsy data from the HDSSs.

Investigation of interviewer characteristics should be investigated using the information about the interviewers collected at the time of their recruitment and the length and duration of the training periods.

Finally, since no explicit evaluation of the AHDSS data was carried out in this study, an evaluation of these data could be done to establish the quality of the AHDSS.

5.8 Conclusions

This research looked at the quality of the data collected using the special demographic questions in surveys, some of the assumptions underlying the indirect methods and the biases in the child and adult estimates derived. Comparing the quality of the survey data used to compute indirect child mortality estimates to that used to compute indirect adult mortality estimates, shows that the former has more errors than the latter, with the consequence that child mortality estimates are bound to be more unreliable than adult mortality estimates. However, this research has similar findings to those by Kahn (2006) in which she identifies that the three sources of data: census data, AHDSS data, and the ASSA2003 demographic model, show some differences in the apparent levels of mortality, but are generally consistent in their accounts of the trends and patterns of mortality. Also, this study confirms the findings from her research that indirect estimates of adult mortality are more robust than indirect estimates of child mortality.

Indirect child mortality estimates are less robust than those for adults because of the relatively poor quality of responses to the CEB/CS questions compared to responses to orphanhood questions. This is coupled with a greater scope for violation of some of the assumptions of CEB/CS than orphanhood. Responses to questions on child mortality are prone to errors, with reporting of children ever born now dead being more problematic, typically understating the number of children ever born now dead resulting in understated mortality. Child mortality estimates for the recent periods cannot be reasonably estimated given the high errors in the data from women in the youngest age group and small sample of women who have given birth in this age group. In addition, the child mortality estimates are affected by HIV/AIDS, requiring some adjustments, and these require further research. Work by Ward and Zaba (2008) and Darikwa and Dorrington (2011) seem promising in this regard.

Recent child mortality may be estimated by asking questions about the births of each woman that occurred in the 24 month period before the survey and asking whether the child is still alive or dead. The data may then be used with the Blacker and Brass (2005) method to produce more robust estimates of infant mortality rate (IMR).

Questions on household deaths in the year before the survey and questions asking whether the deaths were registered have the potential to allow estimation of recent mortality levels at all ages and the completeness of death registration. However, there is a need to train the enumerators to be able to elicit accurate information, in particular on ages at death.

Given the results from this research, one might conclude that, although the data collected in surveys and censuses may still be prone to some errors, the data and mortality estimates derived can be considered reasonable, provided the planning and conducting of the actual census/survey is done with due care.

REFERENCES

Adetunji, J. A. 1996. "Infant mortality levels in Africa: Does method of estimation matter?", *Genus* **52** (3/4):89-106.

Aguirre, G. P. 1995. *Child mortality and reproductive patterns in urban Bolivia.* Center for Demography and Ecology Working Paper No. 95-28. Madison: University of Wisconsin-Madison. www.ssc.wisc.edu/cde/cdewp/95-28.pdf.

Bah, S. 1999. "Diagnostic tests on assessing the quality of maternal orphanhood data from the 1996 South African census and implications for the indirect estimation of adult mortality", *Population Studies Centre Discussion Papers Series* **13** (5):1-41.

Bangha, M., Diagne, A., Bawah, A. *et al.* 2010. "Monitoring the millennium development goals: the potential role of the INDEPTH Network", *Glob Health Action* **3** (1):5517-25.

Bhuyan, K. C. 2000. "Differential in child mortality by fertility in North-Eastern Libya", *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* **62** (2):317-326.

Blacker, J. 2004. "The impact of AIDS on adult mortality: evidence from national and regional statistics", *AIDS* **18** (2):19-26.

Blacker, J. and Brass, W. 1993. "Experience of retrospective demographic enquiries to determine vital rates," in Bulmer, M. and Warwick, D. P. (eds). *Social research in developing countries: surveys and censuses in the Third World.* London: UCL Press, pp. 195-206.

Blacker, J. and Brass, W. 2005. "The estimation of infant mortality from proportions dying among births in the past 24 months", *Southern African Journal of Demography* **10** (1/2):25-42.

Blacker, J. G. 1977. "The estimation of adult mortality in Africa from data on orphanhood", *Population Studies* **31** (1):107-28.

Blacker, J. G. 1981. "Experience in the use of special mortality questions in multi-purpose surveys: the single-round approach," Paper presented at the Meeting of the United Nations/World Health Organization Working Group on Data Bases for Measurement of Levels, Trends and Differentials in Mortality. Bangkok, 20-23 October 1981.

Bradshaw, D., Laubscher, R., Dorrington, R. *et al.* 2004. "Unabated rise in number of adult deaths in South Africa", *South African Medical Journal* **94** (4):278–279.

Brass, W. 1996. "Demographic data analysis in less developed countries: 1946-1996", *Population Studies* **50** (3):451-467.

Brass, W. and Coale, A. J. 1968. "Methods of analysis and estimation," in William Brass, Coale, A. J., Demeny, P. *et al.* (eds). *The Demography of Tropical Africa.* New Jersey, Princeton: Princeton University Press, pp. 88-139.

Chen, X., Ender, P., Mitchell, M. *et al.* 2008. "*Logistic Regression with Stata*," Los Angeles: UCLA: Academic Technology Services, Statistical Consulting group http://www.ats.ucla.edu/stat/stata/webbooks/logistic/chapter3/statalog3.htm Accessed: 10 March 2012

Chitiyo, V. 2011. "The impact of HIV on the summary birth history method of estimating child mortality: A Zimbabwean demographic surveillance case study." MPhil in Demography, Cape Town: University of Cape Town.

Chowdhury, A. K. M. A., Khan, A. R. and Chen, L. C. 1976. "The effect of child mortality experience on subsequent fertility: in Pakistan and Bangladesh", *Population Studies* **30** (2):249-261.

Clark, S. J., Collinson, M. A., Kahn, K. *et al.* 2007. "Returning home to die: circular labour migration and mortality in South Africa", *Scandinavian Journal of Public Health* **35** (69 suppl):35-44.

Clark, S. J., Jasseh, M., Punpuing, S. *et al.* 2008. INDEPTH Model Life Tables 2.0. Available: http://samclark.net/CV/MANUSCRIPTS%20IN%20PREPARATION/clark%20et%20al_2009_indepth%20model%20life%20tables%202.0.pdf Accessed: 20 March 2012.

Cleland, J. 1996. "Demographic data collection in less developed countries 1946-1996.", *Population Studies* **50** (3):433-450.

Coale, A. J., Kirk, D., Hauser, P. M. *et al.* 1971. "The determination of vital rates in the absence of registration data.", *The Milbank Memorial Fund Quarterly* **49** (4):175-191.

Collinson, M. A. 2010. "Striving against adversity: the dynamics of migration, health and poverty in rural South Africa", *Glob Health Action* **3** (1):5080-5094.

Collumbien, M. and Sloggett, A. 2001. "Adjustment methods for bias in the indirect childhood mortality estimates," in Zaba, B. and Blacker, J. (eds). *Brass Tacks: Essays in Medical Demography.* London: Athlone, pp. 20-42.

Darikwa, T. B. and Dorrington, R. 2011. "The level and trends of child mortality in South Africa, 1996-2006", *African Population Studies* **25** (1 Supplement):159-172.

DeMaris, A., Teachman, J. and Morgan, S. P. 1990. "Interpreting logistic regression results: A critical commentary", *Journal of Marriage and Family* **52** (1):271-277.

Department of Health. 2011. "Epidemiological Comments. Progress on HIV Incidence Estimation Methods in South Africa.," Paper presented at The National Consultation on HIV Estimation in South Africa Durban, March 2011.

Dorrington, R. 1999. "To count or to model that is not the question: Some possible deficiencies with the 1996 census results," Paper presented at PRU Arminel Roundtable on: "The Reception of the Census of 1996". East London, April 9th -11th 1999.

Dorrington, R. and Bradshaw, D. 2011. "Maternal mortality in South Africa: Lessons from a case study in the use of deaths reported by households in censuses and surveys", *Journal of Population Research* **28** (1):49-73.

Dorrington, R., Moultrie, T. A. and Timaeus, I. A. 2004. *Estimation of mortality using the South Africa census 2001 data.* CARe Monograph No. 11. Cape Town: Centre for Actuarial Research. http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono11.pdf.

Dorrrington, R., Johnson, L. and Budlender, D. 2010. *ASSA2008 AIDS and Demographic models user guide.* Beta Version. Cape Town: AIDS Committee of the Actuarial Society of South Africa.

Ewbank, D. C. 1982. "The sources of error in Brass's method for estimating child survival: The case of Bangladesh", *Population Studies* **36** (3):459-474.

Feeney, G. 1976. "Estimating infant mortality trends from child survivorship data", *Asian and Pacific Census Newsletter* **3** (2):12-16.

Feeney, G. 1980. "Estimating infant mortality trends from child survivorship data", *Population Studies* **34** (1):109-128.

Feeney, G. 2001. "The impact of HIV/AIDS on adult mortality in Zimbabwe", *Population and Development Review* **24** (4):771-780.

Gakidou, E., Hogan, M. and Lopez, A. D. 2004. "Adult mortality: time for a reappraisal", *International Journal of Epidemiology* **33**:710–717.

Garenne, M. and Gakusi, E. 2006. "Health transitions in sub-Saharan Africa: Overview of mortality trends in children under 5 years old (1950-2000)", *Bulletin of the World Health Organisation* **84** (6):470-478.
Gil, B. and Omaboe, E. N. 1993. "Population censuses and national sample surveys in developing countries," in Bulmer, M. and Warwick, D. P. (eds). *Social research in developing countries : surveys and censuses in the Third World.* London: UCL Press, pp. 43-54.
Girma, B. and Berhane, Y. 2011. "Children who were vaccinated, breast fed and from low parity mothers live longer: a community based case-control study in Jimma, Ethiopia", *BioMed Central Public Health* **11** (197):1-7.
Gómez-Olivé, F. X., Clark, S. J., Houle, B. *et al.* 2012. *The Prevalence of HIV among adults 15 and older in rural South Africa: A cross-sectional study.* San Francisco: Population Association of America. http://paa2012.princeton.edu/papers/122637.
Hallett, T. B., Gregson, S., Kurwa, F. *et al.* 2010. "Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection", *Bulletin of the World Health Organization* **88** (10):761-8.
Hansen, M. H. and Waksberg, J. 1970. "Research on non-sampling errors in censuses and surveys", *Review of the International Statistical Institute* **38** (3):317-332.
Hartmann, M. 2009. *Demographic Methods for the Statistical Office.* Sweden: Statistics Sweden. http://www.scb.se/statistik/_publikationer/OV9999_2009A01_BR_X103BR0902.pdf.
Hill, K. 1991. "Approaches to the measurement of childhood mortality: A comparative review", *Population Index* **57** (3):368-382.
Hill, K. 2006. "Making deaths count", *Bulletin of the World Health Organization* **84** (3):162-162.
Hill, K. and Figueroa, M.-E. 1999. *Child mortality estimation by time since first birth.* Baltimore: Johns Hopkins University. http://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-population-center/publications/pdf/WP99-05.pdf.
Hill, K. and Stanton, C. 2011. "Measuring maternal mortality through the census: rapier or bludgeon?", *Journal of Population Research* **28** (1):31-47.
Hill, K. H. 2013. "Indirect estimation of child mortality," in Moultrie, T. A., Dorrington, R. E., Hill, A. G. *et al.* (eds). *Tools for Demographic Estimation.* Paris: International Union for the Scientific Study of Population, pp. 148-164.
INDEPTH Network 2011. *Agincourt HDSS profile.* Accra: INDEPTH Network. http://www.indepth-network.org/Profiles/Agincourt%20HDSS.pdf.
INDEPTH Network. 2012. INDEPTH makes headway in data preparation for all-cause and cause-specific mortality analysis. *Better Health information for Better health Policy* 3.12, http://www.indepth-network.org/index.php?option=com_content&task=view&id=1491 Accessed: 19 June 2012.
Kahn, K. 2006. "Dying to make a fresh start: Mortality and health transition in a new South Africa." PhD Thesis, Umeå: Umeå University.
Kahn, K., Collinson, M. A., Gómez-Olivé, F. X. *et al.* 2012. "Profile: Agincourt Health and Socio-demographic Surveillance System", *International Journal Of Epidemiology* **41** (4):988-1001.
Kahn, K., Garenne, M. L., Collinson, M. A. *et al.* 2007. "Mortality trends in a new South Africa: Hard to make a fresh start", *Scandinavian Journal of Public Health* **35** (Suppl 69):26-34.

Kiregyera, B. 1982. "Some sources of nonsampling errors in African census and survey work: A review", *International Statistical Review* **50** (3):305-316.
Mahy, M. 2003. "Measuring child mortality in AIDS-affected countries," Paper presented at Training Workshop on HIV/AIDS and Adult Mortality in Developing Countries. New York, 8-13 September 2003.
Masiteng, K. and Kekovole, J. 2010. *Planning of Census 2011: Approach and implementation strategies.* New York: United Nations. https://unstats.un.org/unsd/demographic/sources/census/2010_phc/South_Africa/South_Africa.pdf.
Mathers, C. and Boerma, T. 2010. "Mortality measurement matters: Improving data collection and estimation methods for child and adult mortality", *PLoS Medicine* **7** (4):e1000265.
Mathers, C. D., Fat, D. M., Inoue, M. *et al.* 2005. "Counting the dead and what they died from: An assessment of the global status of cause of death data", *Bulletin of the World Health Organization* **83** (3):171-177.
Mba, C. 2006. *Civil registration system and census exercise in Nigeria: The challenges of demographic estimation.* Legon: University of Ghana. http://uaps2007.princeton.edu/download.aspx?submissionId=70109.
Mee, P., Mgiba, N. and Collinson, M. 2006. *Agincourt Demographic Surveillance System: Tables Guide (Data Dictionary v1.1).* Accra: INDEPTH Network. http://www.agincourt.co.za/datasection/datadocumentation/ahdss_overviewdatadictionary.pdf.
Menard, S. W. 2002. *Applied logistic regression analysis.* Thousands Oaks, Califonia: Sage Publications Inc.
Moultrie, T. A. 2013. "General assessment of age and sex data," in Moultrie, T. A., Dorrington, R. E., Hill, A. G. *et al.* (eds). *Tools for Demographic Estimation.* Paris: International Union for the Scientific Study of Population, pp. 3-21.
MRC/Wits Rural Public Health and Heath Transitions Research Unit 2012. "*Agincourt: Overview*," http://www.agincourt.co.za/index.php/about/ Accessed: 12 April 2012
Ng'weshemi, J., Urassa, M., Isingo, R. *et al.* 2003. "HIV impact on mother and child mortality in rural Tanzania", *Journal of Acquired Immune Deficiency Syndromes (JAIDS)* **33** (3):393-404.
O'Reilly, D., Rosato, M. and Connolly, S. 2008. "Unlinked vital events in census-based longitudinal studies can bias subsequent analysis", *Journal of Clinical Epidemiology* **61** (4):380-385.
Obermeyer, Z., Rajaratnam, J. K., Park, C. H. *et al.* 2010. "Measuring adult mortality using sibling survival: A new analytical method and new results for 44 countries, 1974–2006", *PLoS Medicine* **7** (4):e1000260.
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.
Phillips, H. 1997. "Demographic survey data in South Africa: An evaluation of methodology and quality", *Southern African Journal of Demography* **7** (1):1-10.
Preston, S. H., Heuveline, P. and Guillot, M. 2001. *Demography: measuring and modeling population processes* Australia: Blackwell Publishing Ltd.
Queiroz, B. L. 2011. "Estimating maternal mortality differentials using census data: experience in Honduras", *Journal of Population Research* **28** (1):75-87.
Reniers, G., Masquelier, B. and Gerland, P. 2011. "Adult mortality in Africa," in Rogers, R. G. and Crimmins, E. M. (eds). *International Handbook of Adult Mortality.* Netherlands: Springer pp. 151-170.

Saloojee, H., Kahn, K., Clark, S. *et al.* 2005. *INDEPTH Verbal Autopsy causes of death MONOGRAPH: Data template for Agincourt.* Accra: INDEPTH NETWORK. http://www.samclark.net/CV/BOOK%20CHAPTERS/Saloojee%20et%20al_2005_chapter%20for%20indepth%20verbal%20autopsy%20causes%20of%20death%20monograph.pdf.

Sankoh, O. and Byass, P. 2012. "The INDEPTH Network: Filling vital gaps in global epidemiology", *International Journal of Epidemiology* **41** (3):579-88.

Sartorius, B., Kahn, K., Vounatsou, P. *et al.* 2010. "Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992-2007", *Glob Health Action* **3**

Stanton, C., Hobcraft, J., Hill, K. *et al.* 2001. "Every death counts: measurement of maternal mortality via a census", *Bulletin of the World Health Organization* **79** (7):657-664.

Statistics South Africa 2012a. *Census 2011 Post-enumeration survey: Results and methodology.* 03-01-46. Pretoria: http://www.statssa.gov.za/Publications/P03014/P030142011.pdf.

Statistics South Africa 2012b. *Census 2011 Statistical release (Revised).* P0301.4. Pretoria: Statistics South Africa. http://www.statssa.gov.za/Publications/P03014/P030142011.pdf.

Stycos, J. M. 1993. "Sample surveys for social science in underdeveloped areas," in Bulmer, M. and Warwick, D. P. (eds). *Social research in developing countries : surveys and censuses in the Third World.* London: UCL Press, pp. 55-66.

Timæus, I. M. 1991a. "Estimation of adult mortality from orphanhood before and since marriage", *Population Studies* **45** (3):455-472.

Timæus, I. M. 1991b. "Estimation of mortality from orphanhood in adulthood", *Demography* **28** (2):213-227.

Timæus, I. M. 1991c. "Measurement of adult mortality in less developed countries: a comparative review", *Population Index* **57** (4):552-68.

Timæus, I. M. 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. M. 1993. "Adult mortality," in Foote, K. A., Hill, K. H. and Martin, L. G. (eds). *Demographic change in Sub-Saharan Africa.* Washington D.C.: National Academies Press, pp. 218-55.

Timæus, I. M., Dorrington, R. E. and Hill, K. H. 2013. "Introduction to adult mortality analysis," in Moultrie, T. A., Dorrington, R. E., Hill, A. G. *et al.* (eds). *Tools for Demographic Estimation.* Paris: International Union for the Scientific Study of Population, pp. 191-194.

Timæus, I. M. and Nunn, A. J. 1997. "Measurement of adult mortality in populations affected by AIDS: An assessment of the orphanhood method", *Health Transition Review* **7** (2):23-43.

Tollman, S. M., Kahn, K., Garenne, M. *et al.* 1999. "Reversal in mortality trends: Evidence from the Agincourt field site, South Africa, 1992-1995", *AIDS* **13** (9):1091-1097.

U.S. Bureau of the Census 1985. *Evaluating Censuses of Population and Housing.* Statistical Training Document ISP-TR-5. Washington, D.C.: U.S. Bureau of the Census. http://www.census.gov/srd/papers/pdf/rr85-24.pdf.

United Nations 1982. *Provisional guidelines on standard international age classifications.* ST/ESA/STAT/SER.M/74. New York: United Nations. http://unstats.un.org/unsd/publication/SeriesM/SeriesM_74e.pdf.

United Nations 1983. *Manual X: Indirect techniques for demographic estimation.* ST/ESA/SER.A/81. New York: United Nations.
United Nations 1989. *Approaches to collecting data on fertility and mortality for the estimation of vital rates.* IIVRS Technical report 37. New York: United Nations Statistical Office. http://unstats.un.org/unsd/demographic/CRVS/IIVRS%20papers/IIVRS_paper37.pdf.
United Nations 1990. *Step-by-Step Guide to the Estimation of Child Mortality.* ST/ESA/SER.A/107. New York: United Nations.
United Nations 2002. *Methods for estimating adult mortality.* ESA/P/WP.175. New York: United Nations.
United Nations 2004. *Handbook on the collection of fertility and mortality data.* ST/ESA/STAT/SER.F/92. New York: United Nations. http://unstats.un.org/unsd/demographic/workshop/Ghana2004/docs/Handbook%5FFertility%5FMortality.pdf.
United Nations 2008a. *Africa addendum to the United Nations principles and recommendations for population and housing censuses.* ACS/WPS2/2008. Addis Ababa, Ethiopia: Economic Commission for Africa.
United Nations 2008b. *Principles and recommendations for population and housing censuses.* ST/ESA/STAT/SER.M/67/Rev.2. New York: United Nations.
Vemuri, M. D. 1994. "Data collection in census: A survey of census enumerators", *Economic and Political Weekly* **29** (51/52):3240-3248.
Ward, P. and Zaba, B. 2008. "The effect of HIV on the estimation of child mortality using the children surviving/children ever born technique", *Southern African Journal of Demography* **11** (1):39-73.
Williams, J. R., Schatz, E. J., Clark, B. D. *et al.* 2010. "Improving public health training and research capacity in Africa: a replicable model for linking training to health and socio-demographic surveillance data", *Glob Health Action* **3** (1):5287-93.
Wits Rural Public Health and Health Transitions Research Unit 2011. "*About the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) and the Agincourt Health and Demographic Surveillance System (HSDSS)*," http://www.agincourt.co.za/index.php/about/#Decentralised Health Systems Development 1992-96 Accessed: 20 Dec 2012
Zaba, B., Whitworth, J., Marston, M. *et al.* 2005. "HIV and mortality of mothers and children: Evidence from cohort studies in Uganda, Tanzania, and Malawi", *Epidemiology* **16** (3):275-280.

4-1 Has any member of this household died in the past 12 months? Y | N | DK

SKIP: If "Yes" - continue to 4-2
If "No" or "DK" - END INTERVIEW

4-2 How many members of this household have died in the last 12 months? [ ][ ] people

	4-2a		4-2b	4-2c	4-2d	4-2e	4-2f	4-2g	
	What were the first names and surnames of the household members who died? *List the names of each person who died*		What was the month and year of <name>'s death? *Answer as precisely as possible. Leave blank for any part of the date not known*	How old was <name> when <name> died? *Leave blank if age not known*	Did <name> die from an accident, or through violence?	What is the sex of the deceased?	Did <name> die while pregnant, or within six weeks of delivery? *Only ask this of women aged between 12 and 50 at the time of death. If not a woman aged 12-50, go to 4.2g*	Did anyone register the death of <name> at Home Affairs, or apply for, or receive, a death certificate for <name>?	
1	FIRST NAME	SURNAME	M M Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	1
2	FIRST NAME	SURNAME	M M Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	2
3	FIRST NAME	SURNAME	M M Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	3
4	FIRST NAME	SURNAME	M M Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	4
5	FIRST NAME	SURNAME	M M Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	5
6	FIRST NAME	SURNAME	D D Y Y Y Y	A A A	Y N	M F	Y N DK	Y N DK	6

4

AGINCOURT HDSS – CENSUS DATA VALIDATION STUDY

Operational data

Household ID	[ ][ ][ ][ ][ ][ ][ ]	**Fieldworker ID**	[ ][ ][ ][ ]
Village	[ ][ ]	**Field Team**	[ ]
		Supervisor	[ ][ ][ ][ ]
Date – 1st visit	D D M M Y Y Y Y	**Outcome**	[ ]
Date – 2nd visit	D D M M Y Y Y Y	**Outcome**	[ ]
Date – 3rd visit	D D M M Y Y Y Y	**Outcome**	[ ]

Outcome Codes
1 – visit successful; completed
2 – visit successful; not complete
3 – no-one present
4 – no adult available to answer
5 - refusal

Questions about the respondent

1-1 What is your date of birth? D D M M Y Y Y Y | DK

What is your age, in completed years? A A A

1.2 Are you male or female? M F

1-3 Is your own biological mother still alive? Y N DK

1-4 When was she born? D D M M Y Y Y Y | DK → 1.4a – If alive: How old is she now? A A A DK

SKIP: If QUESTION 1-3 ="Yes" (Mother alive), GO TO QUESTION 1-9

1-5 Was your mother alive on your 18th birthday? Y N DK NA — NA if under 18

1-6 Was your mother alive when your first child was born? Y N DK NA — NA if no children

1-7 Was your mother alive at the time of the first elections of 1994? Y N DK NA — NA if mother born after 1994

1-8 When did she die? D D M M Y Y Y Y | DK → 1.8a - How old was she when she died? A A A DK

1-9 Is your own biological father still alive? Y N DK

1-10 When was he born? D D M M Y Y Y Y | DK → 1.10a – If alive: How old is he now? A A A DK

SKIP: If QUESTION 1-9 ="Yes" (Father alive), GO TO PAGE 2

1-11 Was your father alive on your 18th birthday? Y N DK NA — NA if under 18

1-12 Was your father alive when your first child was born? Y N DK NA — NA if no children

1-13 Was your father alive at the time of the first elections of 1994? Y N DK NA — NA if father born after 1994

1-14 When did he die? D D M M Y Y Y Y | DK → 1.14a - How old was he when he died? A A A DK

Administrative data

First entry data typist [ ]

Second entry data typist [ ]

1

	2-1		2-1a	2-2		2-3	2-4	2-5					2-6	2-7				
	Write the names of all people who slept here last night. Members of the household who are absent overnight, for example working, travelling or at an entertainment venue, are to be counted in the household if they return to it the next day *Start with respondent. Underline the name of the household head*		Did <name> usually live here between August and November 2008?	What is <name>'s date of birth and age in completed years? *Answer as precisely as possible. Leave blank for any part of the date, or age, not known*		What is <n>'s sex?	Is <name>'s own biological mother alive? *If yes or DK, go to Q2-6*	When did <name>'s mother die? *Answer as precisely as possible. Leave blank for any part of the date not known*	How old was <name>'s mother when she died? *Leave blank if not known.*	Did <name>'s mother die before <name> turned 18? *NA if <name> is 17 or younger*	Was <name>'s mother alive when <name>'s first child was born? *NA if <name> has had no children*	Was <name>'s mother alive at the time of the first elections in 1994? *NA if <name>'s mother born after 1994*	Is <name>'s own biological father alive? *If yes or DK, go to Q3-1*	When did <name>'s father die? *Answer as precisely as possible. Leave blank for any part of the date not known*	How old was <name>'s father when he died? *Leave blank if not known*	Did <name>'s father die before <name> turned 18? *NA if <name> is 17 or younger*	Was <name>'s father alive when <name>'s first child was born? *NA if <name> has had no children*	Was <name>'s father alive at the time of the first elections in 1994? *NA if <name>'s father born after 1994*
1	RESPONDENT	SURNAME	Y N DK															
2	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
3	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
4	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
5	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
6	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
7	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
8	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
9	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
10	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
11	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA
12	FIRST NAME	SURNAME	Y N DK	D D M M Y Y Y Y	A A A	M F	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA	Y N DK	D D M M Y Y Y Y	A A A	Y N DK NA	Y N DK NA	Y N DK NA

	3-1			3-2	3-3				3-4				3-5			
	Check whether <name> is a WOMAN AGED 15-49 (or with birth year between 1959 and 1994). Mark Q3-1 column accordingly. *Write names exactly as above*		*If Yes, go to Q3-2. If No, go to next person*	Has <name> ever given birth? *If Yes, go to Q3-3. If No or DK, go to next person*	How many sons has <name> ever given birth to ... in total? *If NONE ("00") or DK, go to Q3-4*	who are alive and living with her	who are alive and living elsewhere	who have died	How many daughters has <name> ever given birth to ... in total? *If NONE ("00") or DK, go to Q3-5*	who are alive and living with her	who are alive and living elsewhere	who have died	What was the date of birth of <name>'s last-born child? *Answer as precisely as possible. If month and year of birth are not known, ask how many years ago <name>'s last-born child was born*	How many children did <name> give birth to at that time?	How many of those children are still alive?	
1	RESPONDENT	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			1
2	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			2
3	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			3
4	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			4
5	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			5
6	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			6
7	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			7
8	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			8
9	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			9
10	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			10
11	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			11
12	FIRST NAME	SURNAME	Y N	Y N DK									D D M M Y Y Y Y or T T			12

2 3