ASSESSMENT OF QUALITY OF DATA:
THE CASE OF THE 2008-09 KDHS

BY

MUGO EDWIN WAWERU

Q56/64546/2010

RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN POPULATION STUDIES AT THE POPULATION STUDIES AND RESEARCH INSTITUTE (PSRI)

UNIVERSITY OF NAIROBI

P.O. BOX 30197

NAIROBI

NOVEMBER 2012

DECLARATION

I declare that this research project is my own original work. It is being submitted for the degree of Master of Science in Population Studies at the University of Nairobi. To the best of my knowledge, it has not been submitted before in part or in full for any degree or examination at this or any other university:

Candidate **Signature** **Date**

MUGO EDWIN WAWERU Nov 29, 2012

(Q56/64546/2010)

This research project has been submitted for examination with our approval as University Supervisors:

Supervisor's Name **Signature** **Date**

Dr. WANJIRU GICHUHI 29/11/12

Population Studies and Research Institute

University of Nairobi

Mr. BEN OBONYO JARABI 29 November 2012

Population Studies and Research Institute

University of Nairobi

DEDICATION

I dedicate this project to my wife Magdaline Chepchirchir, daughter Cicily Wanjiku and son Davidson Mugo Jnr. for your presence, encouragement and moral support, patience and understanding throughout the period when I wrote this report. Your constant cheer and prayers inspired me to push on.

To my parents Davidson Mugo and Patricia Wairimu, without whom I never would have been - you shaped my life through immense sacrifice, care and support. And through guidance and example, taught me many a thing, values and virtues that continue to shape my character.

To my little sister Purity Waringa and brother Ben Ateku and your very own, Claire and Collins, for your loads of support and encouragement. Not even distance ever mattered.

And not to forget, to *cucu* Milka Wanjiku and *cucu* Elizabeth Waringa, great women who have had a big influence on my life, ever since.

ACKNOWLEDGEMENT

I am greatly thankful to the Almighty God, my Wonderful Counsellor and ever present help, for whom I continue counting the blessings.

I am greatly indebted to my supervisors, Dr Wanjiru Gichuhi and Mr Ben Obonyo Jarabi for their inspirational guidance that made me realise the completion of this project. Your perseverance and understanding is greatly appreciated.

Special thanks go to PSRI staff, Dr Kimani Murungaru and Mr Andrew Mutuku who listened and offered advice when it mattered most, and Dr Anne Khasakhala for her role at the formative stage of the study. I am also grateful to them and the rest of the staff at PSRI: Prof John Oucho, Dr Lawrence Ikamari, Dr Agwanda Otieno, and Mr George Odipo for their academic influence-cum-nourishment that moulded and provided me with strong background knowledge in Population Studies and Research during my master's degree programme.

And to the PSRI library staff – Jane Ochieng, Mildred Namukoa and Alice Okong'o, thank you for your varied assistance and sacrifice to accord an extra minute and hour when it mattered most.

Many thanks to my colleague students at PSRI: Janet Naisoi, Victoria Amwoliza, Francis Kamunya, Milton Adieri, Chris Agweyu, Irene Maeri, Charles Ogolla, Beatrice Odaa and Dhillon Nyachieo. Not to forget Fidelis Aluoch, Charles Obila, Cate Mbaire, Emma Watathi, Ken Lwaki, Janeth Nyarinda, Sharon Lidoro, Faith Atieno, Naomi Makau, Naomi Kiraka, Roy Meoly and W. Ohon. I treasure the valuable experiences shared in and out of class and in those invaluable group discussions. To Dennis Odago, I still remember your humility, smiles and the moments shared, however short-lived. Sad you left us early; may your soul rest in peace brother, and as you would intonate, "My friend..."

To you all, I once again say, "Thanks and God bless you abundantly!"

ABSTRACT

This study focused on the assessment of the quality of the Kenya Demographic Health Survey (KDHS) data, specifically the 2008-09 survey. It set out three objectives: to determine the extent of age heaping or digit preference for males and females in the 2008-09 KDHS; to examine age misreporting and transfers of respondents across age boundaries; and to determine sex ratios by age in the 2008-09 KDHS data. It utilised the Myers' Blended Method for data quality checks for ages given in single years and the age and sex ratios and the United Nations Joint Score methods to assess the quality of age reporting in five-year groups.

The study established that the 2008-09 KDHS age data is highly inaccurate, with more women than men having their ages reported as either unknown or not given at all and age heaping rampant among males and females across the ages. Higher heaping was observed in even age groups compared to the odd age groups for both sexes. Females generally misreported their ages more compared to the males in the survey. Respondents had preferences for ages ending in terminal digits 0 and 5 for the males and 0, 5 and 8 for the females, with the exception (for both sexes) of ages 5 and 15 years. Ages 55, 48 and 68 years for the females were other exceptions to this observation. Overall, males and females avoided ages ending in terminal digits 1, 7, and 9.

The data is also characterised by systematic errors brought about by age misreporting as is evidenced by age ratio values. For the males, there were preferences for the age groups 30-34 and 70-74 years resulting in unusually more than the expected numbers while age groups 65-69 and 75-79 years were avoided giving way to unusually fewer than the expected numbers. On the other hand, females reported preference for the 10-14, 20-24, 50-54 and 70-74 years age groups, and avoided the 15-19, 55-59 and 75-79 years age groups. In terms of numbers, females outweighed males all through except at birth. This in turn suggests that individuals concerned had their ages carried across age group boundaries, either to the next lower or higher age group, a character more pronounced among females compared to males. The errors detected in the 2008-09 KDHS data are therefore likely to have compromised its quality and the accuracy of the various demographic measures derived out of it.

The study therefore calls for intensive training of KDHS enumerators in order to reduce errors pertaining to respondents' ages in the future. Often, in estimation of ages, they base their figures on physical attributes, marital status among others, but it would be desirable if they sought documentary proof when in doubt. The study recommends too that populations be educated through mass media on the need to report their ages as accurately as is possible. Cultures or traditions that influence misinformation on age should be discouraged.

It is prudent too that other methodology be employed to assess KDHS data to confirm the study findings and correct the errors for better and quality DHS data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1: Background

The quality of any data is of paramount interest in that data is the foundation upon which all scientific research is built. Once data is processed, inferences, generalisations or conclusions are made and whose reliability or validity is dependent on the quality of data. Quality data is most likely to lead to objectivity in problem analysis which in turn will lead to reaching objective decisions. Poor quality data will most likely lead to incorrect inferences while decisions based upon such data will be misleading.

Data on age by sex are important for the description and analysis of various types of demographic data (mortality, fertility, nuptiality and migration), and for the evaluation of the quality (that is, completeness and accuracy) of the census counts on population (Shryock and Siegel, 1976; Siegel and Swanson, 2004). For example, social scientists have an interest in population's age structure, planning for community institutions and services are dependent on age composition, and age is important in measuring potential manpower, school or voting population. Age data are also required for preparing current population estimates and projections- such as of households, school enrolment and health services requirements (Shryock and Siegel, 1976). A population's age data helps in calculating the dependency ratio, especially when the actual size of the working population is unknown (CBS, 2002).

According to Magadi (1990), data quality for censuses and surveys is influenced directly or indirectly by demographic, environmental, socio-economic and cultural factors; and demographic data compiled by national population censuses in developing countries such as Kenya are often subject to various limitations arising out of age misreporting and coverage errors. For example, Ewbank (1981) established that some Asian and African populations experienced overreporting of females in childbearing years, which indeed would have an effect on the quality of censuses and surveys through a tendency to exaggerate the ages of women

15-29 years — by pushing them into the 25-34 year age groups, probably to make their ages consistent with expectations regarding age at marriage and fertility. Elsewhere, a study in India by Ghosh (1967) pointed out that age misstatement was influenced by the age, sex and marital status of the person. Of the problems emanating from faulty data, Bairagi et. al. (1982) would acknowledge, age misreporting stands out as major concern.

Complete and accurate reports of ages are critically important for demographic and health surveys (DHS) surveys, with eligibility for inclusion in the survey of women age 15-49, as well as most surveys of men and special surveys, dependent on the age given in the household survey. Both the numerators and the denominators of age-specific fertility rates, infant mortality rates, and other rates depend on reported age. In addition, the quality of the reports of ages reflects on the quality of other information in the surveys (Pullum, 2006).

Errors on age may arise from failure to record age and misreporting (by respondent or erroneous estimation and/or allocation by enumerator and/or office respectively) of age (Shryock and Siegel, 1976). The UN (1955) states that statistics classified by age groups may be affected by errors in age reporting and by variations in completeness of enumeration, or of recording of vital events, for the different age groups.

The quality of DHS data is of utmost importance to researchers, policymakers and programme managers as well as for planning purposes in developing countries like Kenya (CBS, 1996b). Indeed, most of the statistics produced by DHS surveys depend on accurate reporting of ages of women age 15-49 years and children (Pullum, 2006). But as the KNBS and ICF Macro (2010) aptly point out, estimates from a sample survey such as the DHS are affected by non-sampling errors. Non-sampling errors - mistakes made in implementing data collection and data processing - include the misunderstanding of the survey questions such as on age by either the interviewer or respondent and data entry errors; all which in turn compromise data quality. Though efforts were made at implementation of the 2008-09 KDHS survey to rid it of non-sampling errors, it was impossible to avoid them altogether (KNBS and ICF Macro, 2010).

Censuses and surveys may have problems that include not only vagueness, the tendency of respondents or enumerators to report certain ages at the expense of others, otherwise called age heaping or age preference or digit preference but also "complete ignorance" (Magadi, 1990; Shryock and Siegel, 1976). Indeed, there is low quality of age data in many censuses and surveys, largely due to a genuine inability of the respondent or the proxy to report the exact age(s) (UNFPA, 1993).

Systematic age preferences can couple with constant age biases to markedly distort age distributions (Bairagi et. al., 1982). Indeed, some African and Asian populations are marked by an overreporting of females in the childbearing years with net transfers out of the 10-14 and 45+ or 49+ age categories (Ewbank, 1981). Such large net transfers of females into the childbearing years have an obvious harmful effect on various fertility estimation procedures. Both bias and random error in the age statements for young children can also have great importance in demographic investigations (Bairagi et. al., 1982).

In an analysis of three sets of demographic estimates for Pakistan, that is the Population Growth Surveys of 1968 and 1971; the Census of 1972 and; the 1973 Housing, Economic, and Demographic Survey, Retherford and Mirza (1982) found systematic distortions in the estimates caused by patterns of age exaggeration that increased with age. They established that in Pakistan, there was very inaccurate reporting of ages of children for whom births were estimated. For children aged 0-14, there was heaping noticeable particularly on ages 8, 10 and 12 years, while for women, age heaping was systematically biased upward, in the form of age exaggeration that increased with age. Further, Retherford and Mirza (1982) state, in the Pakistan Fertility Survey of 1975, only 6 per cent of women knew their birth date, and either the respondent or interviewer had to guess the ages of the remaining 94 per cent.

One of the major difficulties in African censuses and surveys is age measurement (Magadi, 1990). In almost all African cultures, numerical age has had no importance over the years, contributing to the many errors in African censuses and surveys (Kpedekpo, 1982). An analysis of Kenya's 1962, 1969, 1979, 1989 and 1999 Population and Housing censuses has repeatedly pointed at poor quality of age data generated, with improvements on 1962, and only slightly between 1969 and 1989 (Central Bureau of Statistics, 1996a; CBS, 2002). In all the years, the Whipple's indices (measures the degree of heaping on ages ending in 0 and 5) obtained were high indicative of either "rough" or "very rough" age data for both male and female populations. For the 1989 census, the CBS (1996a) analysis of age reporting by sex found preference for digits 0 and 5 by females that were higher than those of the males. Similarly, the 1999 census had higher concentration of age misplacement in ages 20, 30, 40, 50 and 60 years (CBS, 2002).

The age-ratio scores for the last three censuses pointed at modest deterioration in age reporting, especially during the 1999 census (CBS, 2002). Analyses using the United Nations index method (measures the fluctuations in age and sex ratios) found all the aforementioned censuses' data either "inaccurate" or "highly inaccurate" (CBS, 1996a; CBS, 2002).

Indeed, the CBS (undated) analysis of the 1979 Population Census established that misreporting of ages distorted the reported age-sex distribution. The report thereof states that in Kenya, many people do not know their ages precisely, implying that entries made of their ages are guesswork, at times assisted by use of event calendars. Further, the evaluation established that among others, there was; a marked age heaping on round numbers ending in 0 or 5; overstatement of ages of young children; general exaggeration of ages among the middle aged and elderly, that is more pronounced for men than for women, giving high sex ratios for the older age groups; and an overstatement of age among adolescent girls and young women, resulting in low sex ratios between the ages of 16 and 30 years (CBS, undated).

1.2: Statement of the Problem

Age by sex data is important in that the estimates so obtained are specific to the population and are used for programmatic planning, policy, research purposes and planning for socio-economic development. Consequently, owing to its importance, KDHS data should be reported as precisely or concisely as possible. Otherwise, age misreporting can adversely affect various demographic measures. For example, children's age 0-5 should be accurate to ensure estimates such as neonatal and child mortality estimates are plausible; while females' ages must be as accurate as possible to avoid unnecessarily transferring women into reproductive years which may in turn affect various fertility estimates Similarly, programmatic planning and its budgeting for children, adolescents, youths, men and women ages 15-49 years will require flawless data on age to inform on who in the population falls into these categories.

However, data on age in Kenya has had quality issues. Age is often misreported owing to the fact that many people do not know their ages precisely, which implies that entries on age and therefore the reported age-sex distribution on census schedules and the KDHS among others are likely to be distorted. This is manifested in age heaping around 0 and 5, overstatement and/or exaggeration of ages of young children and of males and females in general.

While analytical quality checks on the accuracy of reported ages are done for each census, no such efforts are evident for the 2008-09 KDHS, an exercise that is necessary to authenticate estimates so obtained. This study seeks to establish the extent of age heaping/digit preferences, misreporting and/or misstatement at the household level and any irregularities in age-sex distribution in the 2008-09 KDHS.

1.3: Research Questions

The study sought to answer the following questions;

a) What is the extent of age heaping or digit preference and age misreporting in the 2008-09 KDHS?

b) Are there significant differentials in age heaping or digit preference by sex in the 2008-09 KDHS data?

c) How consistent are the sex ratios from the 2008-09 KDHS data?

1.4: Objectives of the Study

I he aim of the study is: To carry out an assessment of the quality of data of the 2008-09 KDHS"

Specific Objectives of the Study

The study specifically aims to:

a) Determine the extent of age heaping or digit preference for males and females in the 2008-09 KDHS.

b) Examine age misreporting and transfers of respondents across age boundaries.

c) Determine the sex ratios through the different ages in the 2008-09 KDHS data.

1.5: Justification of the Study

The age-sex structure determines a population's needs and the potential for future growth of the total population and of specific age groups. Consequently, information on age reporting is pivotal to effective development planning, programming and research, as well as in policy formulation. Indeed, planning by and for private and public sectors, such as the military, community institutions, and services like health and sales programmes require separate age by sex data. For example, the government will require age and sex classification of data to plan for a roll out of a HIV and AIDS programme, while an hotelier will require similar disaggregated data to guide market research and therefore estimate or project demand for services within a community. Consequently, an evaluation of the extent of distortions in reported age by sex data will inform users of the data of its limitations and guide future censuses and surveys and specifically in our case, future DHSs in Kenya.

On the other hand, many demographic measures are age-specific, such as estimates of age-specific fertility rates, neonatal and child mortality. That is, information on age is a basic variable in constructing many demographic parameters. Further, tabulations on age are required in the computation of basic measures relating to population change factors, in the analysis of labour supply factors and in economic dependence studies. Similarly, most indicators produced by DHS surveys depend on accurate reporting of ages of women and children. However, estimates of levels (or differentials) and trends in such rates may be affected by misreporting of the ages of populations, affecting reliability of derived estimates. For example, since standard age intervals begin with (preferred) numbers ending in digits 0 and 5, misreporting can shift women into the next higher age interval. Age displacement of women can seriously distort estimates of current levels and recent trends in fertility and mortality.

An assessment of the quality of age reporting will therefore make aware users of the KDHS of its limitations. This is despite the oft highly quality of training enumerators are taken through and which is aimed at bringing out DHS data of high quality. The study is indeed timely in that it will inform programmes on among others, child health as well as on reproductive health touching on women that include family planning needs for specific age groups.

1.6: Study Limitations and Assumptions

The study examined the quality of age and sex reporting in the 2008-09 KDHS data to identify evidence of heaping and misreporting of ages. The study was only limited to omission and misreporting of ages. And whereas Myers', Whipple, Bachi, Carrier and Ramachandran indices are often used to detect digit preferences or irregularities in reporting age in single years, this study only utilised Myers' index for which literature suggests is most preferable.

For the 2008-09 KDHS, a representative sample of 9,057 households was drawn with a total of 38,515 persons analysed for age reporting, which is only about a 0.10 per cent of the total population of 38.6 million people (based on the 2009 population estimates by KNBS). Of the total, 18,774 are male and 19,741 female. It is however assumed that the degree of reporting for those interviewed is the same as for the rest of the national population. Otherwise, this will be a limitation if the assumption is not true.

CHAPTER TWO

LITERATURE REVIEW

2.1: Introduction

This chapter reviews findings of past research and studies on age by sex reporting, and the attendant data quality issues. Generally, the same issues cutting across census data are found in surveys like the Demographic and Health Surveys (DHS). The chapter discusses issues such as underreporting (avoidance of certain age with a consequent lower number of people than expected) or overstatement of age (stating an age higher than the actual), heaping, age misstatement and misreporting and their influences on data quality, starting with the general and moving through, the specific categories of individual populations. Specifically, the chapter examines general issues of data quality, age reporting for children, males and females and among the aged.

2.2: General Issues on Data Quality

Sources of errors in surveys and censuses are numerous. And whereas it is easy to obtain information on sex, this is not always the case in reporting of age, for there arises various forms of error and bias. According to Bairagi et. al. (1982) and UNFPA (1993), simple random age errors develop from among others, design of the questionnaire, coding errors, data processing, the interviewer and respondent's (or that of proxy) inaccuracies in reporting the correct age. In retrospective inquiries, recall lapses further contribute to response errors.

Further, a large amount of random error is common in age data in developing countries (Bairagi and Rahman, 1974; You, 1959), while questions on age may elicit different interpretations in different cultures (UNFPA, 1993) In many developing countries, the UNFPA (1993) goes on, exact knowledge of age is not important and birth registration is rare, rendering it difficult to obtain information on age. In such scenarios, age is often approximated or even non-numeric; whereas in situations that a main respondent has to supply information on household members' ages, the proxy reporting weighs in to age misreporting. This is borne out of the proxy's ignorance of own and other household members' ages. Matters are not made any easier in such countries where literacy levels are still low, birthdays are rarely noted, and if noted, at times in local calendar systems (different from the western "solar" calendar) and enumerators often guess respondents' ages.

According to Wamai (2004), the importance of age as a variable in demographic analysis cannot be underrated. Poor quality of age data, she reckons, will certainly and significantly reduce the accuracy of such important population estimates as fertility and mortality. She suggests that evaluation of age data quality be carried out prior to carrying out any analytical work. Further, Pardeshi (2010) states that age-related data often suffer from misstatements and irregularities compromising accuracy in censuses and surveys. Age heaping which is one such irregularity is considered to be a measure of data quality and consistency. Heaping as well as other constant age biases, Bairagi et. al. (1982) continue, act to shift age distribution either up or down contributing to gross movements in and out of specific age categories frequently distorting the age distributions in population censuses and sample survey data.

From age data collected during a community survey in the Yavatmal District, Maharashtra state in India, Pardeshi (2010) established that there was age heaping at ages with terminal digits '0' and '5', indicating a preference in reporting such ages while 42 percent of the population in the six villages sampled reported ages with an incorrect final digit. A UNFPA (1993) report cites considerable heaping at past censuses for digits '0' and '5', with digits '2' and '8' also evidencing some overstatement. Digit '1', the report says showed the greatest amount of understatement, with most of the preference for digit '0' seemingly due to digit '1' rather than digit '9'. At the household level, Pullum (2006) states that the household head or spouse is expected to report information more accurately about himself or herself than about other household members. But in his study of three Bangladesh's DHS, it emerged that the level of heaping at digits '0' and '5' was very low when the respondent gave his or her own age, but very high level of four to six times higher when the same person reported the age of other household members. On the other hand, early American censuses were found to suffer from underreporting of infants, distinct overstatement among those at advanced ages, heaping and the reporting of some individuals as being of unknown age (UNFPA, 1993). Wamai (2004) points out that age misreporting errors in males are lower compared to in females due to the higher literacy level among males.

Such selective under or overenumeration by age, Bairagi et. al. (1982) argues, is a form of aggregative age misreporting. Selective enumeration by age can adversely affect standard Bourgeois-Pichat and Brass fertility estimates and may also lead to peculiarities in selecting stable or quasi-stable populations.

2.3: Age Reporting for Children

In the analysis of age data for 3,393 children six years of age and under in rural Bangladesh for the level and pattern of age misstatement, Bairagi et. al. (1982) found random error, age heaping at whole years, and preferences for particular ages in the data. Variation in age reporting was discovered to increase monotonically with age. Systematic errors in age misstatement displayed modest overstatement for the first four years of life and more pronounced understatement for ages 4, 5, and 6 years. Elsewhere, the UNFPA (1993) points out, early American censuses characteristically suffered from underreporting of the number of children at ages '0' and '1' years.

In an analysis of Turkish censuses for 1935-40 and 1955-60, Demeny and Shorter (1968) concluded that there was exaggeration of age of young children. They had established that there was a deficit at ages 0-4 years, which in turn resulted into an excess of both males and females in ages 5-9 years. Further, and according to Ewbank (1981), an Office of Population Research at Princeton study directed by Ansley Coale and Paul Demeny based on stable population analysis of more than 150 age-sex distributions from censuses and surveys established that the ages of infants and children are "probably" reported more accurately than the age of adults. This, he attributes to their reporting by parents or other adults who remember the birth and that the rapid physiological and psychological changes during childhood making it easier to guess the age with reasonable accuracy. However, and citing various studies, he points out that parents often exaggerated the ages of their children through rounding off, rather than truncating it to the number of completed years, that is, an overstatement by a year for those within 6 months of the reported age. Indeed, the study by Coale and Demeny concluded that the distortions had the tendency to exaggerate the ages of children aged 0-4 years. Varied data from Senegal, Gambia and Ghana showed higher chances of understating age than overstating age above about age 6 years in some cases and 7 or 8 years in other cases.

There was consistency in various countries for preference for even-numbered ages above age 5, that is, for ages 6, 8, 10 and 12 years. Further, a simulation of Coale and Demeny's "North" model life table showed a clear preference for ages 4, 6, 7 and 10 years. In some instances, preference for 7 replaced the usual preferences for 6 and 8. A survey of 4 censuses from North Africa, 6 from South America, and 15 from Asia demonstrated that preference for even-numbered ages is very strong "almost everywhere" Countries also varied in the degree of underreporting of the population aged 0 years and in the degree of accuracy of age reporting among children aged 0 and 1 year. World Fertility Surveys attested to the great variation in the ratios of the population aged 0 years to that aged 1 year and the population aged 1 to that aged 2 years.

2.4: Reporting Age among Males

According to Myers (1951), the percentage of men reported as being of unknown age is higher than that of women owing to the fact that in most cases, wives at home do most of the reporting and may not know their spouses' exact ages.

In a study of the 1989 and 1999 Kenyan census data, Wamai (2004) found that males had the tendency to heap on ages ending in terminal digits 0 and 5, except age 5 years in 1999. The highest heaping occurred in ages 60, 70, 50, 40, 30, 45, 35, 25 and 65 years Ages most avoided were 73, 66, 44, 74, 34, 33, 64, 31, 41 and 46 years, an indication of avoidance for ages ending with terminal digits 1, 3, 4 and 6. Characteristically, the report adds that there was a decrease in age heaping since there was no age heaping for age 5 and 6 years in 1999 compared to 1989.

Preference for terminal digits 7 and 8 was similar in both censuses whereas there was no avoidance of age 2 and 3 years in 1999 but preference for terminal digits 1, 4, 8 and 9 in 1999.

2.5: Age Reporting among Females

In a study of age at marriage in India's West Bengal, Ghosh (1967) established that unmarried women around age 15 years tended to understate their age seriously distorting the age distribution of unmarried females. Coale and Demeny's Princeton study (Ewbank, 1981) established a so-called African pattern (typical of populations in Africa and Southern Asia) in which females were characterized by a "surplus" at 5-9 years, and a deficit in the adolescent age intervals (10-14 and 15-19 years), followed by a surplus in the central ages of child bearing (25-34 years). This latter study found evidence of exaggeration of the ages of girls 10-14 years if they passed puberty and an understatement of the ages of those who had not reached puberty; and a tendency to exaggerate the ages of women 15-29 years, "probably to make their ages consistent with age at marriage and fertility". The study attributed this to the fact that in societies in which age is unimportant, ages of young women are frequently estimated by their physical maturity, their union status, or their parity, while migration and formation of new households are thought to be responsible for the relatively large under-enumeration in the 15-29 years age group. On the other hand, the so-called Latin American pattern had general preference for age groups 25-29 and 35-39 years over the age groups 30-34 and 40-44 years, with women surplus reported at 20-29 years.

In the study by Wamai (2004), the highest age heaping among females in the Kenyan censuses of 1989 and 1999 occurred in the ages 60, 70, 50, 40, 30, 45, 35, 20, 25, 65 and 10 years, an indicator of tendency to heap on ages ending with 0 and 5, except in age 5 years. Ages most avoided were 73, 66, 74, 34, 44, 33, 62, 41, 43, 53, 51, 72, 63 and 52 years, a pointer to avoidance for ages ending with terminal digits 1, 2, 3, 4 and 7. However, preferences for terminal digits 7, 8 and 9 were similar for the two censuses.

2.6: Age Reporting Among the Elderly (85+ years)

According to Hill et. al. (2000), age inconsistencies tend to increase slightly with age among those aged 85 years and above. Although apparent for both sexes, this age pattern is more pronounced for males. Myers (1951) agrees on this male-female comparison, saying that there is considerable amount of age overstatement among persons aged 90 years and over. In a study, Rosenwaike and Logue (1983) further established that for the extreme old persons, the older the age at death reported on the death certificate, the greater the average error-the curve of average error plotted against reported age at death rises nearly exponentially

CHAPTER THREE

DATA AND METHODOLOGY

3.1: Introduction

This section examines the data used for the study and methods utilized to analyse the data. It describes in detail the sources of data, its composition and analysis of the quality using various methodologies. Different methodologies are chosen depending on whether the age is in single years or five-year groups. Myers' blended method is used to analyse for quality of age data in single years. This in turn reveals preferences for or avoidance of terminal digits 0 to 9. For age in grouped data, the age and sex ratios and the UN Joint Score method are used to evaluate the quality.

3.2: Data Sources

The study utilized data from the 2008-09 KDHS. Specifically, the household file was used. The data was availed in the form of Statistical Package for Social Scientists (SPSS) software for windows version 16.0. The Household Questionnaire was used to list all the usual members and visitors, with basic information collected on the characteristics of each person listed, including age and sex (KNBS and ICF Macro, 2010). The question that was asked during the enumeration was "How old is (*NAME*)?, where the *NAME* referred to each of all persons who usually lived in the household and guests or temporary visitors of the household who stayed there the night before the survey. Age was recorded in years.

Information on completeness of age data for males was derived from the male file. The men eligible for the individual interviews were actually identified using the Household Questionnaire. The Men's Questionnaire was administered to all men age 15-54 years living in every second household in the sample.

Notably, the questionnaires were translated from English to 10 other local languages- Kalenjin. Kamba, Kikuyu, Kisii, Luhya, Luo, Maasai, Meru, Mijikenda, and Somali- to ensure clarity and ease of understanding of questions by the respondents.

3.3: Sampling

A representative sample of 10,000 households in the country was drawn for the 2008-09 KDHS. The sample allowed for separate estimates for key indicators for each of the eight provinces in Kenya, and for rural and urban areas separately. Fewer households and clusters were surveyed for North Eastern province owing to its sparse population, while urban areas were oversampled to obtain enough cases for analysis.

The Kenya National Bureau of Statistics current master sampling frame for household based surveys- the fourth National Sample Survey and Evaluation Programme (NASSEP IV) was developed (in 2002 from a list of enumeration areas covered in the 1999 population and housing census) on the platform of a two-stage sample design; and the 2008-09 KDHS adopted this design The first stage had selection of 400 data collection points (clusters) - 133 urban and 267 rural- from the national master sample frame. The second stage of selection involved systematic sampling of households from an updated list.

All women age 15-49 years who were either usual residents or visitors present in sampled households on the night before the survey were eligible to be interviewed in the survey. All men age 15-54 years in every second household selected for the survey were eligible to be interviewed.

3.4: Assessing the Quality of the Data

A total of 38,515 responses were analyzed for age reporting, that is, heaping and for evidence of transfers outside the age range of eligibility. Of this total, 18,774 (about 49%) were male and 19,741 (or 51%) were female. The figures include each of all persons (children, men and women), irrespective of age, who usually lived in the household and guests or temporary visitors of the household present the night before the survey.

3.5: Data Analysis

Frequencies and percentages were be used to establish the proportions of incompleteness of data in the reporting of age. Age displacement across all ages for males and females was analyzed using age ratios and sex ratios. A graphical analysis of the respective age ratios (plotted against age-groups) typically highlights the probability of any errors in the reported age data

Myers' Blended Index was be used for analysis of heaping and digit preferences or avoidances for each terminal digit. The UN index of age-sex composition was also used to establish the extent of heaping.

3.6: Completeness of Data on Age

Completeness of any data is a very important indicator of its quality. A person's age, was considered complete if age was indicated and incomplete if the returns on age were missing or if it was not known. Such responses could have been due to ignorance of age or lack of knowledge of exact age and carelessness in reporting and recording (Kpedekpo, 1982; Pollard et al. 1974).

On the other hand, every DHS through the Household Questionnaire examines the completeness of reporting age and birth date among others for women aged 15-49 years (Pullum, 2006). Similarly, the DHS' Men's Questionnaire addresses the completeness of age and date of birth (month and year) for men aged 15-54 years. Ideally, each woman (and man) provides her (his) age in completed years, a year of birth, and a month of birth. At a minimum, there should be an age or a birth year. Noting that some women do not provide all three items, and even if all information is provided, Pullum (2006) observes that there may be inconsistencies that require the imputation of one or even two of the items. As for women, men's age is also imputed for those values with inconsistencies or not provided.

3.7: Measurement of Age and Digit Preference for Age Given in Single Years

Although age in single years is prone to different types of errors such as age misreporting, net underenumeration, and nonreporting or misassignment of age, age heaping remains outstandingly rampant (Shryock and Siegel, 1976). Populations with low education status report high levels of age heaping while patterns for age or digit preference vary from one culture to another with preference for "0" and "5" digit endings for age. On the other hand, digit avoidance may be specific to a people, with the West avoiding 13 and the Orient shunning '4'. Age "0" years is grossly underreported because parents often do not take newborns as regulars in the family and many people disregard 0 as a number like any other

To obtain indices of age preferences, the arithmetic devices developed depend on the assumption of a true distribution of population by age over a part or all of the age range (Shryock and Siegel, 1976). That is, that the true figures form an arithmetic progression or are rectangularly or linearly spread over this range (say 3-year, 5-year or any other age range) which includes and, preferably, is centred on the age under examination.

For example, over a 3-year range, the index of heaping on age say 32 years will be calculated as the ratio of the enumerated population aged 32 years to one-third of the population aged 31, 32 and 33 years. Still, it may be calculated over a 5-year range as the ratio of the enumerated population aged 32 years to one-fifth of the population aged 30, 31, 32, 33, and 34 years Usually, this index is often calculated as a percentage.

Therefore, for any age *x* whose population is P_x, the index of heaping in a 3-year and 5-year ranges respectively will be;

$$\frac{P_x}{1/3(P_{x-1}+P_x+P_{x+1})} \text{ x } 100 \qquad \text{and} \qquad \frac{P_x}{1/5(P_{x-2}+P_{x-1}+P_x+P_{x+1}+P_{x+2})} \text{ x } 100$$

Often, the two indexes are approximately the same, whether a 3-year or 5-year group is used (Shryock and Siegel, 1976).

3.7.1: Myers' Blended Method

Myers' blended method computes for preferences and avoidance of all terminal digits "0" to "9" where age is given in single years. The method derives a blended population that is essentially a weighted sum of the number of persons reporting ages ending in each of the 10 terminal digits (Kpedekpo. 1982, Yusuf F. 1967). It is assumed that barring any irregularities, the blended sum at each of the digits should be 10 percent of the total blended population. Therefore, any excess reflects preference while any shortfall implies avoidance. The Index of Preference, or the overall measure of the extent of digit preference or avoidance in a population is then obtained as the absolute sum (or in some case, half the absolute sum) of deviations for each of the terminal digits In theory, Myers' index can vary between 0 for ages that are reported accurately and 180, for where all ages are reported with the same terminal digit (Pollard et.al., 1974; Kpedekpo, 1982).

In computing the blended populations (Yusuf, 1967), a decision is made about the age range on which to base the computations. Usually, the limits are not less than 10 years and 80 years for the lower and the upper limit respectively. This is because the age-reporting at less than 10 and more than 80 years of age is affected by causes other than digital preference.

Taking the age range 10-79 years, the Myers' Index involves the computation of two series of population totals with a time lag of 10 years. In our case, one series will have a range 10-69 years while the second will have the range 20-79 years. If P(x) is the population at age x, the ten population totals in the first series will be:

Total for digit 0 = P(10) + P(20) +...........+ P(60)

Total for digit 1 = P(11) + P(21) +...........+ P(61)

.

.

Total for digit 9 = P(19) + P(29) +...........+ P(69)

Similarly, the ten population totals in the second series will be:

Total for digit 0 = P(20) + P(30) +............+ P(70)
Total for digit 1 = P(21) + P(31) +...........+ P(71)

.

.

Total for digit 9 = P(29) + P(39) +............+ P(79)

The ten population totals (one for each digit) of the first series (10-69) are then multiplied by weights *1, 2, 3,.........., 10* while the totals of the second series (20-79) are multiplied by *9, 8, 7,......, 1, 0* respectively. The two sets of products are then summed for each terminal digit to arrive at the blended population for that digit. The blended populations for the ten digits are then converted into percent of the total blended population. Myers' Index is derived by summing the absolute differences of the percent blended populations for each terminal digit from 10 percent. This method effectively gives equal weights to each terminal digit. Algebraically, the sum of blended populations (age range 10-79 years) corresponding to the ten terminal digits is equal to the sum of the populations in the ranges 10-69, 11-70, 12-71,......, 19-78 and 20-79 years.

To interpret the results, it is noted that the percent deviation for each digit will be a measure of preference or avoidance for ages ending in each terminal digit. Positive deviation will imply preference while negative deviation is synonymous with digit avoidance.

3.8: Measurement of Age Accuracy for Grouped Data Using Indices

There are several indices for evaluating the age and sex composition. The age, and sex ratios and indices for detecting digit preference in age reporting are some of the principal indices used for evaluating the age and sex composition. They rely on an expected pattern reflecting the distribution of a population without migration and in which mortality and fertility have changed in only one direction. The two indices may be used either separately or jointly in evaluating the quality of a census or survey returns by age groups (Kpedekpo, 1982).

3.8.1: Age Ratio

According to Arriaga (1994), age ratios for 5-year age groups may be used as indices for detecting possible age misreporting in populations where fertility has not fluctuated greatly during the past and where international migration has not been significant. Age ratio is defined as the ratio of the population in the given age group to one half the population in the two adjacent age groups (Kpedekpo, 1982). Mathematically, if ${}_5P_x$ is the age group from age x years to age $x+5$ years, ${}_5P_{x-5}$ and ${}_5P_{x+5}$ the preceding and the following age groups respectively, then

$$\text{Age Ratio} = \frac{{}_5P_x}{1/2({}_5P_{x-5} + {}_5P_{x+5})} \text{ x } 100$$

However, Shryock and Siegel (1976) define the age ratio as the ratio of the population in the given age group to one-third of the sum of the populations in the age group itself and the preceding and following groups, times 100. Consequently for the same age group above,

$$\text{Age Ratio} = \frac{{}_5P_{.}}{1/3({}_5P_{x-5} + {}_5P_x + {}_5P_{x+5})} \text{ x } 100$$

However, in the UN procedure, the previous definition suffices for the age ratio.

The age ratios so computed are then compared with the expected value, usually 100.0, with discrepancies at each age group the measure of net age misreporting. In both cases, the three age groups form a nearly linear series, assuming no extreme fluctuations in past births, deaths or migration. By expecting a value of 100.0, it is assumed that coverage errors are about the same for all age groups. The larger the fluctuations of age ratios, and the larger their deviation from 100, the greater is the probability of errors in the data.

Shryock and Siegel (1976) further came up with an overall age-accuracy index equivalent to taking the average deviation (irrespective of sign) from 100.0 of the age ratios over all ages. Mean deviations are separately calculated for males and females (by dividing the sum of deviations from 100.0 by the number of age groups) and the average of the two mean deviations taken as the overall accuracy of the particular age data.

3.8.2: Sex Ratio

The sex ratio is calculated by taking the number of males in a population and dividing it by the number of females in the same population, usually expressed as the number of males per 100 females (Pollard et.al., 1974). Sex ratios may be calculated separately for various ages or age groups to give age specific sex ratios, with the sex ratio at birth being fairly constant for most countries of the world at around 105 male births per 100 female births. Naturally, mortality is usually higher for males than females, and the sex ratio is reduced continuously up to the oldest ages (Arriaga, 1994).

The age specific sex-ratios obtained are then compared to expected values, the latter being carefully developed estimates (developed principally from vital statistics) or theoretical figures based on a population model (Shryock and Siegel, 1976).

Sex ratios depend largely on the on the number of male and female births (Arriaga, 1994) and the relative mortality of the population and where there is substantial migration, on the age-sex distribution of the migrant intake or outflow (Shryock and Siegel, 1976). In a study among populations with African origin in the US and Europe, Garenne (2003) established that they had lower sex ratios compared with those from other parts of the world. The general pattern of the age specific sex ratios is such that they approximate to the sex ratio at birth in the younger ages, and fall gradually with advanced age (Kpedekpo, 1982). Further, Arriaga (1994) states that the larger the abrupt departure of this ratio from values close to 100, the larger the possibility of errors in the data.

3.8.3: United Nations (UN) Joint Score

The United Nations has further developed an index incorporating measures of accuracies of the age and sex ratios, otherwise called the **UN age-sex accuracy index**, also referred to as the UN joint score. In the index (Shryock and Siegel, 1976), the mean of the successive differences from one age group to the next in reported sex ratios, irrespective of the sign, are taken as a measure of the accuracy of the observed sex ratios, on the assumption that these age-to-age changes should approximate zero.

The UN age-sex accuracy index combines the sum of:

a) the mean deviation of the age ratios for males from 100.0;

b) the mean deviation of the age ratios for females from 100.0, and:

c) three times the mean of the age-to-age differences in reported sex ratios.

That is,

Joint Score = (3 x {sex ratio score}) + (male age ratio score) + (female age ratio score)

According to the UN, the Joint Score is judged on the following scale; Data whose accuracy index is below 20 is termed accurate, from 20 to 40 inaccurate and anything over 40 as highly inaccurate. The age ratios, sex ratios and the UN age-ratio scores for both male and female and the age-sex accuracy indices can be obtained using the computer software programme AGESEX spreadsheet (Arriaga, 1994).

3.9: Correcting Age Distributions

To correct for age misreporting, smoothing or graduation techniques are used. Techniques available either slightly modify or not the total population size. The errors in the age distribution arising from the age heaping are corrected by assuming that the excesses should be redistributed to adjoining ages or age-groups, thereby preserving the unique shape of the age distribution curve and eliminating the irregularities (Bogue and Arriaga, 1993).

Smoothing formulas include those that do not modify totals, that is, the Carrier-Farrag, Karup-King-Newton, and the Arriaga, all which give rather similar results (Arriaga, 1994). The United Nations formula however slightly modifies the total population. In the case where age misreporting is not severe, light smoothing is done to correct the not so significant irregularities, while in the event that there is age misreporting coupled with digit preferences, together leading to severe irregularities, strong smoothing procedures are encouraged. Computer programmes available may be used to forego very extensive numerical computations, one such being the spreadsheet AGESMTH that smoothes population age structure by 5-year age groups (Arriaga et al. 1994).

Single age distribution may be smoothed and adjusted by among others: fitting of a stable population; fitting a succession of polynomials; comparisons with a standard age distribution (UN, 1983). In order to adjust the 2008-09 KDHS data, the spreadsheet AGESMTH was used to smooth age in 5-year groups.

CHAPTER FOUR

RESULTS OF THE ASSESSMENT

In this chapter, results of the analysis of the quality of the 2008-09 KDHS are discussed. Graphical methods, descriptive analysis and population analysis spreadsheets were used to come up with the results.

4.1: Completeness of Age Data

There were a total of 18,774 males and 19,741 females for whom age analysis was done. Table 4.1 presents the frequency distribution of males and females and data on persons whose ages were missing or who did not know their ages.

Table 4.1: Distribution of Completeness of Age by Sex

	Frequency		
	Males	Females	Total
Age reported	18,768	19,722	38,490
Don't Know	1	9	10
Missing	5	10	15
Total	**18,774**	**19,741**	**38,515**

For the male population, it is evident that one person did not know ("DK"- don't know) his age while another five (5) had their ages missing. On the other hand, nine females did not know their ages while 10 had their ages missing in the returns from the household questionnaire.

4.1.1: Completeness of Age Data

Demographic Health Surveys examine completeness of age data for women in the reproductive years 15-49 and that of men in the years 15-54. For the 2008-09 KDHS, the observations in Table 4.2 were made on how age was reported and any subsequent imputation done. Notably, all women and men in the 15-49 and 15-54 age groups sampled for the survey had their ages falling into the categories shown in the table. Consequently, none had age reported in such a manner that; *Year and age were given, year ignored*; *Year given, age and month imputed*; *Month given, age and year imputed*, and; *None given, all imputed.*

Table 4.2: Types of Age Imputations for Men Aged 15-54 and Women Aged 15-49 Years

	Frequency		Percent	
	F	M	F	M
Age, month and year given and okay	6190	2588	69.24	74.69
Month and age given, year imputed	94	8	1.05	0.23
Year and age given, month imputed	2145	850	23.99	24.53
Age given, year and month imputed	511	19	5.72	0.55
Total	**8940**	**3465**	**100**	**100**

From Table 4.2, more males (nearly 75%) compared to females (at 69%) had their responses on age fully complete such that no imputation had to be made. Responses that had age and year given but month imputed had near equal proportions for both males and females (23.99% and 24.5% respectively). That for which only age was given and year and month of birth being imputed had males proportionately outnumbering females 10 times; and the case for which month and age were given and only the year had to be imputed had the proportion of males outweigh females by nearly five times (1.05% and 0.23% respectively).

It can, therefore, be concluded that men had their ages reported more completely compared to females An assumption is made that the degree of completeness in reporting the ages for both males and females is uniform throughout the ages, this considering that the 2008-09 KDHS considered completeness only for males aged 15-54 years (in the men's file) and females aged 15-49 years (household file) only.

4.1.2: Completeness of Information by Educational Level

Table 4.3 below presents completeness of age information cross-tabulated by highest educational level The percentages represent the proportions for males and females out of the respective totals for each. That is, any percentage for males is taken out of the total number of males aged 15-54 years while percentages for males represent individual figures taken out of the total number of females aged 15-49 years.

Table 4.3: Percent Completeness of Information by Highest Education Level

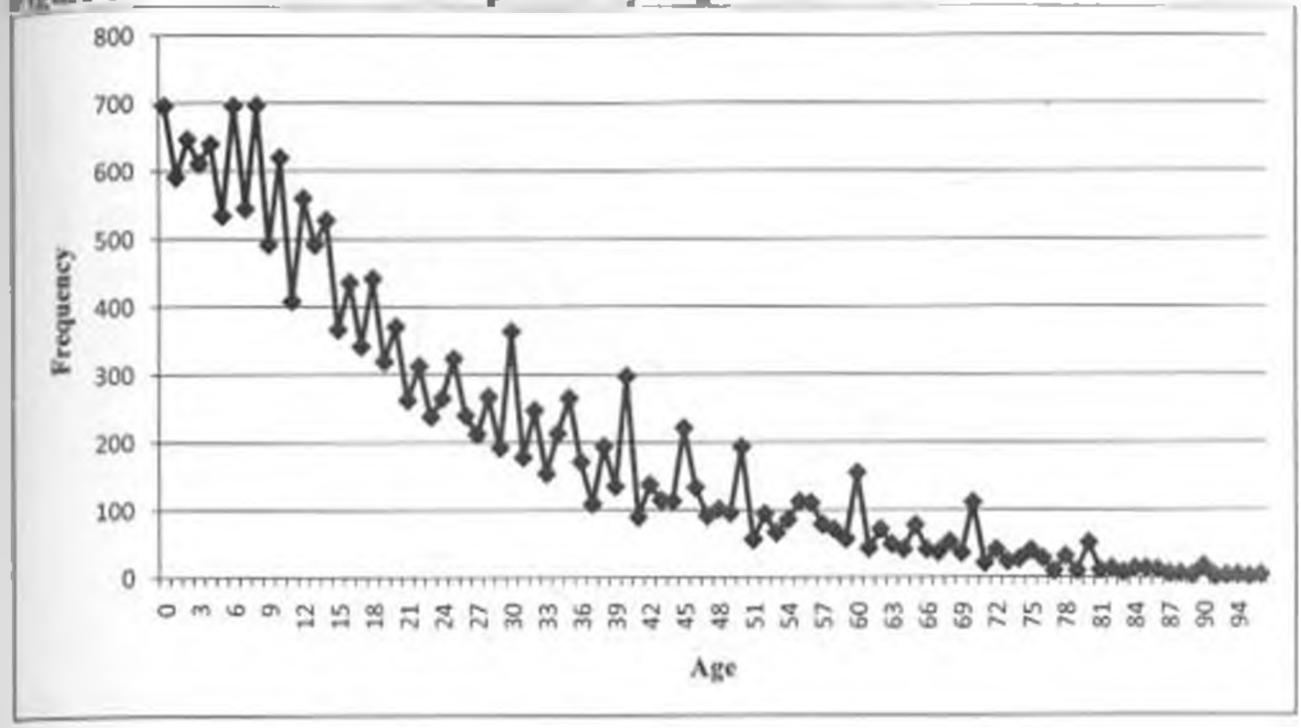
	Highest Educational Level							
	No education		Primary		Secondary		Higher	
	Male	Female	Male	Female	Male	Female	Male	Female
Age, month and year given and okay	**1.10**	3.80	**36.80**	37.00	**26.50**	20.90	**10.30**	7.50
Month and age given, year imputed	**0.00**	0.10	**0.20**	0.60	**0.10**	0.20	**0.00**	0.10
Year and age given, month imputed	**4.50**	9.60	**14.40**	11.90	**4.80**	2.40	**0.80**	0.10
Age given, year and month imputed	**0.40**	1.30	**0.10**	2.50	**0.00**	1.20	**0.00**	0.50

It can therefore be concluded that for persons without any education, the proportion of females was higher than that of males in each of the four age reporting categories. It appears that overall, for persons with at least some education, men outnumbered women whose age, month and year were given in such a way that no imputation had to be made. And across all education levels, there were more women compared to men whose age data was most incomplete, that is where age only was given and year and month of birth were imputed.

4.2: Digit Preference in Single Years

To examine the extent of digit preference for age in single years, a summary of responses on age and sex distribution is presented in graphical form. This is done separately for males and females. From the graphs below (Figures 4.1 and 4.2), evidence of heaping may be deduced from the sharp peaks, while digit avoidance is noted where the troughs are sharpest.

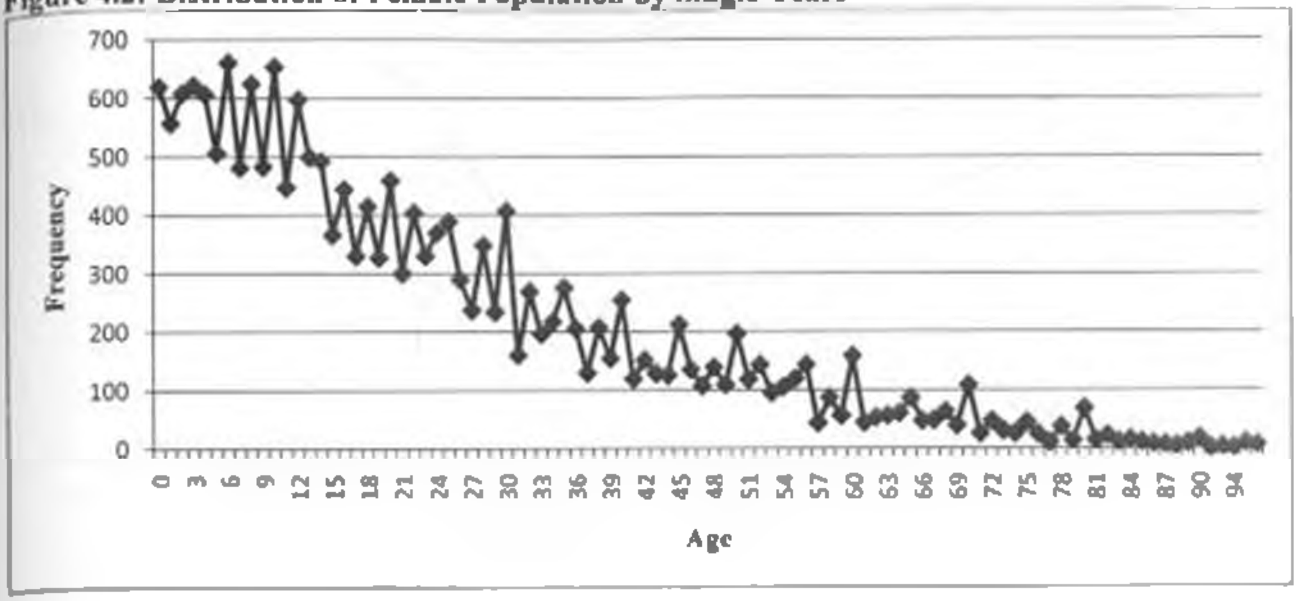
igure 4.1: Distribution of Male Population by Single Years



For the male population, heaping is observed in the ages 0, 2, 6, 8, 10, 12, 16, 18, 20, 22, 25, 28, 30, 32, 35, 38, 40, 42, 45, 50, 60, 65, 70, 80 years. The highest or major heaping is observed in the ages 0, 6, 8, 10, 30, 35, 40, 45, 50, 60, 65, 70 and 80. Hence it can be argued that from the general heaping, males covered in the 2008-09 KDHS had a preference for ages ending in 0, 1, 2, 5, 6 and 8. However, looking at the ages with the highest heaping, it emerges that ages with terminal digit 0 and 5 were the most preferred with the notable exception at ages 5 and 15 that were actually avoided altogether. Further, the results indicate that males avoided ages 1, 7, 9, 11, 17, 19, 21, 29, 31, 33, 37, 39, 41, 51, 61. It can therefore be argued that there was a tendency by the males to avoid age ending with terminal digits 1, 7 and 9.

A similar plot for the distribution of females yielded the graph shown in Figure 4.2. From this graph, it can be deduced that females on the other hand had their ages heaping on 0, 6, 8, 10, 12, 16, 18, 20, 22, 25, 28, 30, 32, 35, 38, 40, 45, 50, 56, 58, 60, 65, 70, 72, 75, 78 and 80 years. The highest heaping occurred at ages 0, 6, 8, 10, 12, 16, 18, 20, 22, 25, 28, 30, 32, 35, 38, 45, 50, 60, 65, 70 and 80 years.

Figure 4.2: Distribution of Female Population by Single Years



Like their male counterparts, it is apparent that the females covered in the 2008-09 KDHS generally preferred ages ending with 0, 2, 5, 6 and 8. But even as it emerges that the most preferred terminal digit was 0, 5 and 8, there was notably an exception for ages 5, 15, 55, 48 and 68 years. The females, it is observed, also avoided ages 1, 7, 9, 11, 14, 17, 19, 21, 27, 29, 31, 37, 53 and 57 years. This reflects avoidance for terminal digits 1, 7 and 9.

4.3: Myers' Index

The computational procedures (adapted from Pollard et. al (1974)) were done separately using the 2008-09 KDHS data for males and females. An age range 10-79 years over which the extent of digital preference is measured is divided into two partly overlapping sub-ranges 10-69 and 20-79. Population totals are then computed for ages ending in each of the ten terminal digits as shown in Table 4.4.

The population totals (columns 2 and 5) are the multiplied by coefficients in columns 3 and 6 to obtain products in columns 4 and 7 and whose sum is indeed the blended population in column 8 (see Table 4.5).

Table 4.4: Myers' Index for Males in the 2008-09 KDHS

Terminal Digit	Numbers at ages specified							Sum of Ages 10-69	Sum of Ages 20-79
	10-19	20-29	30-39	40-49	50-59	60-69	70-79		
0	619	370	363	297	193	155	111	1997	1489
1	408	262	178	89	56	43	21	1036	648
2	559	312	247	137	94	71	41	1420	902
3	491	238	154	114	66	49	23	1112	644
4	527	265	213	113	85	41	27	1244	744
5	366	323	265	221	112	77	40	1364	1038
6	435	240	171	133	110	41	28	1130	723
7	341	212	109	91	78	37	9	868	536
8	441	267	194	101	71	53	31	1127	717
9	319	192	135	95	57	36	9	834	524

For each of the digits, it is apparent as seen in Table 4.5 that the sum of the coefficients in columns 3 and 6 is 10. Multiplication by these coefficients is done to ensure that each digit has equal weight. In the case of an age distribution with no digital preference or avoidance, the blended population total for each digit would be approximately 10 percent of the total population for all digits. The percentage for each final digit and deviations from 10 percent are shown in columns 9 and 10 respectively. The total of absolute deviations gives Myers' Index.

From Table 4.5, for the males, Myers' Index results suggest preference for terminal digits 0 and 5, owing to the high values of the positive percent deviation. There is avoidance for ages with the terminal digits 1, 3, 7 and 9. Overall, the percentages suggest that the terminal digits in the order of preference are 0, 5, 8, 2, 6, 4, 9, 3, 7 and 1.

Table 4.5: Deriving the Blended Male Population and Percent Deviations

Terminal Digit	Age group 10-69			Age group 20-79			Blended Population (4)+(7)	% distribution	Deviation from 10%
	Sum	Coefficient	Product (2)x(3)	Sum	Coefficient	Product (5)x(6)			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	**1997**	1	1197	**1489**	9	13401	14598	14.55	**4.55**
1	**1036**	2	2072	**648**	8	5184	7256	7.23	**-2.77**
2	**1420**	3	4260	**902**	7	6314	10574	10.54	**0.54**
3	**1112**	4	4448	**644**	6	3864	8312	8.29	**-1.71**
4	**1244**	5	6220	**744**	5	3720	9940	9.91	**-0.09**
5	**1364**	6	8184	**1038**	4	4152	12336	12.30	**2.30**
6	**1130**	7	7910	**723**	3	2169	10079	10.05	**0.05**
7	**868**	8	6944	**536**	2	1072	8016	7.99	**-2.01**
8	**1127**	9	10143	**717**	1	717	10860	10.83	**0.83**
9	**834**	10	8340	**524**	0	0	8340	8.31	**-1.69**
Sum							**100,311**		**16.53***

Represents sum of absolute deviations

The procedures above are repeated for the female population, giving the results presented in Tables 4.6 and 4.7.

Table 4.6: Myers' Index for Females in the 2008-09 KDHS

Terminal Digit	Numbers at ages specified (Females)							Sum of Ages	Sum of Ages
	10--19	20-29	30-39	40-49	50-59	60-69	70-79	10--69	20-79
0	653	459	407	254	196	158	108	**2127**	**1582**
1	447	300	160	119	118	44	25	**1188**	**766**
2	597	404	269	151	143	53	47	**1617**	**1067**
3	499	330	196	128	95	56	30	**1304**	**835**
4	492	370	217	125	105	60	25	**1369**	**902**
5	366	389	275	212	119	86	44	**1447**	**1125**
6	444	289	205	135	143	47	24	**1263**	**843**
7	330	237	129	107	43	48	10	**894**	**574**
8	415	348	207	139	87	63	37	**1259**	**881**
9	327	234	154	109	56	39	13	**919**	**605**

Table 4.7: Deriving the Blended Female Population and Percent Deviations

Terminal Digit	Age group 10-69			Age group 20-79			Blended Population	% distribution	Deviation from 10%
	Sum	Coefficient	Product (2)x(3)	Sum	Coefficient	Product (5)x(6)	(4)+(7)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	**2127**	1	2127	**1582**	9	14238	16365	14.48	**4.48**
1	**1188**	2	2376	**766**	8	6128	8504	7.52	**-2.48**
2	**1617**	3	4851	**1067**	7	7469	12320	10.90	**0.90**
3	**1304**	4	5216	**835**	6	5010	10226	9.05	**-0.95**
4	**1369**	5	6845	**902**	5	4510	11355	10.05	**0.05**
5	**1447**	6	8682	**1125**	4	4500	13182	11.66	**1.66**
6	**1263**	7	8841	**843**	3	2529	11370	10.06	**0.06**
7	**894**	8	7152	**574**	2	1148	8300	7.34	**-2.66**
8	**1259**	9	11331	**881**	1	881	12212	10.80	**0.80**
9	**919**	10	9190	**605**	0	0	9190	8.13	**-1.87**
Sum							**113,024**		**15.91***

***Represents the sum of absolute deviations**

It is clear from Table 4.7 that females preferred stating ages with terminal digits 0 and 5 and avoided ages with terminal digits 1, 7, and 9. This is as evidenced by the very high values for the deviations, with a positive deviation implying preference while a negative value signifies avoidance of a digit. The percentage distribution in column 9 suggests that the terminal digits in the order of preference for females in the 2008-09 KDHS are 0, 5, 2, 8, 6, 4, 3, 9, 1 and 7.

4.4: National Age Ratios, Sex ratios and Joint Score

For the 2008-09 KDHS data, the computation for age ratios by sex and the sex ratios is done as shown in Table 4.8. This together with a computation for the deviations from 100 of respective age ratios and the sex ratio differences are utilized in the calculation for age and sex ratio scores and the UN Joint Accuracy Index.

Table 4.8: Computing Age and Sex Ratios

Males				Females				
Age Group	Number	Age Ratio	Deviations from 100	Number	Age Ratio	Deviations from 100	Sex Ratio	First differences
0-4	3180			3011			105.61	-1.87
5–9	2960	102.35	2.35	2754	96.65	-3.35	107.48	10.61
10–14	2604	107.12	7.12	2688	115.96	15.96	96.88	-4.19
15-19	1902	93.90	-6.10	1882	82.71	-17.29	101.06	23.39
20-24	1447	92.28	-7.72	1863	110.27	10.27	77.67	-4.76
25-29	1234	94.85	-5.15	1497	96.21	-3.79	82.43	-10.04
30-34	1155	109.58	9.58	1249	101.26	1.26	92.47	2.37
35-39	874	91.76	-8.24	970	95.76	-4.24	90.10	-6.42
40-44	750	99.01	-0.99	777	92.94	-7.06	96.53	5.21
45-49	641	103.05	3.05	702	97.91	-2.09	91.31	16.12
50-54	494	92.42	-7.58	657	114.26	14.26	75.19	-20.35
55-59	428	100.35	0.35	448	87.16	-12.84	95.54	-1.23
60-64	359	106.85	6.85	371	101.50	1.50	96.77	10.55
65-69	244	83.85	-16.15	283	93.40	-6.60	86.22	-8.67
70-74	223			235			94.89	
75+	273			335				
Total	**18768**			**19722**				
Absolute Total			**81.23**			**100.53**		**125.79**
Mean			**6.25**			**7.73**		**8.98**

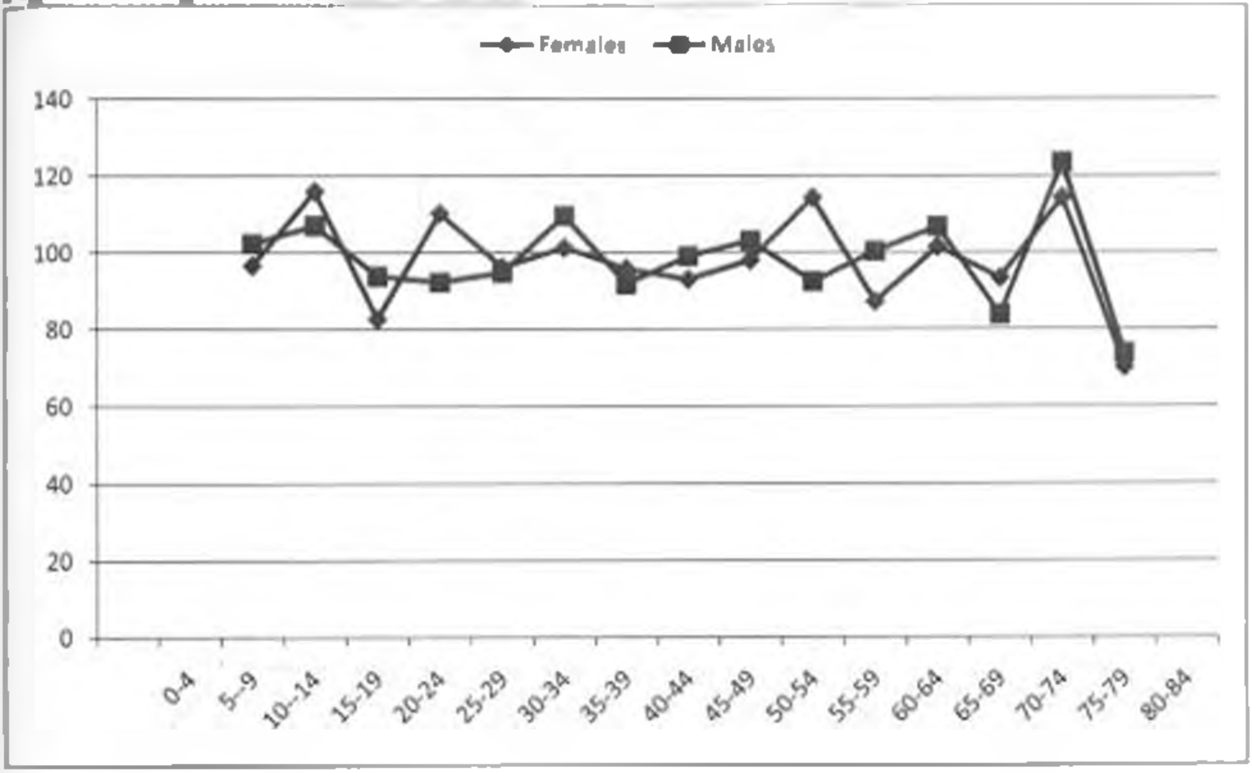
In the above Table 4.8, Age Ratio Score for males (ARSM) = 6.25

Age Ratio Score for females (ARSF) = 7.73

Sex Ratio Score (SRS) = 8.98

The age and sex ratios in Table 4.8 may be represented graphically as below;

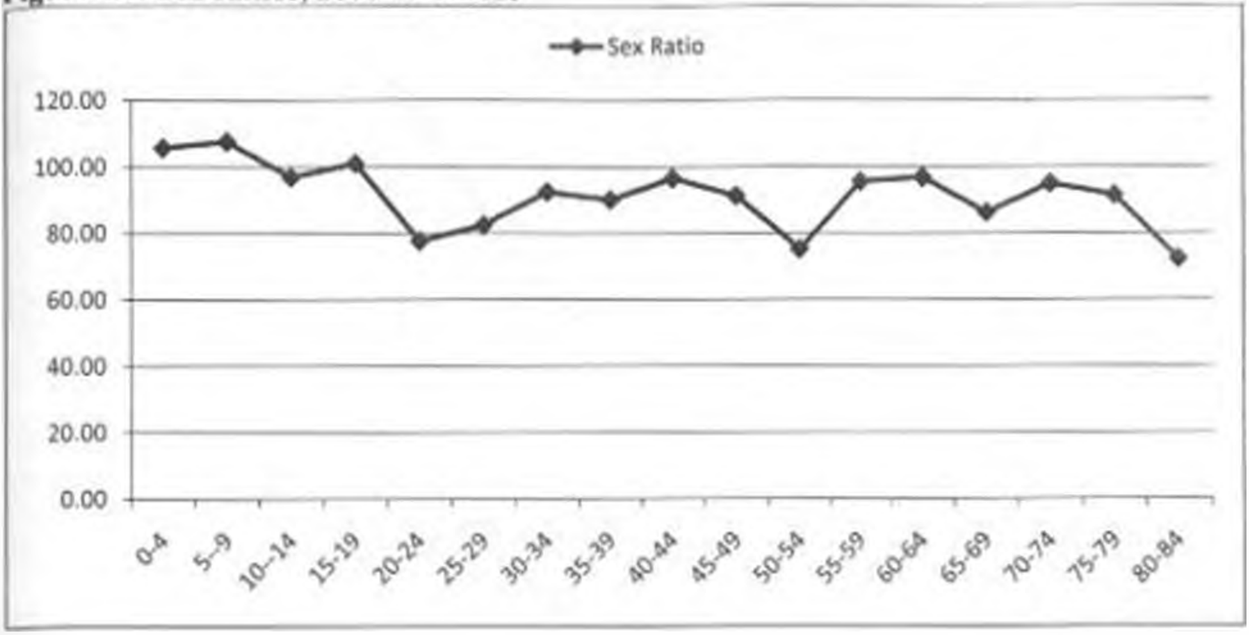
Figure 4.3: Age Ratios, 2008-09 KDHS



From the age ratio curve, it emerges that compared to the expected value of 100, the level of misreporting appears more pronounced for females compared to males in the 2008-09 KDHS. Males tended to markedly "overreport" (highly prefer or heap) their ages in the age-groups 30-34 and 70-74 years and "underreported" (avoided) their ages in the 65-69 year age group. The females on the other hand had their ages concentrated in the 10-14, 20-24, 50-54 and 70-74 age groups, while they avoided stating ages in the 15-19, 55-59 year age groups. Further, both males and females showed a very big dislike or avoidance of ages in the 75-79 year group.

The sex ratios observed in Table 4.8 are plotted against respective age groups and presented in Figure 4.4 below. As an analytical too, the larger the abrupt departure of this ratio from values close to 100, the larger the possibility of errors in the data.

Figure 4.4: Sex Ratios, 2008-09 KDHS



From the graph (Figure 4.4), it is observed that sex ratio at birth is reasonably normal (that is, slightly over 100), while other sex ratios suggest men only outnumber women at birth up to age nine only to be outnumbered all through the years after with the exception of near equal numbers for the age groups 10-14, 15-19 and 40-44. Errors in the 2008-09 KDHS data may be deemed to be in the age groups 20-24, 50-54 and 80-84 where the abrupt departure of sex ratio from values close to 100 is largest

The UN Joint Accuracy Index or the Joint Score is calculated as follows;

Joint Score = 3 x (sex ratio score) + (male and female age ratio scores)

= 3 x SRS + ARSM + ARSF

= (3 x 8.98) + (6.25 + 7.73) = 26.94 + 13.98 = 40.92

According to the UN suggestions (Arriaga, 1994), a Joint Score index value below 20 indicates that the data is accurate, while for a value between 20 and 40, the data is inaccurate, and for an index value above 40 the data is highly inaccurate. In our case, the Index value of 45.86 is too high suggesting highly inaccurate 2008-09 KDHS data.

4.5: Regional Age and Sex Ratios and the UN Joint Score

Regional population distribution, split into males and females is used as input in the AGESEX computer software programme. (See *Appendix 2* for population distribution at the national and regional levels). A summary of the output is presented in Table 4.9.

Table 4.9: Summary of Age and Sex Ratios and Age-Sex Accuracy Index by Province

	Region								
	Kenya	Nairobi	Central	Coast	Eastern	Nyanza	RVP*	Western	NEP*
Age Ratio Score for Males	6.25	15.45	11.04	8.46	13.76	5.46	8.08	11.50	27.99
Age Ratio Score for Females	7.73	19.60	8.44	14.52	9.18	10.70	11.60	7.93	24.00
Sex Ratio Score	8.98	28.26	11.69	18.09	11.58	13.80	12.62	10.12	29.03
Age-Sex Accuracy Index	**40.93**	**119.82**	**54.54**	**77.25**	**57.68**	**57.55**	**57.54**	**49.77**	**139.09**

**RVP- Rift Valley Province, NEP- North Eastern Province*

From Table 4.9, it can be concluded that regionwise, Western province's data appear better when compared with the rest of the regions, followed by Central, Rift Valley, Nyanza and Eastern provinces in that order. However, Coast, Nairobi and North Eastern provinces (in that order) have the worst age by sex data. As per the UN suggestion, it is evident that the age-sex data used in the 2008-09 KDHS is highly inaccurate, calling for strong smoothing.

4.6: Corrected Age Distributions

Table 4.10 presents results obtained from using the AGESMTH Software Programme/ spreadsheet to correct the male age data in the 2008-09 KDHS. It shows that the smoothed population for males using various methods is almost the same. However, it is observed that the variation from the reported data is characteristically different for the age groups 10-14 and 15-19. This implies age misreporting in these age categories by the males.

Table 4.10: Reported and Smoothed Population of Males by Age and Sex, Kenya

		Smoothed				
Age	Reported	Carrier Farrag	K.-King Newton	Arriaga	United Nations	Strong
Total, 0-79	18,612			18,612		18,612
Total, 10-69	12,132	12,132	12,132	12,132	12,130	12,132
0-4	3,180			3,266		3,311
5--9	2,960			2,874		2,829
10--14	2,604	2,486	2,469	2,473	2,554	2,364
15-19	1,902	2,020	2,037	2,033	1,939	1,972
20-24	1,447	1,474	1,495	1,471	1,453	1,583
25-29	1,234	1,207	1,186	1,210	1,248	1,310
30-34	1,155	1,098	1,095	1,095	1,112	1,075
35-39	874	931	934	934	905	902
40-44	750	764	765	761	744	760
45-49	641	627	626	630	630	634
50-54	494	509	510	507	507	513
55-59	428	413	412	415	425	420
60-64	359	339	338	337	348	336
65-69	244	264	265	266	264	264
70-74	223			201		199
75-79	117			139		141
80+	162					

Similar correction procedure was performed for the female population. Results presented in Table 4.11 show that the smoothed population for the females using the five methods is nearly the same. However, it is observed that the variation of the smoothed population from the reported is noticeably different for the age groups 10-14, 15-19 and 50-54 years. This implies age misreporting in these age categories for the females in 2008-09 KDHS.

Table 4.11: Reported and Smoothed Population of Females by Age and Sex, Kenya

Age		Smoothed				
	Reported	Carrier Farrag	K.-King Newton	Arriaga	United Nations	Strong
Total, 0-79	19,515			19,515		19,515
Total, 10-69	13,387	13,387	13,387	13,387	13,335	13,387
0-4	3,011			3,031		3,046
5-9	2,754			2,734		2,719
10-14	2,688	2,439	2,435	2,436	2,534	2,395
15-19	1,882	2,131	2,135	2,134	2,048	2,095
20-24	1,863	1,831	1,827	1,826	1,763	1,796
25-29	1,497	1,529	1,533	1,535	1,535	1,525
30-34	1,249	1,223	1,227	1,219	1,232	1,245
35-39	970	996	992	1,000	975	1,036
40-44	777	804	809	802	785	842
45-49	702	675	670	678	709	702
50-54	657	609	604	606	626	586
55-59	448	496	501	499	475	482
60-64	371	372	373	370	359	383
65-69	283	282	281	284	292	299
70-74	235			211		219
75-79	128			152		144
80+	226					

Consolidating the various indices for the 2008-09 KDHS data, a summary of the findings is made in Table 4.12 below.

Table 4.12: Summary of Indices Measuring the Accuracy of Data

		Smoothed				
Index	Reported	Carrier Farrag	K.-King Newton	Arriaga	United Nations	Strong
Sex ratio score	9.51	5.58	5.98	5.58	6.33	1.96
Male age ratio score	5.56	2.78	3.07	2.99	2.94	1.64
Female age ratio score	7.46	2.29	2.43	2.53	3.70	1.28
Accuracy index	41.56	21.79	23.44	22.25	25.62	8.79

It is observed that all the smoothing methods bring down the accuracy index by nearly half, with the exception of the strong method that is very accurate compared to the rest of the methods. The imputed accuracy indices based on the smoothed data from the various methods fall between 8.79 for the Strong Method to 25.62 for the United Nations method. This implies that the smoothed data is only of fairly good quality. The 2008-09 KDHS data may not be deemed as satisfactory reporting.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1: Summary

The study focuses on assessment of the quality of 2008-09 KDHS data. While the main objective was to carry out the assessment of the quality of data with a particular focus on the 2008-09 KDHS, specifically the study aimed at determining the extent of age heaping or digit preference for males and females in DHS, examining the age misreporting and determining the sex ratios through the different ages in the 2008-09 KDHS.

It emerged from the study that heaping of ages was rampant in the 2008-09 KDHS with a similar pattern for both males and females. Preference was observed in even age groups compared to the odd age groups for both sexes. Females were found to generally misreport their ages compared to the males in the 2008-09 KDHS. Males overreported their ages at 30-34 and 70-74 and underreported their ages in the 65-69 and 75-79 age groups. Females on the other hand overreported their ages in the 10-14, 20-24, 50-54 and 70-74 age groups, while they underreported in the 15-19, 55-59 and 75-79 age groups. In terms of numbers, females outweighed males all through except at birth.

Women were also characteristically found to trail males in literacy as evidenced by the analysis on education. For example, females without any education far outnumber males, while on the overall, males with at least secondary education and higher outweigh the females in the same category.

5.2: Conclusion

It is clear that the 2008-09 KDHS data is characterised by age misreporting errors. Age heaping is widespread with preferences for ages ending in terminal digits 0 and 5 for males and 0, 5 and 8 for females, but with exception for ages 5 and 15. Both males and females avoid ages ending in terminal digits 1, 7, and 9. The 0 and 5 preferences are in tandem with the trends in reporting ages across countries, giving credibility to the methodologies used in this project. Compared to results from the Kenya Population and Housing Censuses, the 2008-09 KDHS data may not be deemed as satisfactory reporting For example, an analysis of the 1979, 1989 and 1999 national censuses had reported accuracy indices of 28.1, 24.9 and 26.4 respectively. The 2008-09 KDHS data's Accuracy Index would attain values in this range only after smoothing.

The 2008-09 KDHS data is also characterised by systematic errors brought about by age overreporting and underreporting. This in turn is suggestive of the individuals concerned having their ages carried across age group boundaries, either to the next lower or higher age group, a character more pronounced for the female lot. The errors detected in the 2008-09 KDHS data are likely to have compromised its quality and the accuracy of the various demographic measures derived out of it. Further, differentials in education between males and females also influenced completeness of information or the way different sexes reported their ages.

5.3: Recommendations

It is recommended that the training of KDHS enumerators is intensified to reduce errors. Often, in estimation of ages, they base their figures on physical attributes, marital status among others, but it would be desirable they endeavour to use documentary proof when in doubt. Similarly, the masses should be educated through mass media on the need to report their ages as accurately as is possible. That women are the "bigger culprits" could be borne out of lack of education, culture or tradition that influences them to wish to conform to certain "accepted" ages. Other methodologies should be employed to assess KDHS data to confirm these findings and correct the errors thereby yielding better quality DHS data

Further studies should also be carried across various KDHS data and varied methods of analysis utilised to assess data quality as well as to adjust the KDHS data. Noting that the various indices computed are useful mainly in comparative analyses, it would be prudent to calculate the indices for various KDHS for this historical series would indicate whether the quality of the population age and sex reporting is improving or deteriorating.

REFERENCES

Arriaga, E. 1994. Population Analysis with Microcomputers; Presentation of Techniques. Vol 1. US Census Bureau. Washington.

___________. P. Johnson and E. Jamison. 1994. Population Analysis with Microcomputers Vol II. US Census Bureau. Washington.

Bairagi, R., K. Aziz, M. Chowdhury, and B. Edmonston. 1982. Age Misstatement for Young Children in Rural Bangladesh: In: Demography, Vol. 19, No. 4, pp. 447-458.

Bairagi, R., and A. Rahman. 1974. Age Reporting in Rural Bangladesh. Rural Demography 1(1):65-85.

Bogue, J. and E. Arriaga. 1993. Correction, Graduation, and Interpolation of population Data. In Readings in Population Research Methodology. UNFPA. Illinois.

Central Bureau of Statistics, 1996a. 1989 Population Census, Analytical Report: Volume III, The Population Dynamics of Kenya. Nairobi.

___________. 1996b. 1989 Kenya Population Census, A Popular Report. Nairobi.

___________. 2002. Kenya 1999 Population and Housing Census. The Popular Report. Nairobi.

___________. Undated. 1979 Population Census. Analytical Report, Volume II. Nairobi.

Ewbank, D. 1981. *Typical Patterns of Distortions in Reported Age Distributions*: In, *Age Misreporting and Age Selective Underenemuration Sources, Patterns and Consequences for Demographic Analysis*. Report No. 4. Committee on Population and Demography. Washington DC. National Academy Press, pp 46-69.

Garenne, M. 2003. Sex Ratios at Birth in African Populations: A Review of Survey Data. Detroit, Michigan: Wayne State University Press.

Ghos, A. 1967. Demographic Trends in West Bengali During 1901-1950. Census of India. Monograph No. 5. New Delhi, India. In, Bairagi R. et. al., 1982. Age Misstatement for Young Children in Rural Bangladesh: In: Demography, Vol. 19, No. 4, pp. 447-458.

Hill, M., S. Preston and I. Rosenwaike I. 2000. Age Reporting among White Americans Aged 85+: Results of a Record Linkage Study. Demography, Vol. 37, No. 2. pp. 175-186.

Kenya National Bureau of Statistics (KNBS) and ICF Macro, 2010. Kenya Demographic and Health Survey 2008-09. Calverton, Maryland: KNBS and ICF Macro.

Kpedekpo, G., 1982. Essentials of Demographic Analysis for Africa. London

Magadi, M. 1990. *Estimation of Age Distributions, Census Coverage and Death Registration Completeness in Kenya.* Unpublished Masters Thesis. PSRI

Myers, R., 1951. Errors and Bias in the Reporting of Ages in Census Data.In Jaffe, A.J. Handbook of Statistical Methods for Demographers. Washington. pp. 395-414.

Oucho, J., 1985. Some Demographic Measures of Rural Migrants in Kenya Based on Survey data. Genus, Vol 41, No 1/2. Pp 77-95. Universita degli Studi di Roma "La Sapienza".

Pardeshi, G. (2010). *Age Heaping and Accuracy of Age Data Collected During a Community Survey in the Yavatmal District, Maharashtra,* Indian J Community Med. Vol 35(3): 391–395.

Pollard, A., F. Yusuf and G. Pollard, 1974. Demographic Techniques. Sydney.

Pullum, T. 2006. *An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003.* Methodological Reports No. 5. Calverton, Maryland: Macro International Inc.

Retherford, R. and G. Mirza, 1982. Evidence of Age Exaggeration in Demographic Estimates for Pakistan. Population Studies, Vol 36. London.

Rosenwaike, I. and B. Logue. 1983. Accuracy of Death Certificate Ages for the Extreme Aged. Demography, Vol. 20, No. 4. pp 569-585

Shryock, H. and J. Siegel. 1976. *The Methods and Materials of Demography.* San Diego, California. Academic Press Inc.

Siegel, J. and D. Swanson (Eds). 2004. *The Methods and Materials of Demography.* Elsevier Science (USA).

UNFPA. 1993. Readings in Population Research Methodology. Social Development Centre. Chicago, Illinois.

United Nations, 1955. Methods of Appraisal of Quality of Basic Data for Population Estimates, Manual II. Population Studies No. 23.

Wamai, A., 2004. *Detection and Correction of Errors: A Case Study of 1989 and 1999 Census Data*. Unpublished Masters Thesis. PSRI.

You, P. 1959. Errors in Age Reporting in Statistically Underdeveloped Countries. Population Studies 12(2):164-182.

Yusuf, F. 1967. On the Extent of Digital Preference in Reporting of Ages in Pakistan. The Pakistan Development Review: pp 519-532

APPENDIX 1: 2008-09 KDHS Data in Single Years, Kenya

Age	Males	Females	Age	Males	Females	Age	Males	Females	Age	Males	Females	Age	Males	Females
0	695	619												
1	590	557	21	262	300	41	89	119	61	43	44	81	10	14
2	646	608	22	312	404	42	137	151	62	71	53	82	13	22
3	610	621	23	238	330	43	114	128	63	49	56	83	7	11
4	639	606	24	265	370	44	113	125	64	41	60	84	13	15
5	534	506	25	323	389	45	221	212	65	77	86	85	13	11
6	695	660	26	240	289	46	133	135	66	41	47	86	11	8
7	544	481	27	212	237	47	91	107	67	37	48	87	5	7
8	696	624	28	267	348	48	101	139	68	53	63	88	5	5
9	491	483	29	192	234	49	95	109	69	36	39	89	2	9
10	619	653	30	363	407	50	193	196	70	111	108	90	16	17
11	408	447	31	178	160	51	56	118	71	21	25	91	1	1
12	559	597	32	247	269	52	94	143	72	41	47	92	2	3
13	491	499	33	154	196	53	66	95	73	23	30	94	3	1
14	527	492	34	213	217	54	85	105	74	27	25	95	1	9
15	366	366	35	265	275	55	112	119	75	40	44	96+	3	6
16	435	444	36	171	205	56	110	143	76	28	24	DK	1	9
17	341	330	37	109	129	57	78	43	77	9	10	Missing	5	10
18	441	415	38	194	207	58	71	87	78	31	37			
19	319	327	39	135	154	59	57	56	79	9	13			
20	370	459	40	297	254	60	155	158	80	51	68			

APPENDIX 2: 2008-09 KDHS Data in Five-Year Groups, Kenya

	Region																	
	Nairobi		Central		Coast		Eastern		Nyanza		Rift Valley		Western		Northeastern		Total	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	P
0-4	199	205	251	266	457	418	392	387	558	505	556	539	436	430	331	261	3180	3011
5-9	147	145	286	249	382	381	435	427	450	410	538	518	385	377	337	247	2960	2754
10-14	114	112	260	268	297	350	370	366	429	406	435	495	364	374	335	317	2604	2688
15-19	88	137	182	203	226	245	285	262	379	314	280	312	274	259	188	150	1902	1882
20-24	167	252	166	187	198	263	208	196	226	330	224	288	173	227	85	120	1447	1863
25-29	194	236	138	170	155	193	134	176	188	247	214	198	147	165	64	112	1234	1497
30-34	193	145	130	145	167	172	143	194	153	162	162	199	141	151	66	81	1155	1249
35-39	129	115	118	112	126	133	95	141	118	136	139	136	93	113	56	84	874	978
40-44	91	75	89	128	99	86	97	107	99	100	119	121	94	107	62	53	750	777
45-49	81	62	79	102	81	89	103	107	82	107	108	106	66	84	41	45	641	782
50-54	58	68	55	86	61	103	71	95	65	92	78	83	56	75	58	56	494	657
55-59	60	32	59	69	68	59	59	75	56	65	57	51	43	62	26	35	428	448
60-64	41	25	45	54	50	48	60	71	45	48	46	46	43	49	29	30	359	371
65-69	17	15	44	52	35	30	40	58	34	37	30	35	32	39	12	17	244	283
70-74	9	10	31	38	24	26	43	41	28	40	23	31	28	27	37	22	223	235
75-79	4	12	19	13	13	15	25	23	17	24	14	18	18	18	7	5	117	128
80-84	4	8	12	24	10	16	16	26	14	10	16	17	13	15	9	14	94	130
85+	1	7	16	13	3	4	15	20	9	5	4	8	6	10	8	10	62	77
Total	1597	1661	1980	2179	2452	2631	2591	2772	2950	3038	3043	3200	2412	2582	1743	1659	18768	19722