DETECTION AND CORRECTION OF AGE ERRORS.

A CASE STUDY OF

1989 AND 1999 KENYAN CENSUS DATA

BY

WAMAI AGNES WAMUYU

Q56/7422/02

This Project is submitted in Partial Fulfillment of the requirement for the award of the degree of Master of Science in Population Studies at the Population Studies and Research Institute, University of Nairobi.

November 2004

DECLARATION

This project is my original work and has not been presented in part or whole, for a degree in this or any other University.

Candidate: Wamai Agnes Wamuyu

Signature	8.12.04 Date

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

Dr. Agwanda

Signature	26/12/2004 Date

Dr. Kimani Murungaru

Signature	17/12/04 Date

DEDICATION

This project is dedicated to my husband Marclus and our son Dominic for their encouragement, dedication, moral support and patience for the entire period I have been working on this report.

ACKNOWLEDGEMENT

Special thanks go to the University of Nairobi for granting me scholarship which enabled me to undertake Masters of Science in Population Studies at Population Studies and Research Institute, University Of Nairobi.

I am greatly indebted to my supervisors, Dr. Agwanda and Dr. Kimani for their inspirational supervision which made it possible for me to complete this project. To them I express my sincere appreciation for their untiring guidance throughout the course of the project.

I also extend my gratitude to all Population Studies and Research Institute staff for their academic nourishments and for their contribution that made this project successful. I also thank all my colleagues at Population Studies and Research Institute for their positive insights, criticism and giving me company during these moments.

To all these parties may God bless you abundantly.

ABSTRACT

This study focused on investigating the quality of age data in 1989 and 1999 Kenyan population censuses through detection and correction of age misreporting. The study had three objectives namely; to identify the nature and extent of age errors, make appropriate adjustment for age misreporting and compare the corrected age data with that of the Central Bureau of Statistic (CBS). The methods which were used were age ratio technique for the detection of age heaping, Feeney method for the correction of age heaping in multiples of five, Demeny-Shorter method for the detection and correction of systematic errors and a combination of Feeney and Demeny-Shorter methods for the correction of age heaping as well as the systematic errors. On the other hand, linear interpolation was used for the estimation of mortality level. The results that were obtained were then compared with the CBS data so as to asses the appropriate method that can be used for the adjustment of Kenyan population census age data.

The data used were the 1989 and 1999 Kenyan population censuses, specifically age data in single years and five years age groups distribution by sex was used. The results obtained by age ratio technique indicated that there was age heaping in the two censuses .Age heaping was almost similar in 1989 and 1999 censuses in that there was preference of digit 0 and 5 with an exception of age 5 in1999 population census, the degree of age heaping for both sexes was observed to increase with age but it was more pronounced in even age groups such as 10-14, 20-24 as compared to odd age groups such as 5-9, 15-19 and also greater in females as compared to males. In addition, age heaping decreased while age avoidance increased in 1999. Overall age misreporting was more pronounced in females as compared to males. The Results obtained by Feeney method indicated that there was systematic age misreporting and there was an increase of age under reporting in 1999 census as compared to 1989 census. From Demeny-Shorter method it was found that males had a high tendency to over report their ages while the females under reported theirs. Over reporting of age in males occurred at older ages but

in females this occurred among the teenagers and very old people 70+. In females under reporting was common in ages 30+ while the same occurred among the teenagers in males.

A comparison of the results in this study and the data corrected by CBS suggest that when Demeny-Shorter and Feeney methods were applied individually and then jointly better results were obtained in the later as compared to the former. However discrepancies were found between the two data and this is attributed to; errors during the data entry which lead to wrong figures particularly for the total, the nature of the data that was used because in case of the CBS the data that was used had already been cleaned from irregularities but raw data was used in this study. However the methods used for the analysis contributed a lot for the discrepancies. The CBS used the Arriaga light smoothing technique which just smoothed data to reduce the effect of digit preferences and random errors. This was insufficient since the distortions that result from systematic over or under reporting was not put into consideration. When the projection of the two censuses data was carried out using the population data already adjusted by the combination of both Feeney and Demeny Shorter methods, the census coverage was found to be 96.44 and 99.7 in 1989 and 1999 censuses as compared to 94.0 and 97.3 obtained by CBS for the similar censuses respectively. In this study the adjustment of age data was carried for both random and systematic errors. Feeney and Demeny-Shorter methods were used and the main reason was that age misreporting in Kenyan census data is not only caused by random errors but also by systematic errors.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

GENERAL INTRODUCTION

Information on age is fundamental in demographic analysis since the study of fertility, mortality, migration and population projection rely on accurate age reporting. Age is also a central variable in demographic estimation in particular mortality and fertility depends on the availability and quality of age data Ewbank (1981), Murkerjee (1976) and Fosu (2001). Data on age is usually gathered during census or survey CBS (1994, 2000) Narasimhan, Rutherford, Mishra Vinod, Anold Fred and Koy (1997). Unfortunately, it is difficult to get accurate data on age especially in developing countries since in such societies accurate data on age of individuals is not emphasized and the policy that puts very little efforts in registering the vital events Bhat, (1990).

There are basically two types of errors in age data namely; random and systematic. The major sources of these errors are; the respondent who willfully misreports his or her age or gives an approximate if the true age is not known, the interviewer estimation of the age of a respondent from respondent physiological features such as height, color of hair, personal characteristics such as marital status and number of children born, preference for digits such as 0 and 5 and subjective biases to either over or underreport age, Bhat, (1990). Other sources of age misreporting are cultural beliefs and low level of education among the individuals, (Rutherford Robert, (2001), Dudley, (2000) and Shryock and Siegel (1976). In order to have accurate age data demographers and statisticians have developed graduating techniques (United Nation, 1955 and CBS, 2000).

1.1 Problem statement

Since the age data is subjected to errors and can be misleading and unreliable, several methods for adjustment of age data have been developed. Whipples index, Myers index, and Bachi index are several indices that are frequently used for checking the quality of age data by detecting digital preferences in age reporting (United Nations, (1955) Shryock and Siegel, (1976) and Central Bureau of Statistics, (2000). Unfortunately, these indices in some cases shows a fairly large amount of age misstatement which has no influence on age grouped data and do not take into account that such digit preferences is usually connected to other causes of inaccuracy in age statement such as preference of a particular age, (United Nation 1990). In addition to these indices the United Nation Accuracy Index is used in grouped data to measure net age misreporting. It is also not a reliable measure since it does not measure net under enumeration by age (United Nation, (1955).This implies that all these measures can only be accepted as a fair measure of the general reliability of age data UN (1955). In addition, the methods do detect only age heaping in terms of digit preference as well as smoothing age to reduce the effect of digit preference and random errors in age data. Since these are not the only errors in age misreporting, it is important to minimize the distortion on age that results from systematic errors as over and under enumeration of age which causes difficulties in the estimation of fertility and mortality. The application of some of these methods on the 1989 and 1999 population censuses has not been undertaken. This study therefore focuses on using these techniques to assess a particular preference of a certain age, and correct both the effect of random and systematic errors on census age data. The age ratio, Feeney (1979) and Demeny-Shorter (1968) methods will be used. The study will answer the following two questions:

1. What is the nature and extent of age errors in the 1989 and 1999 Kenyan population censuses; and how can those age errors be corrected?

2. What conclusions can be made by comparing the smoothed age data in this study with Central Bureau of Statistic smoothed data?

1.2 General Objective

The study will assess the extent of age misreporting and correct the errors in 1989 and 1999 Kenyan population censuses and compare the results with the Central Bureau of Statistics adjusted data.

1.3 Specific Objectives

1. Identify the nature and extent of age errors in 1989 and 1999 Kenyan population censuses.
2. Make appropriate adjustment for age misreporting errors in 1989 and 1999 Kenyan population censuses.
3. Compare the corrected age data with the Central Bureau of Statistic (CBS) corrected data.

1.4 Justification/ Rationale of the Study

Age is an important variable in demographic analysis and especially in estimation of fertility and mortality. If age data is of poor quality the accuracy of population estimates would certainly be reduced, perhaps significantly. Hence in order to get a better picture of fertility and mortality, it is necessary therefore to first evaluate the quality of age data before undertaking any analytical work.

1.5 Scope and Limitation of the Study

The analysis was carried out at the national level since the techniques that are used are based on the assumption that the population under study is closed to migration. The effect of immigration and emigration at national level can be ignored since it is insignificant at national level migration. The data from the 1989 and 1999 Kenyan population censuses was utilized in the study. The specific data needed is age in single years and five years age groups distribution, classified by sex. The study will focus on detection and correction of age misreporting .The methods to be used are age ratio for the detection of age heaping while the Feeney method will be used for the correction of age heaping in multiples of five and Demeny- Shorter methods will be used for detection and correction of under or over enumeration of age. Demeny-Shorter method has a limitation in that its assumptions that the model mortality level of the population under study is known, total size of the population was enumerated correctly and that the degree of coverage of the two censuses was reasonably close would have affected the accuracy of the results since different results would give different results. The census coverage obtained after projecting the population using the adjusted population in this study was 96.44 and 99.7 as compared with 94 and 97.3 obtained by CBS in 1989 and 1999 population censuses respectively. Therefore it was necessary to adjust for under coverage but this was not carried out in this study since the time allocated for the study was limited and the results that were obtained were efficient for comparison and making appropriate suggestion on whether the methods used in this study were better as compared to what the CBS is using.

1.6 Literature Review

Data on age in developing countries are subject to errors. There are basically two types of errors in the reporting of age namely random and systematic errors. A common error caused by the former in age reporting is the tendency of rounding the ages to the nearest figure in 0 and 5 or to a lesser extent in even numbers or preference of a certain age United Nation (1955). The latter occurs when individuals of a given age have been missed or counted more than once during enumeration. In view of these considerations, the sources of age errors are age heaping, over-enumeration or under-enumeration, cultural beliefs and low level of education Beckett et al, (1999), Shryock and Siegel (1976) United Nation (1955). The literature reviewed focuses on the methods for the detection and correction of age errors.

1.6.1 Detection of age heaping

The common measure of age heaping are age ratios, Myers index, whipples index, Bachi index, Carrier and Farrag index which provides an overall summary of preferences for, or avoidance of ages ending in specific digits in single years, United Nation (1955) and Shryock and Siegel (1976). In addition, United Nation Accuracy Index is used to detect and measure the overall extent of age misreporting by examining the age and sex ratios of the population. However it is not very exact and should be regarded as an order of magnitude rather than a precise measurement. The index is primarily a measure of net age misreporting but does not measure net under enumeration by age United Nation (1955). However, such digit preference is usually connected with other sources of inaccuracy in age statement and the indexes can only be accepted as a fair measure of the general reliability of the age distribution United Nation (1990). According to United Nation (1955) the age ratio has an advantage over these other indices because it is affected by differential omissions of persons in various age groups from the census count and by the tendency of age misstatement as well as by digital preferences and is therefore a more truly reflection of the general accuracy of age heaping statistics.

1.6.2 Correction of Age Errors

Due to the various errors in age data it is important to adjust the data in order to make it suitable for various demographic analyses. Various techniques have been developed for the adjustment of the data.

Feeney (1979) discovered that age distribution obtained from censuses and surveys in most developing countries exhibits heaping on ages that are multiples of five. He suggested a technique for redistributing the excess numbers of persons at these ages to the adjacent ages. His method involves redistribution of excess persons from a given multiple of five to the eight adjacent ages, the four years on either side of the multiple of five. It was based on two assumptions that ; the adjusted numbers in the surrounding ages are proportional to the original numbers and that the adjusted numbers of the lower four ages ,the central age, and the upper four ages form a linear progression. In his preliminary analysis of Indonesian data Feeney claimed that his procedure was remarkably effective. Instead his method had a weakness in that it was complicated and involved forcing the age distribution into a series of straight-line segments centered at multiples of five.

Saxena Gogte (1985) investigated the possibility that a simpler procedure for correcting age heaping on multiples of five could yield equally effective results as the Feeney's (1979) procedure. Nine point moving average was selected for comparison. The nine point moving average was applied once and then twice to census age distributions for India (1971), Tanganyika (1967) and Indonesia (1971), and the results were compared with those obtained by applying Feeney's method. It was indicated that all the three methods performed equally well for the three countries considered since the results were very close. They concluded that when the nine point moving average was applied once it was more preferable because it was simple to apply. The method has an advantage in that it is easier to compute, can be used for the same purpose and gives almost similar results as Feeney method. One of the disadvantages as compared with Feeney method is that Saxena and Gogte method rationale is not specific.

Demeny and Shorter (1968) developed his method by putting earlier Carrier and Farrag (1959), ideas into a more general form. They devised a new technique to separate the true irregularities in age distributions from reporting errors and applied it to six census population. This was based on three assumptions namely; the pattern of misreporting is systematic and therefore repeats itself from one census to the next; that the mortality schedule of the population can be represented by an appropriately selected model life table and that the total size of the population is enumerated correctly. The method thus required two census data simultaneously and a life table that represents the average mortality situation during the intercensal period. Demeny and Shorter applied the method to the pairs of Turkish censuses for 1935-40 and 1955-60. Their findings were that there was a relative deficit at ages 0-4 which indicated that there was exaggeration of age for young children. This contributed to an excess of 5-9 for both male and females. On the other hand while those women at puberty and those who married when at younger ages exaggerated their ages. This brought about deficit in age groups 10-14 and 15-19 and over-reporting in ages 25-29 and 30-34. In addition, age misreporting errors in males were lower as compared with that for the females. This is allowed due to the higher literacy level among males because age was very important in determining the time for military service. Males that were transferred downward across the boundaries of ages 15 and 20 could have been a tendency to assign lower than true ages to men who were late in maturing physically or in military service. The major advantage of the method is that it is easy to compute and can also be applied in many developing countries specifically in Kenya where there is need to correct for systematic errors. The D-S method uses two census populations simultaneously, a life table that represents the average mortality situation in the intercensal period and a data grouped in five years age groups distribution by sex.

Gupta (1975) developed a general method in order to eliminate the limitation of Demeny-Shorter method .The base of this method was that the Demeny-Shorter method could only give excellent results if the two censuses are identical and reasonably close. However if the disparity between them becomes larger, it becomes increasingly difficult for the method to satisfy simultaneously the three underlying assumptions .The method was based on three assumptions whereby two are similar to Demeny-Shorter assumptions and the third assumption is a modification of Demeny-Shorter assumption that the error in any age group bears a constant ratio to the reported population in the same age group. Gupta assumptions were; the population has experienced mortality according to suitably chosen model mortality, the total size of the population was enumerated correctly and that person's age is reported either in the correct age group or in the next higher or lower age group. The findings were that Gupta method had several advantages which are, the method is based on a set of internally consistent assumptions whereas in general the Demeny –Shorter method fails to satisfy all its assumptions. According to Gupta, the method can be applied to any population, whereas the Demeny-Shorter method is more suitable for population with unchanging age structures. The correction factors in the Demeny-Shorter method cannot be interpreted in terms of transfer of population to and from adjacent age groups; lastly, the method is quite flexible and does not require prior knowledge of the nature of age biases. However the major advantage of Demeny-Shorter method over Gupta method is its computational simplicity and yield almost similar results when applied in 1955 and 1960 female census data in Turkey. However the final results were very close and this could have been attributed to the fact that the true age structures in the two censuses are the same Gupta (1975). This implies that Demeny-Shorter method satisfies its first underlying assumption that the error in any age group bears a constant ratio to the reported population in the same age group when the age structures in the two censuses are identical or reasonably close.

Ntozi (1978) developed a method from Brass (1969) which had suggested an extension which uses age data from three censuses. The method involved an averaging of two successive census technique and was only applicable when error changes are systematic and small. It was a more flexible method which takes accounts of changes in the age reporting and coverage errors and which conveniently uses information in three censuses held at equal intervals. Unlike Demeny-Shorter method, which examined coverage errors, the three-census method estimated the likely changes in census coverage and different patterns of age errors in three successive censuses. The method was applied to the census data in 1955 and 1960 female census data in Turkey. The method is based on three assumptions two of which are similar to D-S assumptions and the last one is a modification of Demeny-Shorter's assumptions which states that the error in any age group bears a constant ratio to the reported population in the same age group. Therefore the underlying assumptions are; the population is assumed to have experienced mortality according to a suitably chosen model mortality schedule; that the total size of the population was enumerated correctly and that there was regularity of change in error or that the error changes was systematic and small in the three censuses. The third assumption replaced the Demeny-Shorter's assumption that the sizes of age errors are constant at successive censuses. The method was applied in 1955 and 1960 censuses of the Turkish female population and both Demeny –Shorter and three censuses method obtained comparison of correction factors. The Demeny –Shorter technique in most age groups are in the same direction as one of the two corresponding correction factors from the three-census method. Therefore, the patterns of age error disclosed by the two approaches are similar. The three census method verifies that the age errors in 1960 were much the same as those in 1955. The Demeny-Shorter method assumptions are therefore, reasonably accurate with regard to age errors. Replacing the Demeny-Shorter assumptions of constant error by a regular trend of error at successive censuses, the three censuses method is likely to be more realistic. The Demeny Shorter method is easy to compute and yield almost the same results as the three-census method and this verifies that Demeny-Shorter method assumption that the sizes of age errors are constant and repeats from census to census. The conclusion that Ntozi made was that the Demeny-Shorter assumptions are, therefore, reasonably accurate with regard to age errors. This makes the Demeny-Shorter method superior to the other two methods.

Bhat (1990), method of estimating probabilities of age misstatement used two methods namely disproportional adjustment algorithm and guess matrix in the study on correction of systematic errors. He emphasized on the need to have information on gross age errors by focusing on how to get gross age errors through matching stated ages in a census or survey with dates of birth from vital registration records. Unfortunately this method is not applicable in developing countries and specifically in Kenya since data on vital registration is incomplete.

In Kenya little has been done in order to detect digit preference and adjustment of age misreporting. K'oyugi (1982) applied Whipple's index and Myers index to Siaya district using data for 1969 and 1979 census. Rono (1982) estimated the quality of age reporting using the 1969 and 1979 census. The United

Nation Accuracy Index (UN 1955), Myers index and Whipple's indices were used. In both studies it revealed that there was age misreporting and specifically age heaping in Kenyan census data.

Magadi (1990), in her study on estimation of age distribution using 1969 and 1979 censuses detected age heaping using age ratios and age specific growth rates, Saxena and Gogte method for the correction of age heaping in multiples of five and Demeny-Shorter method for correcting systematic age errors. From the results it was indicated that the 1969 and 1979 census age data were encountered by digital preferences and systematic errors.

1.6.5 Summary of Literature Review

The literature review above, suggest how the quality of age data in censuses from developing countries and Kenya in particular suffer from age misreporting errors. Digit preference occurs when most of the people reports their ages with a similar digit which they likes or the enumerator estimates the age of the respondent and this results in age heaping. Age misreporting also occurs when the respondent is enumerated more than once or omitted completely thus leading to systematic errors. There are various common indices for detection of age heaping such as Whipples index, Myers index, Bachi index, and Carrier and Farrag which gives an overall summary of preferences and avoidance of ages ending in specific digits. United Nation joint score index is also useful for the detection of age heaping. Unfortunately it is primarily a measure of net age misreporting but does not measure net under enumeration by age. In general all these are indices for the detection of digital preferences and random errors in age data and only give a fair indication of age misreporting. As a result other techniques that can be used for detection of age heaping is age ratio. The method is preferred over the other methods and will be applied in this study since it is affected by differential omissions of persons in various age groups from the census count and by the tendency of age misstatement as well as by digital preferences and is therefore a more truly reflection of the general accuracy of age data statistics. The methods that are used for adjustment of age misreporting are Feeney method and Saxena and Gogte methods that are used for the correction of age heaping in multiples of five while Demeny –Shorter method, Gupta, general method and Ntozi, three census methods are used for the correction of systematic errors. For the purpose of this study, Feeney method would be used in the correction of age heaping in multiples of five since its rationale base is more specific as compared to Saxena and Gogte method. In addition, Demeny-Shorter method will be used for the correction of systematic errors because the other two methods that is the general and the three census methods have verified that the Demeny-Shorter method has satisfied its three underlying assumptions

CHAPTER TWO

DATA SOURCE AND METHODS OF ANALYSIS

This section describes the data utilized in the study and the methods of data analysis.

2.1 Data Source

The data used in this study was obtained from 1989 and 1999 Kenyan Population Censuses concluded in 1989 and 1999 respectively. In particular, age data in single years and five years age groups distributed by sex was used. The question that was asked during the enumeration in order to capture age data was "How old is ....?"and age was recorded in completed years using two digits and if it was under one year it was recorded as "00".This data was collected during the night of 24th and 25th Aug in 1989 and 1999 censuses, CBS (1994 and 2001).

2.2 Methods of Analysis

Understanding of the sources of age misreporting is important for improving age reporting and adjusting for the effects of misreporting during data analysis. Age misreporting can be as a result of digital preferences which lead to heaping in multiples of five or due to systematic errors that leads to over – enumeration and under – enumeration of ages across age groups.

The main focus of this section is on three methods, one method for the detection of age heaping, the second method for the correction of age heaping in multiples of five and the third method for the correction of systematic errors in that order. Age ratio technique was used to detect age heaping. Data in single years from age 5 to age 74 was used. Feeney method was used for the correction of age heaping in multiples of five and utilizes data in single years and five years age groups distributed by sex while Demeny-Shorter method requires data in five years age groups distributed by sex and a suitably chosen model mortality schedule. The method detects and corrects for systematic age errors. Age ratio technique and Demeny-Shorter methods were useful in achieving the first objective since they will help in identifying the nature and extent of age errors in 1989 and 1999 Kenyan population censuses. Feeney (1979) and Demeny and Shorter method are used for making appropriate adjustment for age misreporting in the two censuses. However, the analysis was carried out in spread sheets.

2.3 Method of detecting age heaping

2.3.1 Age Ratios

Age ratios are indices used for detecting possible age misreporting in population where fertility has not fluctuated greatly during the past and where international migration has not been significant. Under these conditions, age ratios should be similar and close to one. The method is preferred for analysis since it measures age heaping as a result of preference of a certain age unlike the other indices which concentrates on age heaping due to digital preferences. Age data in single years staring from age 5 was used. Age ratios are calculated by dividing the population of a specific age by the average population of two adjacent five years age groups. The numerator is the number enumerated in the given single age in question and the denominator is the mean of the numbers enumerated in the immediately lower and immediately higher age intervals that is (i – 1) th and (i + 1) th age intervals. The higher the fluctuations of the age ratios from values close to 1 the higher the possibility of errors in the data. Age 5 is the lowest age in which an age ratio can be computed. This is obtained by having the number of persons aged five as the numerator divided by one tenth of the sum reporting ages,0,1,2,3,4,5,6,7,8,9 and 10. An age ratio less than one indicates avoidance for that particular age, whereas an age ratio greater than one indicates preference for the age in question while an age ratio equal to one indicates accurate age reporting. The following formula was used in the calculation of age ratios. Computations of age ratios consist of three steps. Suppose the age ratio R(X) for age x is given by the following formula;

$$R(x)=\frac{M(x)}{\left[\left(\frac{1}{10}\right)*\left(Mx_{-5}+Mx_{-4}+Mx_{-3}+Mx_{-2}+Mx_{-1}+Mx_{+1}+Mx_{+2}+Mx_{+3}+Mx_{+4}+Mx_{+5}\right)\right]}$$

Where;

R(x) = Age ratio for age x in question for x = 5, 6, 7, 8, 9… A – 6. I.e. x=5,

M(x) = Recorded population of a single year age x in question.

M x-5….M x-1 = Recorded population for the previous five ages in single years immediately smaller than the age in question. For x=5 it is recorded population in ages 0, 1, 2, 3 and 4.

M x+5 …, M x+5 = enumerated population for the five ages in single years immediately greater than the age in question. For example for age 5 these figures are obtained from ages 6, 7, 8, 9 and 10.

Step 1:

The above procedure for calculating age ratios is repeated up to and including age A – 6, A is the age group beginning the open age interval, and in this case A = 80.This implies that age ratios can only be computed up to age 74.

Step 2:

To get mean age ratios;

In order to get more precise results geometric mean age ratios for five ages in single years starting from age 5 were obtained by adding the age ratios in five ages for example age ratios for ages 5, 6, 7, 8, 9 are added and then dividing the sum by five and presenting the results in five years age groups. This will be useful in determining the age groups that gained persons as a result of age heaping.

Step 3:

To get the absolute mean deviation;

Deviations of age ratios from 1 both positive and negative were then summed up regardless of the sign. Absolute Mean deviation was calculated for five ages and ten ages separately by taking the summation of absolute deviations and dividing this by five and ten respectively. The importance of the results obtained gives the pattern of age heaping in odd and even age groups.

2.4 Methods of correcting age misreporting

2.4.1: Feeney's Method

Age distributions obtained from censuses and surveys in most developing countries exhibits heaping on ages that are multiples of five. Feeney (1979) suggested a technique for redistributing the excess numbers of persons at these ages to the surrounding ages .The procedure involves the transfer of excess persons from a given multiple of five to the eight immediately surrounding ages, the four years on either side of the multiple of five, in such a way that ;the adjusted numbers in the surrounding ages are proportional to the original numbers and that the adjusted numbers at the lower four ages, the central age, and the upper four ages form a linear progression. This method involves two steps;

Step 1:

Considering all the multiples of five together, the increments to the numbers of persons intermediate between two successive multiples of five are made independently from the lower and upper multiple of five. That is, ages x-4 and x+4 are given a double reallocation. This is represented by the following equation;

$$M'_x = M_x - (B_x - 1)\left(M_{(x-5)^+} + M_{x^+}\right)$$

Where

$$B_x = \frac{8}{9}\,\frac{M_{x^-} + M_x + M_{x^+}}{M_{x^-} + M_{x^+}}$$

M'_x = Adjusted number of persons aged x in completed years

M(x-5) + = Number of persons aged x-4 to x-1 in completed years.

Mx+ = Number of persons aged x+1 to x+4 in completed years.

Bx = Constant of proportionality

Step 2:

The adjusted number of persons aged x+1 to x+4 in completed years is obtained by;

$$M'x^+ = M_{x^+} + (B_x - 1)M_{x^+} + (B_{x+5} - 1)M_{x^+}$$

$= (B_x + B_{x+5} - 1)M_{x^+}$ where;

Bx = constant of proportionality for the age in question.

BX+5 = constant of proportionality for the multiple of five following the age in consideration.

Where x is a multiple of 5(i.e. x=5, 10, 15...).

The equations above are valid for x=0, 5... N-5, where B_0 and B_N are taken to be equal to one denoting the age (also a multiple of five) that begins the open ended age group. The adjustment procedure is repeated until the values of B_x converge as close to one as desired. Numbers of persons at single years of age are found by interpolating (e.g., linearly) the final adjusted numbers of persons at ages that are multiples of five. Since the first adjusted population is not equal to the recorded population it is advisable to get the ratio for the recorded and the adjusted population. This ratio is then multiplied by the first adjusted population in order to get the final adjusted population. The discrepancies in the final results could have been attributed by the Feeney's method which gives results that are too low at ages 0-4 and too high in open ended age group while it already known that the two ages suffer from severe systematic errors.

It is important to note that Feeney method can only be used to adjust data starting from age 4 up to age 74. Age groups 75 – 79 and 80+ were left unadjusted since the method does not allow for this computation. The values in decimal places were approximated to four decimal places for easier presentation of the results. This implies that there are other values in decimal places which are not appearing in this final computation.

2.4.2 Demeny - Shorter Method

Demeny-Shorter method is a census survival technique that is used to adjust the age distributions produced by two consecutive censuses. The method assumes that the probability of intercensal survival are known, that both censuses achieved the same level of coverage and that both suffered from similar proportionate age-reporting errors; that is, the recorded population of age group x is equal to the true Population of the age group multiplied by an adjustment factor typical only of the age group which is assumed not to change from census to census. Given two censuses that were held ten years apart, the first at time t2 and the second at time t1, the method is best applied when the intervening age distribution is projected thus creating two five years intervals and applying the method twice. This is carried out by computing the growth rate for five years and then applying it to the initial population in order to get the estimated population. Having the two censuses five years apart, the population aged 0-4 at the time of the first census is assumed to be correct and is projected forward to second census. The population result in the first census is multiplied by the survival ratios in the same age group. This becomes the projected population in the age group 5-9 in the second census. An adjustment factor for age group 5-9 in the first census is then obtained as the ratio of the projected to the reported population aged 5-9 at the second census. This ratio is then multiplied by the enumerated population in the first census. Then the adjusted population for age group 10-14 is found by multiplying the adjusted population in age group 5-9 in the first census by the survival ratio in the same age group. The result will give the adjusted population in age group 10-14 in the second census. Then an adjustment factor for age group 10-14 in the first census is found as the ratio of the projected to the actual population and this ratio is multiplied by the enumerated population in the same age group in the first census. This chaining process continues from age group to age group except the open ended age group where a different procedure is used whereby the adjustment factor in this case, M(A+), is equal to the ratio of the adjusted to the reported population aged A and over in both censuses. After this process there are two adjusted age distribution for the first and second census. However, the last process is to plot out a graph for the presentation of the correction factors in both censuses. The method involves five key steps;

Step 1:

The estimation of the intervening age distribution which was 1994 was obtained for each age group from x to x+4 by the following;

$$5M_x = {}_5M1_x \exp[5r(x)]$$

$$r(x) = \frac{(\ln {}_5M2_x - \ln {}_5M1_x)}{(t_1 - t_2)}$$ Where,

r (x) = growth rate for five years is applied to the initial population in order to get the estimated population using the following formula.

${}_5M1_x$ and ${}_5M2_x$ denote the reported populations in the age group from x to x + 4 in the first and second censuses respectively.

Step 2:

Estimation of the mortality level using the appropriate life table model was carried out since poor choice of a model will give unreliable results as well as wrong interpretations will be made. Linear interpolation method was applied and the data used was adult mortality based on ophanhood information and estimates of child mortality using, Coale Trussel estimates obtained from CBS (2002) Vol v which gives the estimates q(x) values for child mortality in both 1989 and 1999 Kenyan population Censuses. However this procedure is well indicated at the appendices where the estimated mortality level obtained for males and females was 15.6326 and 15.4697 respectively using North Model. This will help to estimate the survival ratios.

Step 3:

The age distribution in age group 0-4 is assumed to be correct such that the enumerated population and the first adjusted population in this age group are equal. Then the population in the same age group (0-4), in the first census (1989) is projected forward to age group 5-9 at the second census (1994), obtained as;

$5M1_0 * {}_5l_0 = {}_5M'2_5$ Where;

$5M1_0$ = The enumerated population in 0-4 age group during the first census (1989)

${}_5l_0$ = The estimated survival ratio in 0-4 age group for the first census (1989)

${}_5M'2_5$ = The corrected population in 5-9 age group during the second census (1994)

To get the corrected population for the age group 5-9 in the first census (1989) the following formula is used;

$\frac{5M'2_5}{5M2_5} * 5M1_5 = 5M'1_5$ Whereas;

M1 and M2 refer to first and second censuses respectively.

$\frac{5M'2_5}{5M2_5}$ = adjustment factor for 5-9 age group in the first census (1989) whereby, the numerator is the corrected population aged 5-9 in the second census and the denominator is the enumerated population aged 5-9 in the same census. This adjusted factor is multiplied by the enumerated population aged 5-9 in the same census.

$5M1_5$ = enumerated population aged 5-9 in the first census (1994).

$5M'1_5$ = corrected population aged 5-9 in the first census (1989).

Similar procedure is repeated for the age group 10-14 and the procedure is summarized as indicated below;

$$5M1_5 * {}_5l_5 = M'2_{10}$$

$$\frac{M'2_{10}}{M2_{10}} * 5M1_{10} = 5M'1_{10}$$

These notations can be defined as above but replacing the population aged 0-4 by 5-9 and population aged 5-9 in the above equation by 10-14.

This chaining process is repeated for each proceeding age groups except in the last open ended interval in this case, 75+ whereby a different process is used.

Step 4:

Suppose that the correction factor for the open ended age group is given by the following formula;

$$K(75+) = \frac{({}_5M1_{70})({}_5l_{70})}{[M2(75+) - ({}_5l_{75+})M1(75+)]}$$

Where;

K (75+) = Correction factor in the open ended age group.

${}_5M'1_{70}$ = First corrected population in 1989 in age group 70 – 74

M2 (75 +) = Second census (1994) enumerated population in open ended age group 75+.

${}_5l_{75+}$ = Survival probability in open ended age group.

$_5l_{70}$ = Survival probability in age group 70-74.

Step 5:

There are two first corrected populations for the first and second census respectively based on the assumption that, the youngest age group 0-4 was accurately recorded by both censuses. However, the enumerated and the corrected population are not equal which implies that there was no validity for the above assumption. It is therefore important to get the ratio for the enumerated and the first adjusted population in each census and in order to reduce the marginal errors, this ratio is multiplied by the correction factors previously obtained by the chaining process and this gives the final correction factors in the two censuses five years apart, that is in 1989-1994. A similar procedure is followed for age group 1994-1999. It is at this stage when the assumption that was previously made that, the population aged 0-4 was correctly enumerated has been dropped since it was not valid. It is from this point that the two final correction or adjustment factors that have already been obtained in 1989-1994 and 1994-1999 are considered by getting their geometric mean. The average correction factors are computed by getting the geometric mean in each age group and this gives the final correction factors. This is multiplied by the enumerated population in each age group for 1989 and 1999 separately. The product of this computation gives the final corrected population. This will be presented in the following chapter. This step will help to identify the nature of age misreporting as well as in the correction of those errors.

It is important to note that the mortality levels for the intervening years were obtained by linear interpolation. In addition, projection of 1989 and 1999 Kenyan population censuses was carried out using the adjusted populations so as to make it possible for the comparison of the results in this study with the CBS data for the same censuses. The age specific fertility rates and the survival ratios were obtained from 1993 and 1998 Kenyan Demographic and Health Surveys.

CHAPTER THREE

AGE MISREPORTING IN KENYA

This chapter presents the results for age heaping by applying age ratio technique.

3.1 Detecting age heaping

Table 3.1.1 gives the summary of age ratios in Kenya 1989 and 1999 censuses. In males' age heaping occurred in terminal digits 10, 12, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 and 70 but those that had highest heaping (with an age ratio of 1.15 and above) were 60, 70, 50, 40, 30, 45, 35, 25 and 65. In 1989 and 1999 age heaping also occurred in terminal digits 5, 9, 26, 28, 67 and 13, 38 47 respectively. In general there was a tendency to heap on ages ending in terminal digits 0 and 5 except age 5 in 1999. However all the other ages were avoided but the ages that were most avoided (with an age ratio of 0.85 and below) were 73, 66, 44, 74, 34, 33, 64, 31, 41, and 46 indicating that there was avoidance of ages ending with terminal digits 1, 3, 4 and 6. Other ages that were most avoided were72, 63, 23, and 62 in 1989 and 61, 58, 59, 53, 68, 48, and 54 in 1999 . In overall there was a decrease in age heaping since there was no age heaping for age 5 and 6 in 1999 as compared to 1989 although preference for terminal digits 7 and 8 was similar in both censuses. Also there was no avoidance of age 2 and 3 in 1999 but instead there was preference of terminal digits 1, 4, 8 and 9 which lead to an overall increase in ages that were avoided in 1999.

In females age heaping occurred in terminal digits 10, 12, 18, 20, 28, 25, 30, 35, 40, 45, 50, 55, 60, 65, 69 and 70 but those that had highest heaping were 60, 70, 50, 40, 30, 45, 35, 20, 25, 65 and 10. In 1989 and 1999 age heaping also occurred in terminal digits 5, 8, 9, 59, 68 and 13, 22, 38, 47 and 56 respectively. In general there was a tendency to heap on ages ending in digits 0 and 5 except age 5 in 1999. However all the other ages were avoided but ages that were most avoided were 73, 66, 74, 34, 44, 33, 62, 41, 43, 53, 51, 72, 63 and 52 giving an evidence that there was avoidance of ages ending with terminal digits 1, 2, 3, 4, and 7 . Other ages that were avoided were 31, 57, 47, 42, 35 in 1989 and 61, 33, 58, 46, 29, 59 in 1999. In overall there was change in age reporting in that there was no age heaping for age 5 in 1999 as compared to 1989 although preference for terminal digits 7, 8 and 9 was similar in both censuses. In addition there was preference of terminal digits 2, 3, 6 and 7 which occurred in 1999. Age avoidance increased in 1999 in that there was avoidance of terminal digits 6, 8 and 9 which was not the case in 1989 while age heaping decreased in the same year.

Age misreporting in males and females was similar in that there was preferences of terminal digits 0 and 5 except the terminal digit 5 in 1999 and a tendency to avoid ages ending with digits 1, 3 and 4. However both age heaping and avoidance was more pronounced in females than in males.

In both censuses age heaping occurred in terminal digits 10, 12, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 69 and 70 but those that had highest heaping were 60, 70, 50, 40, 30, 45, 35, 20, 25, 65 and 10. In 1989 and 1999 age heaping also occurred in terminal digits 5, 8, 9, 28 and 13, 38 and 47 respectively. In general there was a tendency to heap on ages ending in digits 0 and 5 except age 5 in 1999. However all the other ages were avoided but ages that were most avoided were 73, 66, 74, 34, 44, 33, 62, 41, 43, 53, 51, 72, 63 and 52 implying an avoidance of terminal digits 1, 2, 3, 4 and 6. Other ages that were avoided were 63, 62, 57, 23, 47 and 61, 58, 46, 59, 29 in 1989 and 1999 respectively overall there was change in age reporting in that there was no age heaping for age 5 in 1999 as compared to 1989 although preference for terminal digits 8 and 9 in both censuses was similar. This was in addition to the preference of terminal digits 3 and 7 which occurred in 1999. Age avoidance also changed from terminal digit 7 in 1989 to 9 in 1999. There was deterioration in 1999 census age data which is reflected by increase in heaping caused by an increase of ages which had age ratios greater than 1.15 1999 as compared to 1989. In addition, age avoidance increased in 1999 censuses.

Table 3.1.1: 1989 and 1999 Kenya Population Census age ratios

	Males age ratios			Females age ratios		
Age	1989	1999	Total	1989	1989	Total
5	1.0928	0.8926	1.0758	0.8804	1.0844	0.8865
6	0.9863	0.9432	0.9909	0.9491	0.9886	0.9461
7	0.9349	0.9545	0.9288	0.9419	0.9318	0.9483
8	0.9916	0.9847	1.0126	0.9976	1.0020	0.9911
9	1.0137	0.9988	1.0153	0.9957	1.0145	0.9973
10	1.1311	1.1604	1.1166	1.1553	1.1239	1.1579
11	0.8894	0.8640	0.9075	0.9167	0.8984	0.8902
12	1.0705	1.1701	1.0480	1.1317	1.0593	1.1510
13	0.9558	1.0262	0.9526	1.0111	0.9542	1.0187
14	1.0346	1.0106	1.0165	0.9774	1.0256	0.9940
15	1.0165	1.0534	0.9979	1.0180	1.0071	1.0355
16	0.9849	0.9770	0.9765	0.9746	0.9806	0.9758
17	0.9152	0.9616	0.8689	0.9384	0.8916	0.9498
18	1.1341	1.0911	1.1081	1.0768	1.1208	1.0838
19	0.8910	0.8760	0.8937	0.8861	0.8924	0.8812
20	1.1792	1.1476	1.3347	1.2819	1.2593	1.2168
21	0.8517	0.8848	0.8676	0.8888	0.8600	0.8869
22	0.8842	0.9500	0.9478	1.0072	0.9173	0.9797
23	0.8277	0.8963	0.8333	0.9198	0.8306	0.9086
24	0.9528	0.9613	0.9341	0.9558	0.9429	0.9584
25	1.2002	1.2231	1.2203	1.2481	1.2108	1.2362
26	1.0146	0.9342	0.9864	0.9054	0.9999	0.9192
27	0.9312	0.9732	0.9092	0.8983	0.9198	0.9343

Table 3.1.1 continued;

Age	Males age ratios 1989	Males age ratios 1999	Males age ratios Total	Females age ratios 1989	Females age ratios 1999	Females age ratios Total
28	1.0119	0.9403	1.1255	1.0183	1.0700	0.9803
29	0.9424	0.8725	0.8827	0.8135	0.9118	0.8423
30	1.5476	1.5700	1.6665	1.6996	1.6079	1.6358
31	0.7788	0.7439	0.6662	0.6388	0.7216	0.6904
32	0.9721	0.9563	0.8783	0.9290	0.9247	0.9425
33	0.7430	0.7350	0.6808	0.6702	0.7116	0.7019
34	0.7219	0.8070	0.6955	0.7430	0.7088	0.7746
35	1.3064	1.3068	1.3358	1.3691	1.3211	1.3384
36	0.8892	0.9945	0.8936	0.9975	0.8913	0.9960
37	0.9980	0.8514	0.9465	0.8561	0.9725	0.8537
38	0.8675	1.0046	0.9528	1.1326	0.9096	1.0686
39	0.8835	0.9400	0.8460	0.8674	0.8648	0.9033
40	1.6431	1.5848	1.8851	1.7559	1.7628	1.6709
41	0.8386	0.7852	0.7362	0.6893	0.7871	0.7366
42	0.9761	0.9784	0.8219	0.8710	0.8984	0.9241
43	0.7990	0.7705	0.7431	0.7290	0.7708	0.7496
44	0.7164	0.6325	0.7247	0.6280	0.7206	0.6303
45	1.3355	1.3741	1.3973	1.4503	1.3668	1.4123
46	0.8486	0.7773	0.9305	0.7772	0.8896	0.7772
47	0.8672	1.0744	0.7976	1.0182	0.8321	1.0464
48	0.9237	0.8383	0.9523	0.8868	0.9382	0.8624
49	0.8916	0.9220	0.8861	0.8574	0.8888	0.8898
50	1.7529	1.7462	2.0661	1.9873	1.9095	1.8654
51	0.6957	0.8490	0.6754	0.7813	0.6855	0.8152
52	0.9206	0.9741	0.7911	0.8316	0.8556	0.9027

Table 3.1.1 continued;

	Males age ratios			Females age ratios		
Age	1989	1999	Total	1989	1999	Total
53	0.8678	0.7856	0.7507	0.7008	0.8087	0.7430
54	0.8780	0.8487	0.8694	0.8535	0.8737	0.8511
55	1.1654	1.0936	1.0639	1.0974	1.1132	1.0955
56	0.9811	0.9529	0.8519	1.0221	0.9154	0.9879
57	0.8664	0.9235	0.7945	0.8531	0.8299	0.8876
58	0.8812	0.7600	0.9276	0.7636	0.9048	0.7618
59	0.8595	0.7817	1.0129	0.8378	0.9377	0.8106
60	1.9084	1.9798	2.4011	2.3584	2.1569	2.1740
61	0.8547	0.6750	0.8659	0.6641	0.8605	0.6693
62	0.8380	0.8564	0.7411	0.7434	0.7878	0.7975
63	0.7283	0.9452	0.6532	0.7996	0.6892	0.8685
64	0.7599	0.8111	0.7134	0.7958	0.7356	0.8030
65	1.3356	1.2180	1.2963	1.2368	1.3151	1.2281
66	0.6635	0.7524	0.5552	0.6383	0.6078	0.6921
67	1.0341	0.9744	0.7853	0.8466	0.9055	0.9068
68	0.9541	0.8268	1.0032	0.9464	0.9793	0.8897
69	0.9971	0.9382	1.2094	1.1262	1.1051	1.0374
70	1.9071	1.9593	2.2794	2.3404	2.0937	2.1585
71	0.9305	0.9270	0.9439	0.8921	0.9372	0.9085
72	0.6261	0.8716	0.6749	0.7939	0.6508	0.8305
73	0.4609	0.7044	0.4451	0.6085	0.4528	0.6533
74	0.7167	0.7103	0.6282	0.6479	0.6713	0.6772

Table 3.1.2 gives the mean age ratios in five years age groups. This is useful in identifying the amount of error in each age interval. From the results, the age groups that gained persons as a result of age heaping in males i.e. those with mean age ratios greater than one are age groups were 10-14, 50-54 and 60-64 in both censuses. The difference in age misreporting between males in both censuses is that in 1989 there was heaping in age groups 5-9 and 25-29 in addition to those similar age groups that had undergone heaping in the two censuses. All the other age groups lost part of their population as a result of age avoidance i.e. mean age ratios below one. In females the age groups 50-54 and 60-64 gained persons as a result of age heaping. The difference in age misreporting for females in both censuses is that in 1999 age groups 5-9, 10-14, 25-29, 20-24 and 35-39 also gained persons as a result of age heaping in addition to similar age groups in the two censuses. All the other remaining age groups lost persons due to age avoidance. In both censuses the age groups that gained persons were 10-14, 50-54 and 60-64. Additional age groups that gained persons as a result of age heaping were 5-9, 10-14, 25-29 and 35-39, 70-74 in 1989 and 1999 respectively. All the other age groups in both censuses lost persons as a result of age avoidance. Therefore it is important to note that age heaping was common in males, females and in both censuses but it was more pronounced in females as compared to males and almost similar in the two censuses.

Table 3.1.2: Mean age ratios in five years age groups

	Males		Females		Total population	
	Mean age ratios	Mean age ratios	Mean age ratios	Mean age ratios	Mean age ratios	Mean age ratios
Age group	1989	1999	1989	1999	1989	1999
5-9.	1.0039	0.9548	1.0047	0.9529	1.0043	0.9539
10-.14	1.0163	1.0463	1.0082	1.0384	1.0123	1.0423
15-19	0.9883	0.9918	0.9690	0.9788	0.9785	0.9852
20-24	0.9391	0.9680	0.9835	1.0107	0.9620	0.9901
25-29	1.0201	0.9887	1.0248	0.9767	1.0224	0.9825
30-34	0.9527	0.9624	0.9175	0.9361	0.9349	0.9490
35-39	0.9889	1.0194	0.9949	1.0445	0.9919	1.0320
40-44	0.9946	0.9503	0.9822	0.9346	0.9880	0.9423
45-49	0.9733	0.9972	0.9928	0.9980	0.9831	0.9976
50-54	1.0230	1.0407	1.0305	1.0309	1.0266	1.0355
55-59	0.9507	0.9023	0.9302	0.9148	0.9402	0.9087
60-64	1.0179	1.0535	1.0750	1.0723	1.0460	1.0625
65-69	0.9969	0.9420	0.9699	0.9589	0.9826	0.9508
70-74	0.9282	1.0345	0.9943	1.0565	0.9612	1.0456

Table 3.1.3 gives Absolute deviations which are very useful in identifying the amount of error in each age interval. The results indicate a considerable amount of age errors arising from age heaping for the even age groups 10-14, 20-24, 30-34, 40-44 .... and so on, as compared to odd age groups such as 5-9, 15-

19, 25-29 among others. This indicates that age heaping was more common in the even age groups which could have been attributed by the preference of even age digits.

Table 3.1.3: Absolute deviations from one in five years age groups.

	Males		Females		Total population	
	Mean Deviation From 1 1989	Mean Deviation From 1 1999	Mean Deviation From 1 1989	Mean Deviation From 1 1999	Mean Deviation From 1 1989	Mean Deviation From 1 1999
5-9.	0.0388	0.0452	0.0368	0.0471	0.0361	0.0461
10-.14	0.0782	0.1007	0.0642	0.0808	0.0712	0.0887
15-19	0.0719	0.0660	0.0742	0.0591	0.0727	0.0625
20-24	0.1326	0.0910	0.1504	0.1050	0.1417	0.0967
25-29	0.0706	0.1006	0.1135	0.1299	0.0899	0.1120
30-34	0.2663	0.2656	0.3491	0.3437	0.3082	0.3053
35-39	0.1336	0.1051	0.1394	0.1562	0.1366	0.1308
40-44	0.2626	0.2836	0.3718	0.3677	0.3172	0.3261
45-49	0.1609	0.1822	0.1662	0.1894	0.1636	0.1859
50-54	0.2781	0.2578	0.3959	0.3640	0.3372	0.3107
55-59	0.1154	0.1351	0.1005	0.1330	0.1051	0.1295
60-64	0.3455	0.3384	0.4855	0.4711	0.4168	0.4071
65-69	0.1510	0.1452	0.2337	0.1863	0.1855	0.1554
70-74	0.4346	0.3492	0.5175	0.4796	0.4763	0.4178

Table 3.1.4 gives the summary of the mean deviation in ten years age groups. The degree of age heaping was observed to increase with age. This is indicated by an increase of mean deviation from one age group to the next following age group except for age group 20-24 for males in 1999. However all the other age groups followed this pattern. This could have been attributed by increase in memory lapse as a person becomes older.

Table 3.1.4: Absolute deviations in ten years age groups

Age	Males		Females		Total population	
	Mean deviation from one	Mean deviation from one	Mean deviation from one	Mean deviation from one	Mean deviation from one	Mean deviation from one
10-.14	0.0702	0.0728	0.0606	0.0639	0.0537	0.0674
20-24	0.1227	0.0661	0.1348	0.0820	0.1072	0.0796
30-34	0.2022	0.1703	0.2776	0.2368	0.1991	0.2086
40-44	0.2377	0.1884	0.3067	0.2619	0.2269	0.2284
50-54	0.2634	0.2122	0.3372	0.2767	0.2504	0.2483
60-64	0.2765	0.2149	0.3516	0.3021	0.2609	0.2683
70-74	0.3513	0.2410	0.4507	0.3330	0.3309	0.2866

CHAPTER FOUR

CORRECTION OF AGE ERRORS

From the previous chapter it is evident that age heaping contributes to age misreporting. The purpose of this section is to correct for age heaping. Feeney's method is used for the correction of age heaping in multiples of five. The steps previously explained were applied and the following results were obtained.

4.1 Application of Feeney's method to Kenyan data

Table 4.1.1 gives the results for age misreporting in 1989 using Feeney's method. The age groups that were over reported were 10-14, 30-34, 40-44, 50-54, 60-64 and 70-74 in males, females and in the total population but age group 20-24 was only over reported in females and total population. Over reporting in age group 20-24 in females could have been attributed by the tendency of women to prefer to appear younger especially if they are not yet married such that those aged between age group 25-29 would prefer to report their ages in age group 20-24. The remaining age groups were under reported. Over enumeration could have been attributed by preferences of even age groups such as 0-4 and 10-14 while under enumeration could be as a result of avoidance of odd age groups such as 5-9 and 15-19 among others.

Table 4.1.1: Adjusted age distribution in 1989, Feeney method

	1989 adjusted population			Differences of enumerated to adjusted		
Age	Males	Females	Total	Males	Females	Total
0-4	1904178	1879546	3806044	+7038	+9282	-6000
5-9	1750894	1733263	3479706	-7248	-7972	-10768
10-14	1491429	1474872	2962501	+12616	+10775	+27192
15-19	1197101	1237652	2431942	-19118	-36940	-53248
20-24	890284	1003153	1890683	-690	+10188	+12251
25-29	805459	878594	1681631	-22985	-31306	-51868
30-34	564248	544858	1108063	+19526	+30794	+51362
35-39	469304	474270	942256	-8356	-16327	-23363
40-44	354353	342630	696483	+13580	+21616	+35696
45-49	286521	304976	590442	-5396	-11571	-15912
50-54	222582	217482	439675	+13325	+23136	+36850
55-59	188546	199788	387814	-9530	-18633	-27642
60-64	141532	151527	292714	+8964	+16375	+25685
65-69	118343	124463	242502	-4655	-7484	-11834
70-74	79711	82819	162346	+3255	+8394	+11832
75+	148809	154546	303355	-	-	-
Age NS	14740	10508	25248	-	-	-
	10628035	10814947	21443404	-	-	-

Table 4.1.2 gives the results of age misreporting in 1999 using Feeney's method. In 1999 The age groups that were over enumerated were i.e. those with a plus sign were 30-34, 40-44, 50-54, 60-64 and 70-74 in males, females and in total population. The other remaining age groups were under enumerated. There was a decrease of age groups that had been over reported and an increase of those that had under reported their ages in 1999 as compared to 1989. Increase in under reporting from age group 5-29 could have been attributed to the fact that this age gap consist of those that were late in

joining primary school as well as those that were late in completion of secondary education thus leading to under reporting of ages, preferences of women in these age group to appear younger especially those that are not yet married thus under reporting their ages and high tendency of males to migrate to other areas in search of job opportunity which leads to their omission in the enumeration exercise.

Table 4.1.2: Adjusted age distribution in 1999, Feeney method

	Adjusted Population			Difference of enumerated to adjusted		
Age	Males	Females	Total	Males	Females	Total
0-4	2059730	2015532	4075941	+232206	+227434	+458961
5-9	2077262	2038008	4115960	-76682	-75452	-152824
10-14	2056243	2020529	4077412	-21263	-16874	-38777
15-19	1725721	1789733	3516440	-43737	-68539	-113262
20-24	1359682	1526928	2886748	-31153	-22539	-53831
25-29	1142599	1223000	2365784	-47690	-58406	-106281
30-34	822931	812796	1636499	+17761	+32434	+49424
35-39	718236	754968	1473056	-22973	-31219	-54044
40-44	506509	497053	1004039	+9993	+19936	+29452
45-49	435949	436891	870282	-16108	-17904	-31455
50-54	331128	315604	644570	+13511	+24563	+40237
55-59	240688	261903	502523	-16997	-25578	-42507
60-64	183543	195852	379603	+10970	+18863	+29625
65-69	149758	174826	324505	-8789	-14462	-23173
70-74	112792	122915	235845	+5809	+12609	+18281
75-79	79166	81620	160786	-	-	-
80+	95300	121038	216338	-	-	-
Age NS	103487	86956	190443	-	-	-
Total	14200725	14476152	28676775	-	-	-

The difference between the two censuses was that in 1999 age group 10-14 and 20-24 were not over reported as in 1989 but instead they were under reported. Females age groups 10-14 and 20-24 were over reported in 1989 whereas in 1999 they were under enumerated. The age group 10-14 for the males was over reported in 1989 while it was under enumerated in 1999. Specifically, it implies that there was an increase in under reporting in males, females and the total population in 1999 census data. This could have been attributed by low level of education among the respondent which could have been caused by high cost of education such that majority of the people are denied to have education, insufficient training that is given to enumerators, ignorance among the respondents and hiring of people who are not demographers or does not know the importance of age such that instead of interviewing the people they just forge their data.

4.2 Application of Demeny-Shorter Method to Kenyan Data.

Due to the systematic age errors already reflected in the previous chapter Demeny- shorter method was applied in this section in order to further detect these errors as well as to correct the same.

4.2.1 Age reporting for males

It is important to note that the correction factor can be less than or greater than one. Correction factor greater than one indicates under – reporting of age in that age group while if less than one, indicates over-reporting of age in that age group.

Table 4.2.1 gives the results obtained after the application of Demeny-Shorter method. The age groups that were under – reported include 0-4, 5-9, 20-24, 30-34, 40-44, 60-64 and 65-69 whereas age groups 10 – 14, 15 – 19, 25 – 29, 50 – 54 70 – 74 and 75+ were over-reported. Under-reporting in age group 0 – 4 could be attributed by omission of young children as well as over – reporting of age among those under five so that they can be considered in the school going age. Over – reporting in age group 15 – 19 could be contributed to those aged 20 – 24 who are physically late in maturing. This leads to under – reporting their ages and thus falling in the 15 – 19 age group. In addition systematic over – reporting of age occurs above age 30 years and this carries people to the next higher age groups, finally leading to considerable excesses in the age groups 70 – 74 and 75+.

Table 4.2.1: Adjusted male population in 1989 and 1999, Demeny-Shorter Method.

	Enumerated population		Estimated Survival	Correction	Corrected population	
Age	1989	1999	Ratios	Factors	1989	1999
0-4	1911216	2291936	0.9478	1.0296	1967820	2420284
5-9	1743647	2000580	0.9787	1.0101	1761242	2094858
10-14	1504045	2034980	0.9826	0.9449	1421135	1914077
15-19	1177983	1681984	0.9762	0.9789	1153169	1616917
20-24	889594	1328529	0.9713	1.0242	911122	1320927
25-29	782474	1094909	0.9697	0.9715	760157	1049905
30-34	583773	840692	0.9667	1.0448	609914	860648
35-39	460949	695263	0.9611	1.0298	474692	693294
40-44	367933	516502	0.9527	1.0651	391869	542537
45-49	281126	419841	0.9393	1.0703	300899	436237
50-54	235906	344639	0.9203	0.9966	235104	335315
55-59	179017	223691	0.8909	1.1241	201238	255311
60-64	150496	194513	0.8434	1.0389	156352	203473
65-69	113689	140969	0.7721	1.0526	119667	150942
70-74	82965	118601	0.6716	0.8989	74574	104654
75+	148809	174466	0.4170	0.5314	79076	95606
	10613622	14102095			10618030	14094985

4.2.2 Age reporting for females

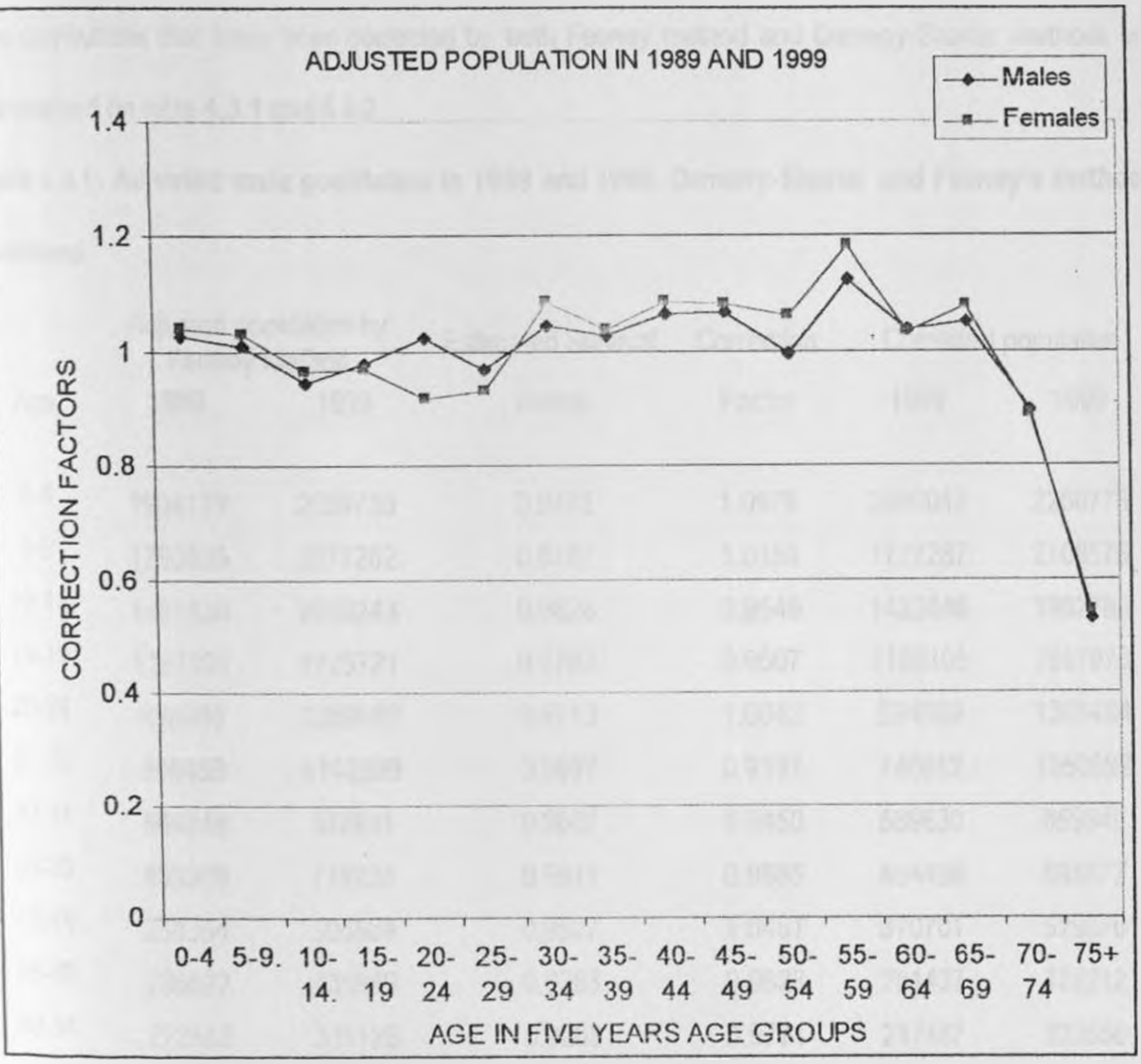
Table 4.2.2 gives the results obtained by Demeny-Shorter methods in males. The age groups that were under – reported were 0 – 4 , 5 – 9, 35 – 39, 40 – 44, 45 – 49, 50 – 54, 55 – 59 and 65 – 69 whereas age groups that were over – reported were 10 – 14, 15 – 19, 20 – 24, 25 – 29, 70 – 74 and 75+. 0 – 4 age group is commonly under – reported due to omissions. This could as a result that in certain communities, children are not considered as full members of the society. Those children living with relatives during the enumeration exercise are not considered in the enumeration exercise.10 – 14 age group gains persons from age group 5 – 9. Young girls who mature early have a tendency to over – report their ages so that they appears older. 15 – 19 age group gains persons from age group 10 – 14 since girls who are past menarche have a tendency to overstate their ages so that they appears older and this carries some to age group 15 – 19. At age 35 onwards , it appears that systematic over – reporting of age carries people across the age group boundaries to the next higher age groups and this finally leads to the age group 75+ being considerably over – reported. This could have been due to the fact that the society gives respect to the elderly and thus the desire for most people above age 35 preferring to appear older.

Table 4.2.2: Adjusted female population in1989 and 1999, Demeny-Shorter Method.

	Enumerated population		Estimated survival	Correction	Corrected population	
Age	1989	1999	Ratios	Factors	1989	1999
0-4	1888828	2242966	0.9525	1.0426	1969241	2338456
5-9	1725292	1962556	0.9803	1.0307	1778330	2022888
10-14	1485647	2003655	0.9845	0.9676	1437493	1938711
15-19	1200712	1721194	0.9816	0.9697	1164321	1669028
20-24	1013340	1504389	0.9783	0.9178	930057	1380748
25-29	847289	1164594	0.9750	0.9334	790891	1087075
30-34	575651	845230	0.9714	1.0877	626129	919347
35-39	457942	723749	0.9668	1.0378	475261	751121
40-44	364245	516989	0.9617	1.0868	395854	561853
45-49	293405	418987	0.9532	1.0842	318106	454260
50-54	240618	340167	0.9380	1.0624	255639	361402
55-59	181155	236325	0.9122	1.1823	214178	279405
60-64	167902	214715	0.8680	1.0342	173638	222051
65-69	116979	160364	0.7999	1.0814	126506	173424
70-74	91212	135524	0.7047	0.8923	81392	120933
75+	154546	202658	0.4424	0.5447	84184	110391
	10804763	14394062			10821220	14391095

Figure 1; Correction factors obtained by the Demeny-Shorter method

Figure 1 consists of correction factors for both males and females against age in five years age group. It is well indicated that under reporting was more pronounced in females as compared to males while over reporting was more common in females as compared to males. The reasons are as explained earlier from the results in table 4.2.1 and 4.2.2.



4.3 Feeney's and Demeny-Shorter Methods combined

Demeny- Shorter method is used to correct data that has already been adjusted by Feeney method in order to correct for both age heaping in multiples of five as well as systematic errors. It could have been important to compare the results obtained by applying Feeney's method to a data that has already been adjusted by Demeny-Shorter method. Unfortunately this was not possible since the Feeney method can only be useful if the data in single years is available and this is not the case with the data that has already been adjusted by Demeny Shorter method which is normally in five years age groups. This step was carried out in order to test whether finer results would be attained and this would help to judge the appropriate method that can be used in adjusting the Kenyan census data. The 1989 and 1999 census age distributions that have been corrected by both Feeney method and Demeny-Shorter methods are summarized on table 4.3.1 and 4.3.2

Table 4.3.1: Adjusted male population in 1989 and 1999, Demeny-Shorter and Feeney's methods combined.

Age	Adjusted population by Feeney Method 1989	1999	Estimated Survival Ratios	Correction Factor	Corrected population 1989	1999
0-4	1904179	2059730	0.9478	1.0976	2090042	2260776
5-9	1750895	2077262	0.9787	1.0151	1777287	2108575
10-14	1491430	2056243	0.9826	0.9546	1423646	1962789
15-19	1197101	1725721	0.9762	0.9607	1150105	1657973
20-24	890284	1359682	0.9713	1.0043	894069	1365464
25-29	805459	1142599	0.9697	0.9197	740812	1050892
30-34	564248	822931	0.9667	1.0450	589630	859949
35-39	469305	718236	0.9611	0.9685	454498	695577
40-44	354354	506509	0.9527	1.0461	370701	529876
45-49	286522	435949	0.9393	0.9823	281437	428212
50-54	222582	331128	0.9203	0.9771	217487	323550
55-59	188547	240688	0.8909	0.9767	184154	235081
60-64	141533	183543	0.8434	1.1424	161693	209688
65-69	118344	149758	0.7721	1.0293	121811	154146
70-74	79711	112792	0.6716	0.9596	76494	108240
75+	148809	174466	0.4170	0.5472	81424	95463
	10613295	14097237			10615290	14046252

Table 4.3.2: Adjusted female population in 1989 and 1999, Demeny-Shorter and Feeney's Methods combined.

Age	Adjusted population by Feeney method 1989	1999	Estimated Survival Ratios	Correction Factor	Corrected population 1989	1999
0-4	1879546	2015532	0.9525	1.0869	2042926	2190732
5-9	1733263	2038008	0.9803	1.0354	1794640	2110176
10-14	1474872	2020529	0.9845	0.9805	1446107	1981122
15-19	1237652	1789733	0.9816	0.9439	1168233	1689349
20-24	1003153	1526928	0.9783	0.9151	918005	1397322
25-29	878594	1223000	0.9750	0.8855	777988	1082956
30-34	544858	812796	0.9714	1.1207	610611	910884
35-39	474270	754968	0.9668	0.9758	462807	736720
40-44	342630	497053	0.9617	1.1088	379905	551128
45-49	304976	436891	0.9532	1.0040	306207	438655
50-54	217482	315604	0.9380	1.1105	241514	350479
55-59	199788	261903	0.9122	1.0156	202904	265989
60-64	151527	195852	0.8680	1.0782	163383	211177
65-69	124463	174826	0.7999	0.9417	117204	164630
70-74	82819	122915	0.7047	0.9168	75926	112686
75+	154546	202658	0.4424	0.5080	78506	102946
	10804439	14389196			10786867	14296948

From the results in tables 4.1.1, 4.1.2 and 4.2.1, 4.2.2 whereby Feeney and Demeny-Shorter methods were used separately and comparing this with the results in tables 4.3.1 and 4.3.2 whereby Feeney's method and Demeny-Shorter methods were combined it is well indicated that a combination of the two methods is more efficient in correcting age errors in Kenya than either of the two individual methods. In fact Feeney itself appears to worsen the data with regard to systematic age misreporting.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This study focuses on detection and correction of age errors using 1989 and 1999 Kenyan population Censuses. The main objective of this study was to detect and correct age errors in 1989 and 1999 Kenyan population censuses. Specifically the study was meant to, identify the nature and extent of age errors in these two censuses, make appropriate adjustments for age misreporting errors in 1989 and 1999 Kenyan population censuses and compare the corrected age data with the Central Bureau of Statistics (CBS) corrected data

5.1 Summary

It is well indicated in this study that there was age heaping in the two censuses .The pattern of age heaping was almost similar in that there was preference of digit 0 and 5 with an exception of age 5 in 1999 population census. This was in both males and females. In addition in both censuses the degree of age heaping for both sexes was observed to increase with age and it was more pronounced in even age groups as compared to odd age groups. However in case of systematic errors there was a difference between males and females. Males had a high tendency to over report their ages as compared to their female counterparts who instead underreported their ages. Over reporting in males occurred at older ages and this could have been contributed by the respect that the older people in the society are given. On the other hand, over reporting in females occurred among the teenagers and very old people 70+. For teenagers this could have been caused by some females who matures very early and as a result would like to appear older while those who engages in first marriages at an early age or bear children would like to over report their ages. Underreporting of age was common among females than males. For females underreporting was common in ages 30+ and the main contributing factor is that women prefer to appear younger than they appear especially those who are almost age 30 or over 30 and are not yet married. For males underreporting occurred among the teenagers and this could have been attributed by age misreporting for those who were late in maturing and those that had delayed in completion of secondary school education.

5.2 Conclusions

It is clear from this study that the Kenyan population census data is composed of age misreporting errors. The major sources of age errors are age heaping and systematic age errors. The former is caused by preferences of terminal digits 0 and 5 while the later is caused by under enumeration and over enumeration which carries people across the age group. The observed pattern of age heaping on terminal digits 0 and 5 appear to be consistent with what had already been observed in other developing countries indicating that the method used for this computation can be effective method. Age heaping plays a major role in reducing the quality of age data and this raises the need for correction of age heaping before the data is used for any analysis. For this purpose Feeney (1979) method was used for the correction of age heaping in multiples of five. Systematic errors were detected and corrected using Demeny – Shorter method. Similarly, in order to improve the quality of age data Demeny –Shorter method was used for the correction of systematic errors. Lastly, Demeny – Shorter method and Feeney (1979) method were combined to correct for both heaping and systematic errors and the results obtained were finer as compared to the results when the two methods were used separately. This made the combination to be effective in smoothening of Kenyan census data. This assumption was made when the census coverage was computed using the already adjusted population in 1979 (Magadi 1990) and 1989 adjusted values in this study so as to project the 1989 and 1999 population. The coverage rates obtained in 1989 and 1999 were 96.44, 99.7 in this study and 94.0 and 97.3 for CBS respectively. It was possible to achieve the objectives already outlined at the beginning of this study. The nature and extent of age misreporting in Kenyan census data was found to be age heaping and systematic errors. The former occurs as a result of preferences of terminal digits 0 and 5 while the later is due to systematic errors which carry people across the age group boundaries to the next higher or lower age group. However age misreporting was greater in females as compared to males. Both heaping and systematic errors could have been caused by socio-economic, socio-cultural, demographic and biological factors. However, there is no reason to suppose that these particular methods are the most accurate ones in adjustment of Kenyan population censuses but it is very important to find out if using alternative methods could have affected the results. Instead these methods have been vital since they have been used to conclude that Kenya census data is encountered by age misreporting errors due to age heaping and systematic errors. This implies that adjustment should not only be carried in consideration of age heaping alone as in the case of CBS but more emphasis should also be put on systematic errors especially when adjusting for census data.

5.3 Recommendations

5.3.1 Policy Makers

a) The population should be educated on the importance of reporting age accurately so as to weaken the traditional beliefs that people have on age and which leads to misreporting of age data.

b) The enumerators should be given sufficient training in order to reduce estimation of respondent's age in terms of physical characteristics and marital status among others. This will enable them to use any available information to obtain a more accurate data.

c) Education of girl child should be encouraged since it is through lack of knowledge that majority of women misreports their age.

d) Appropriate methods of analysis should be applied and the focus should not only be the correction of age heaping but instead should also be useful in correction of systematic errors which is also an error in age misreporting in Kenyan census data. This would help to come up with quality data. Appropriate data leads to comprehensive planning which is very crucial to the ability the government meet the needs of its people.

e) Data in regional levels should be made available for the analysis to be carried out. It is only through this that it is possible to know the differential factors that lead to age misreporting. Through this, it is possible for the government to handle the cause and get a solution and also have that equitable distribution of resources.

5.3.2 Further Research

a) Further study should be carried in order to assess how the socio-economic, social-cultural and demographic factors may lead to age misreporting as a result of age heaping and systematic errors.

b) A similar study should be carried out at the regional level and similar methods of analysis used.

c) Further study on adjusting age data using the other methods already discussed in this study is also useful in order to examine whether finer results can be obtained.

REFERENCES

Beckett Morgan, et al (1999), the quality of retrospective reports in the Malaysian Family Life Survey. Labor and population working paper series 99-13, National Institute of child health and Human Development, Dec1999.

Bhat (1990), Estimating Transition probabilities of age misstatement, Demography Vol 27 No 1,149-163.

CBS (2002), Kenya 1999 Population and housing. Analytical Report on Population Projection vol v11, Ministry of Finance and Planning. Nairobi.

CBS, (2001), The 1999 Population and Housing Census. Vol 1, Ministry of Finance and Planning.

CBS, (1994), Kenya Population Census, 1989, Vol 1, Ministry of Planning and National Development.

Carrier and Farrag, (1959). The Reduction of Errors in Census Populations for statistically Underdeveloped Countries: Population Studies, Vol 12:240-285.

Coale and Demeny, (1966), Regional Model Life Tables and Stable populations. Princeton University Press.

Dudley I. Poston,Jr (2000).An analysis of the quality of the age and sex data of the republic of South Korea and its Provinces 1970-1995, Department of Sociology U.S.A

Ewbank.D.C (1982). The sources of error in Brass method of estimating child survival. The case of Bangladesh, Population studies Vol 36: 459-474.

Ewbank.D.C (1981), Age misreporting and age selective under-enumeration; Sources patterns Washington D.C, national Academy press.

Fosu.B.G, (2001), Evaluation of Population Census Data Through Demographic Analysis. Symposium of Global Review of 2000 Round of population and Housing Censuses. Mid Decade Assessment and Future Prospects, Statistics Department, United Nations Secretariat. New Yolk.

Gupta (1975). A general method of correction for age misreporting in census populations. Demography. Vol12.no 2 303-312.

Kizito, M.L, (1985): The estimation of Adult Mortality Differentials in Kenya Using life table Technique", University of Nairobi, M.Sc.Thesis at P.S.R.I.

K'oyugi (1983) "Mortality and Morbidity Situation in Siaya District "MSC thesis, PSRI, University of Nairobi.

Magadi (1990),"Estimation of age distributions, census coverage and death completeness in Kenya."MSC thesis, PSRI, University of Nairobi.

Mudaki S.J, (1986)," Estimation Of Adult Mortality in Kenya Using Ophanhood Data" M.Sc. Thesis University Of Nairobi at P.S.R.I.

Murkherjee (1976), The age distribution of the Indian population. Reconstruction for the state and territories, 1881-1961 East West Centre U.S.A

Narasimhan et al (1997),Comparison of Fertility Estimates From India's Sample Registration System and National Family Health Survey International Institute for Population Sciences Mumbai, India.

Ntozi (1978), the Demeny Shorter and three-census methods for collecting age data. Demography Vol.4 509-521.

Retherford Robert,(2001),National fertility Health Survey. Subject Reports No.17.International Institute for Population Sciences, East West Center, Population and Health Studies, USA

Rono(1982),"Evaluation of various techniques for estimating mortality in Kenya' MSC Thesis.PSRI University of Nairobi.

Saxena and Gogte (1985), Feeney's method for correcting age distributions for heaping on multiples of five. Asian and Pacific census Forum Vol 11 No 3

Shryock and Siegel (1975), The methods and materials of Demography.Vol 1 Washington, D.C United States Government printings press.

United Nations, (1990), 1988 Demographic Yearbook, New Yolk, United States.

United Nations (1983), Manual X, Indirect Techniques For Demographic Estimation. Population Studies, No. 81, New Yolk, United Nations.

United Nations, (1955) Manual 11, Methods of appraisal of quality of basic data for population estimates. Population Studies, No 23, New Yolk. United Nations.

APPENDICES

1989 AND 1999 KENYAN POPULATION CENSUS.

DISTRIBUTION OF POPULATION BY AGE IN SINGLE YEARS AND FIVE YEARS AGE GROUPS DISTRIBUTION BY SEX.

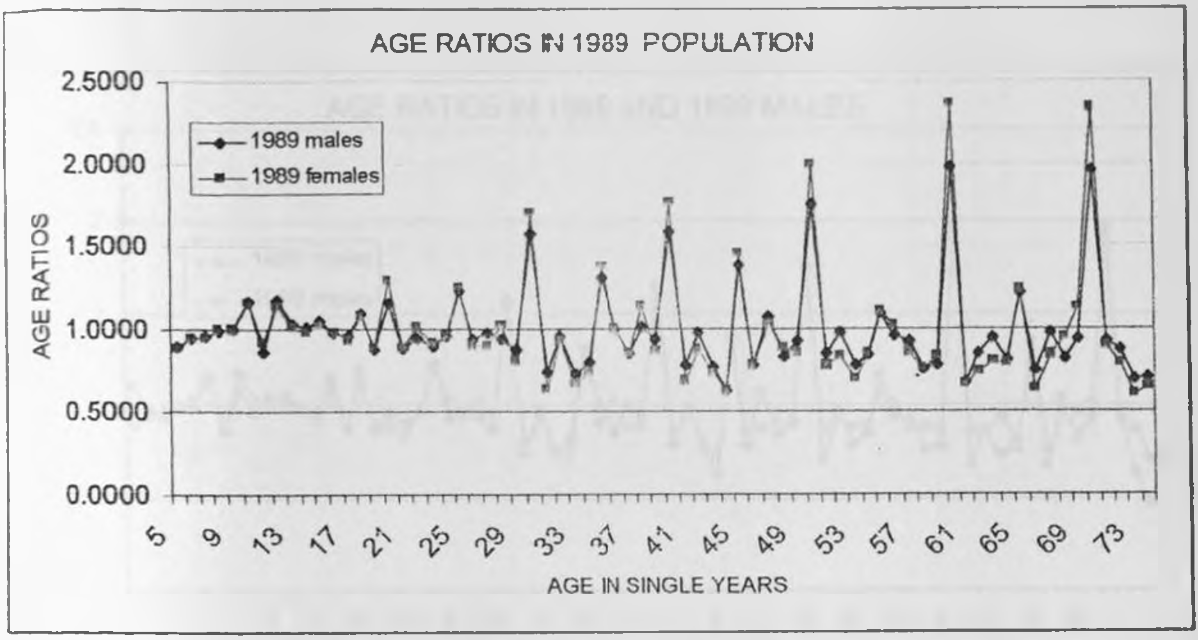
	1989 Population			1999 Population		
Age	Male	Female	Total	Male	Female	Total
Under 1	406759	400913	807672	555876	540722	1096598
1	339262	332254	671516	383031	373480	756511
2	395265	394215	789480	443325	436227	879552
3	393903	390271	784174	451109	446384	897493
4	376027	371175	747202	458595	446153	904748
0-4	1911216	1888828	3800044	2291936	2242966	4534902
5	395278	385008	780286	389416	376714	766130
6	348366	346316	694682	391682	388188	779870
7	329529	324146	653675	401992	390680	792672
8	336942	339652	676594	409378	407425	816803
9	333532	330170	663702	408112	399549	807661
5-9.	1743647	1725292	3468939	2000580	1962556	3963136
10	357423	349611	707034	459800	450283	910083
11	274993	278370	553363	348009	362962	710971
12	313189	305194	618383	451935	433903	885838
13	275699	274127	549826	396613	389279	785892
14	282741	278345	561086	378623	367228	745851
10-14	1504045	1485647	2989692	2034980	2003655	4038635
15	269559	269822	539381	386684	381489	768173
16	245055	250065	495120	342868	351348	694216
17	220135	217975	438110	330280	333913	664193
18	250956	258386	509342	349169	361998	711167

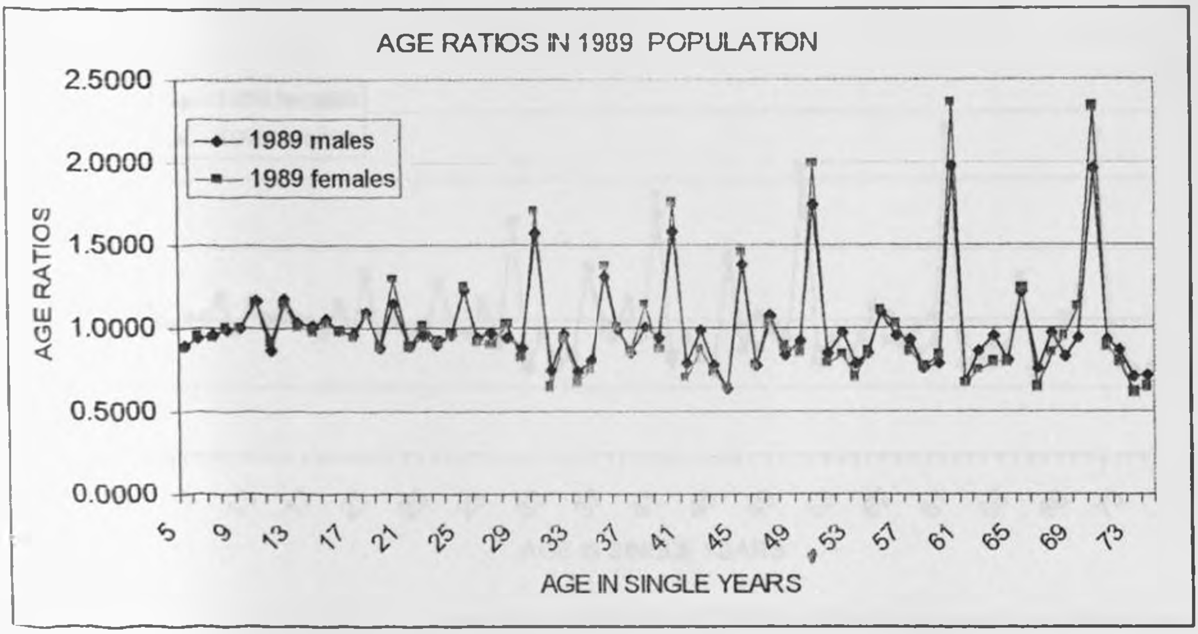
	1989 Population			1999 Population		
Age	Males	Females	Total	Males	Females	Total
19	192278	204464	396742	272983	292446	565429
15-19	1177983	1200712	2378695	1681984	1721194	3403178
20	239272	287393	526665	340365	404228	744593
21	169420	186822	356242	254758	277607	532365
22	167076	193975	361051	260746	298644	559390
23	151644	168736	320380	236190	265559	501749
24	162182	176414	338596	236470	258351	494821
20-24	889594	1013340	1902934	1328529	1504389	2832918
25	200743	226620	427363	295668	332006	627674
26	159744	169229	328973	214409	224972	439381
27	143250	150502	293752	214514	213940	428454
28	147107	171218	318325	196012	223049	419061
29	131630	129720	261350	174306	170627	344933
25-29	782474	847289	1629763	1094909	1164594	2259503
30	201279	223097	424376	289616	323016	612632
31	100133	88588	188721	136811	121911	258722
32	116720	107719	224439	163916	163540	327456
33	86433	80603	167036	122638	116450	239088
34	79208	75644	154852	127711	120313	248024
30-34	583773	575651	1159424	840692	845230	1685922
35	135636	138899	274535	198702	213706	412408
36	84660	83062	167722	137267	138854	276121
37	91408	85115	176523	116444	118515	234959
38	75528	80835	156363	127760	144039	271799
39	73717	70033	143750	115090	108635	223725
35-39	460949	457944	918893	695263	723749	1419012
40	129170	145133	274303	183275	205622	388897
41	64184	56984	121168	87980	79270	167250
42	70614	60314	130928	103341	94009	197350
43	55644	52641	108285	79361	76361	155722

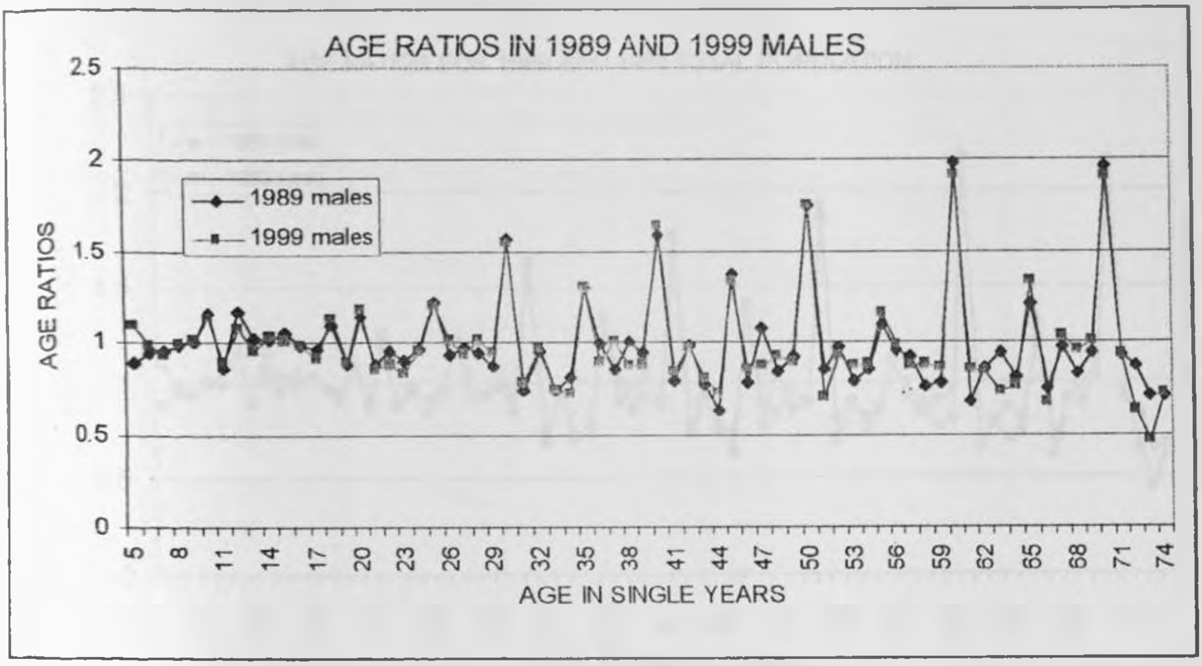
	1989 Population			1999 Population		
Age	Males	Females	Total	Males	Females	Total
44	48321	49173	97494	62545	61727	124272
40-44	367933	364245	732178	516502	516989	1033491
45	86422	92843	179265	127990	135889	263879
46	49908	54921	104829	67413	66414	133827
47	49152	46147	95299	88377	82880	171257
48	49364	52023	101387	66244	68978	135222
49	46280	47471	93751	69817	64826	134643
45-49	281126	293405	574531	419841	418987	838828
50	84481	99619	184100	123091	136869	259960
51	33636	33043	66679	58507	53768	112275
52	42113	36505	78618	64286	55118	119404
53	38450	33918	72368	48835	43978	92813
54	37226	37533	74759	49920	50434	100354
50-54	235906	240618	476524	344639	340167	684806
55	49409	47757	97166	63513	66427	129940
56	37079	33378	70457	47652	52222	99874
57	32365	30580	62945	44028	42382	86410
58	31137	33885	65022	34557	36530	71087
59	29027	35555	64582	33941	38764	72705
55-59	179017	181155	360172	223691	236325	460016
60	58470	74870	133340	75974	95240	171214
61	26105	28143	54248	26504	28304	54808
62	24531	23344	47875	31558	29763	61321
63	20682	20206	40888	32760	30901	63661
64	20708	21339	42047	27717	30507	58224
60-64	150496	167902	318398	194513	214715	409228
65	35346	37625	72971	41508	48028	89536
66	16074	14226	30300	22957	21878	44835
67	22813	18545	41358	28548	27844	56392
68	19748	21818	41566	23314	29575	52889

	1989 Population			1999 population		
Age	Males	Females	Total	Males	Females	Total
69	19708	24765	44473	24642	33039	57681
65-69	113689	116979	230668	140969	160364	301333
70	35807	42988	78795	46801	61262	108063
71	16807	17477	34284	21747	23510	45257
72	11189	12360	23549	19963	20425	40388
73	7842	7958	15800	15340	15104	30444
74	11320	10429	21749	14750	15223	29973
70-74	82965	91212	174177	118601	135524	254125
75	26913	23389	50302	26100	27135	53235
76	9218	8688	17906	12176	12053	24229
77	8526	7086	15612	15621	12541	28162
78	10910	9783	20693	12669	13465	26134
79	11032	11530	22562	12600	16426	29026
75-79	66599	60476	127075	79166	81620	160786
80&over	82210	94070	176280	95300	121038	216338
Age NS	14740	10508	25248	103487	86956	190443
Total	10628368	10815268	21443636	14,205,582.00	14481018	28686600

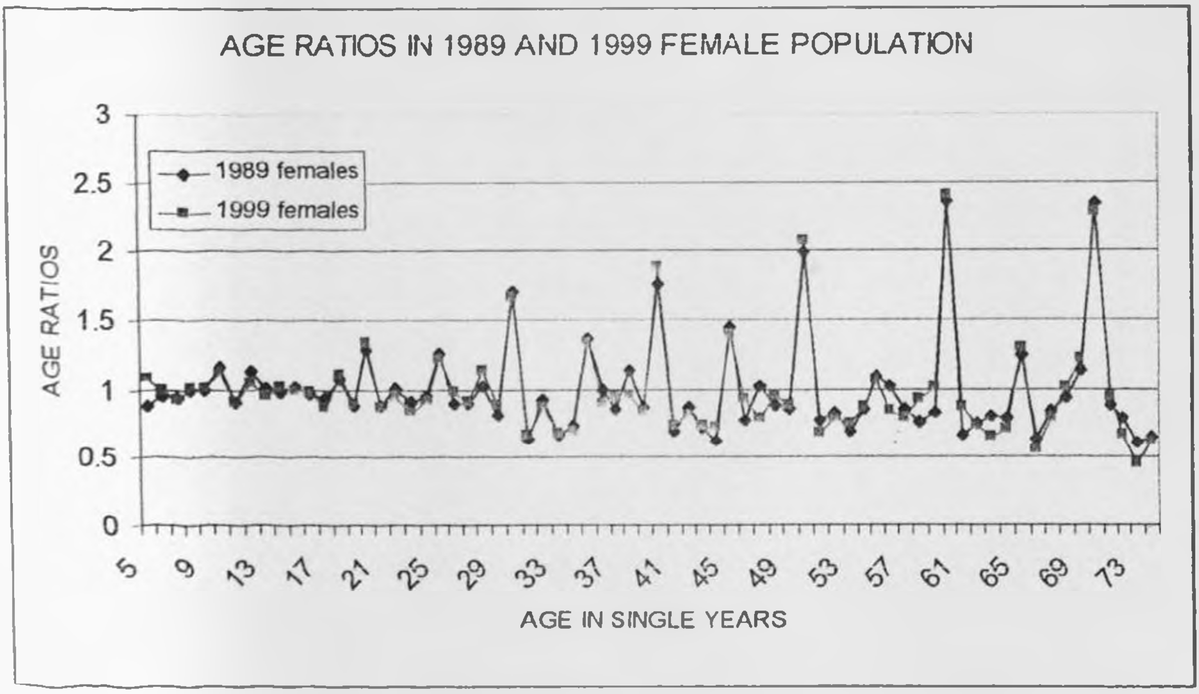
Source: CBS (2001) and CBS (1994)


AGE RATIOS IN 1989 POPULATION
1989 males
1989 females
AGE RATIOS
2.5000
2.0000
1.5000
1.0000
0.5000
0.0000
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
AGE IN SINGLE YEARS

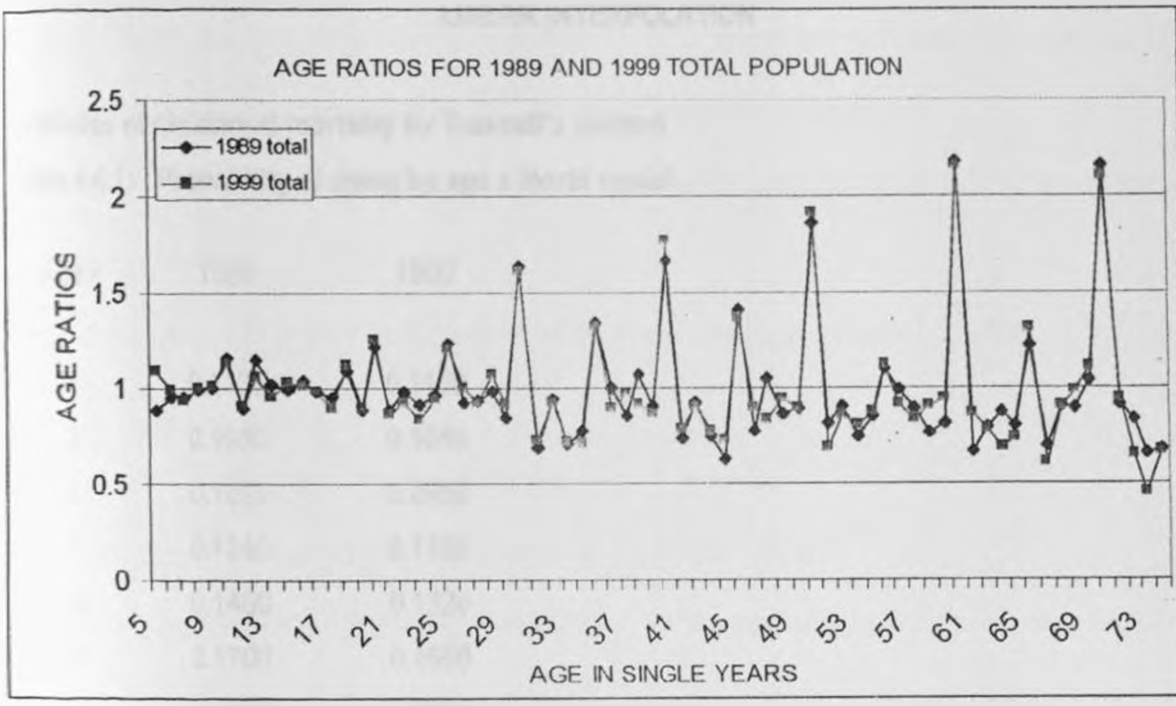

AGE RATIOS IN 1989 POPULATION
1989 males
1989 females
AGE RATIOS
2.5000
2.0000
1.5000
1.0000
0.5000
0.0000
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
AGE IN SINGLE YEARS



AGE RATIOS IN 1989 AND 1999 MALES
1989 males
1999 males
AGE RATIOS
AGE IN SINGLE YEARS



AGE RATIOS IN 1989 AND 1999 FEMALE POPULATION
1989 females
1999 females
AGE RATIOS
AGE IN SINGLE YEARS


AGE RATIOS FOR 1989 AND 1999 TOTAL POPULATION
1989 total
1999 total
AGE RATIOS
2.5
2
1.5
1
0.5
0
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
AGE IN SINGLE YEARS

4.4: COMPUTATION OF ESTIMATED MORTALITY LEVEL BY LINEAR INTERPOLATION

Estimates of childhood mortality by Trussell's method

Table 4.4.1: Probability of dying by age x North model

Age x	1989	1999
1	0.1130	0.1120
2	0.1030	0.1040
3	0.1020	0.0960
5	0.1240	0.1160
10	0.1450	0.1320
15	0.1700	0.1550
25	0.1890	0.1850

Source: CBS (2002) Vol V pg , 12

Table 4.4.2 : Estimating probability of dying at age x in 1989.

Age x	Combined Sexes (qx)	Female qx / 1.02	Males qx*1.02
1	0.113	0.1108	0.1153
2	0.103	0.1010	0.1051
3	0.102	0.1000	0.1040
5	0.124	0.1216	0.1265
10	0.145	0.1422	0.1479
15	0.17	0.1667	0.1734
25	0.189	0.1853	0.1928

It is important to note that to get the probability of dying for each sex the estimated q(x) for both sexes is divided by 1.02 to get q(x) for females and multiplied by 1.02 to get q(x) for males. The main reason is that the sex ratio at birth in Kenya is approximated at 102 males to 100 females.

In calculating the level for q(x) to estimate mortality level the following formula is used for the estimation of interpolated level.

Interpolated Level = Lower Level + (Calculated p(x) – Lower Level p(x) / (Upper Level p(x) – Lower Level p(x))

For example in table 3 at age x = 3 the interpolated level is given by;

Interpolated Level = 16 + (0.9000 – 0.8939) / (0.9087 – 0.8939) = 16.4127

Since the interpolated level lies between level 16 and 17, the next step is to calculate the interpolated probabilities of survival using the following formula:

Interpolated p(x) = Lower p (x) + (Upper p (x) – Lower p (x) * (Average Level – Lower Level) / (Upper Level – Lower Level).

Interpolated p (x) for age x = 5

= 0.9804 + ((0.9833 – 0.9804) * (16.9951 – 16) / (17 – 16)) = 0.9833

Table 4.4.3 : Levels for q (x) ,females ,1989

Age x	(qx) (1)	px=1-q(x) (2)	Lower (px) (3)	Lower Level (4)	Upper (px) (5)	High Level (6)	Interpolated Level (7)
1	0.1108	0.8892	0.8831	12	0.8945	13	12.5380
2	0.1010	0.8990	0.8925	15	0.9061	16	15.4823
3	0.1000	0.9000	0.8939	16	0.9087	17	16.4127
5	0.1216	0.8784	0.8771	16	0.8945	17	16.0765
10	0.1422	0.8578	0.8574	16	0.8779	17	16.0236
15	0.1667	0.8333	0.8227	15	0.8460	16	15.4566
25	0.1853	0.8147	0.8172	16	0.8423	17	15.8989

Average level for x = 2,3 and 5 = 15.9905

The lower and upper levels are obtained from the Coale and Demeny Model Life Tables, Manual X page 284 and 288 for females and males respectively.

Table 4.5.4 : Levels for q(x), males, 1989.

Age x	qx	px=1-q(x)	Lower (px)	Lower Level	Upper (px)	High Level	Interpolated Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.1187	0.8892	0.8759	13	0.8877	14	14.1264
2	0.1082	0.8990	0.8901	16	0.9047	17	16.6110
3	0.1071	0.9000	0.8771	16	0.8937	17	17.3805
5	0.1302	0.8784	0.8597	16	0.8786	17	16.9937
10	0.1523	0.8578	0.8381	16	0.8599	17	16.9038
15	0.1785	0.8333	0.8022	15	0.8263	16	16.2899
25	0.1985	0.8147	0.7894	16	0.8154	17	16.9737

Average Level for x = 2,3 and 5 = 16.9951

Since the interpolated level lies between level 16 and 17, the next step is to calculate the interpolated probabilities of survival using the following formula:

Interpolated p(x) = Lower p (x) + ((Upper p (x) – Lower p (x) * (Average Level – Lower Level) / (Upper Level – Lower Level)) .

For example the interpolated p (x) for age x = 5 is given by :

=0.9804 + (0.9833 – 0.9804) * (16.9951 – 16) / (17 – 16) = 0.9833

Table 4.4.5: Interpolated p (x), males, 1989.

Age	Lower Level 16 (1)	Upper Level 17 (2)	Average level (3)	Lower level (4)	High Level (5)	Survival Ratios (6)
0	0.9524	0.9603	16.9951	16	17	0.9603
5	0.9804	0.9833	16.9951	16	17	0.9833
10	0.9837	0.9856	16.9951	16	17	0.9856
15	0.9774	0.9798	16.9951	16	17	0.9797
20	0.9728	0.9756	16.9951	16	17	0.9756
25	0.9713	0.9743	16.9951	16	17	0.9743
30	0.9685	0.9717	16.9951	16	17	0.9717
35	0.9631	0.9669	16.9951	16	17	0.9668
40	0.9550	0.9594	16.9951	16	17	0.9593
45	0.9420	0.9468	16.9951	16	17	0.9468
50	0.9233	0.9291	16.9951	16	17	0.9291
55	0.8946	0.9016	16.9951	16	17	0.9016
60	0.8479	0.8565	16.9951	16	17	0.8565
65	0.7778	0.7888	16.9951	16	17	0.7887
70	0.6786	0.6921	16.9951	16	17	0.6920
75+	0.4226	0.4338	16.9951	16	17	0.4338

The lower and upper levels are obtained from the Coale and Demeny Model Life Tables, Manual X pages 284 and 288 for females and males respectively.

Table 4.4.6 : Survival ratios ,Females ,1989.

Age	16	17	Average Level	Lower level	High Level	Average Survival Ratios
	(1)	(2)	(3)	(4)	(5)	(6)
0	0.9570	0.9646	15.9905	16	17	0.9569
5	0.9821	0.9851	15.9905	16	17	0.9820
10	0.9857	0.9877	15.9905	16	17	0.9857
15	0.9829	0.9851	15.9905	16	17	0.9829
20	0.9797	0.9822	15.9905	16	17	0.9796
25	0.9766	0.9795	15.9905	16	17	0.9766
30	0.9732	0.9766	15.9905	16	17	0.9732
35	0.9687	0.9723	15.9905	16	17	0.9687
40	0.9637	0.9674	15.9905	16	17	0.9637
45	0.9552	0.9591	15.9905	16	17	0.9552
50	0.9405	0.9452	15.9905	16	17	0.9404
55	0.9155	0.9217	15.9905	16	17	0.9154
60	0.8723	0.8806	15.9905	16	17	0.8723
65	0.8055	0.8161	15.9905	16	17	0.8054
70	0.7113	0.7240	15.9905	16	17	0.7111
75+	0.4483	0.4598	15.9905	16	17	0.4482

Estimates of childhood Mortality by Trussell's Method, 1999.

Table 4.5: Probability of dying by age x North model.

Age x	1999
1	0.1120
2	0.1040
3	0.0960
5	0.1160
10	0.1320
15	0.1550
25	0.1850

Source CBS (2002)

Table 4.5.1: Estimating probability of dying at age x in 1999

Age x	combined sexes (qx)	Female qx/1.02	Males qx*1.02
	(1)	(2)	(3)
1	0.112	0.1098	0.1142
2	0.104	0.1020	0.1061
3	0.096	0.0941	0.0979
5	0.116	0.1137	0.1183
10	0.132	0.1294	0.1346
15	0.155	0.1520	0.1581
25	0.185	0.1814	0.1887

Table 4.5.2: Level for q (x) Females 1n 1999 census.

Age x	(qx)	px=1-q(x)	Lower (px)	Lower Level	Upper (px)	High Level	Interpolated Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.1098	0.8902	0.8831	12	0.8945	13	12.6236
2	0.1020	0.8980	0.8925	15	0.9061	16	15.4101
3	0.0941	0.9059	0.8939	16	0.9087	17	16.8093
5	0.1137	0.8863	0.8771	16	0.8945	17	16.5273
10	0.1294	0.8706	0.8574	16	0.8779	17	16.6450
15	0.1520	0.8480	0.8460	16	0.8680	17	16.0930
25	0.1814	0.8186	0.8172	16	0.8423	17	16.0554

Average Level for x = 2, 3 and 5 16.2489

Table 4.5.3: Survival ratios, Females, 1999

Age	Lower level 16	High Level 17	Average Level	Lower level	High Level	Survival Ratios
	(1)	(2)	(3)	(4)	(5)	(6)
0	0.9570	0.9646	16.2489	16	17	0.9589
5	0.9821	0.9851	16.2489	16	17	0.9828
10	0.9857	0.9877	16.2489	16	17	0.9862
15	0.9829	0.9851	16.2489	16	17	0.9834
20	0.9797	0.9822	16.2489	16	17	0.9803
25	0.9766	0.9795	16.2489	16	17	0.9773
30	0.9732	0.9766	16.2489	16	17	0.9740
35	0.9687	0.9723	16.2489	16	17	0.9696
40	0.9637	0.9674	16.2489	16	17	0.9646
45	0.9552	0.9591	16.2489	16	17	0.9562
50	0.9405	0.9452	16.2489	16	17	0.9417
55	0.9155	0.9217	16.2489	16	17	0.9170
60	0.8723	0.8806	16.2489	16	17	0.8744
65	0.8055	0.8161	16.2489	16	17	0.8081
70	0.7113	0.7240	16.2489	16	17	0.7144
75+	0.4483	0.4598	16.2489	16	17	0.4512

Table 4.5.4: Estimating probability of dying at age x, males, 1999

Age x	qx *1.02	px=1-q(x)	Lower (px)	Lower Level	Upper (px)	High Level	Interpolated Level
1	0.1142	0.8858	0.8759	13	0.8877	14	13.8343
2	0.1061	0.8939	0.8901	16	0.9047	17	16.2636
3	0.0979	0.9021	0.8937	17	0.9094	18	17.5354
5	0.1183	0.8817	0.8597	16	0.8786	17	17.1660
10	0.1346	0.8654	0.8599	17	0.8808	18	17.2603
15	0.1581	0.8419	0.8263	17	0.8495	18	17.6721
25	0.1887	0.8113	0.7894	16	0.8154	17	16.8426

Average Level for x = 2, 3 and 5 16.9883

Table 4.5.5: Computation of p(x)

Age	16 (1)	17 (2)	Average Level (3)	Lower Level (4)	High Level (5)	Survival Ratios (6)
0	0.9524	0.9603	16.9883	16	17	0.9602
5	0.9804	0.9833	16.9883	16	17	0.9832
10	0.9837	0.9856	16.9883	16	17	0.9856
15	0.9774	0.9798	16.9883	16	17	0.9797
20	0.9728	0.9756	16.9883	16	17	0.9756
25	0.9713	0.9743	16.9883	16	17	0.9742
30	0.9685	0.9717	16.9883	16	17	0.9717
35	0.9631	0.9669	16.9883	16	17	0.9668
40	0.9550	0.9594	16.9883	16	17	0.9593
45	0.9420	0.9468	16.9883	16	17	0.9468
50	0.9233	0.9291	16.9883	16	17	0.9290
55	0.8946	0.9016	16.9883	16	17	0.9016
60	0.8479	0.8565	16.9883	16	17	0.8564
65	0.7778	0.7888	16.9883	16	17	0.7887
70	0.6786	0.6921	16.9883	16	17	0.6919
75+	0.4226	0.4338	16.9883	16	17	0.4337

4.6: ESTIMATING MORTALITY LEVEL, ADULT MORTALITY

The same procedure as in child mortality is used such as to calculate lower and upper p (x), average level and interpolated level. Survival probabilities for adult males and females for hypothetical cohorts are used to give calculated p (x), CBS (2002) Vol, v, page, 26. Brass and Hill (1973) method was used in their formulation. For adult mortality, the conditional probabilities of survival are also available in Coale Demeny Model Life Tables. Using these model life tables along with the adult mortality estimates from ophanhood information it is possible to estimate mortality level for adult mortality.

Table 4.6.1 : Estimated mortality level for adults using ophanhood information, females for 1989.

Age n	25+n (1)	1(25+n) 1(25) (2)	Calculated px (3)	Lower Level (4)	Lower Level px (5)	Upper Level px (6)	Estimated Level (7)
5							
10	35	0.9718	0.9718	19	0.9701	0.9750	19.3469
15	40	0.953	0.9530	19	0.9522	0.9600	19.1071
20	45	0.9262	0.9262	18	0.9186	0.9299	18.6741
25	50	0.8929	0.8929	18	0.8898	0.9040	18.2170
30	55	0.8509	0.8509	17	0.8341	0.8516	17.9600
35	60	0.798	0.7980	17	0.7816	0.8026	17.7811
40	65	0.7218	0.7218	17	0.7076	0.7324	17.5724
45	70	0.6386	0.6386	18	0.6319	0.6605	18.2347
50	75	0.5091	0.5091	18	0.4962	0.5269	18.4206

Average Level = 17.7711

Table 4.6.2: Estimated mortality level for male adults using ophanhood information, 1999

Age n	25+n (1)	1(25+n) / 1(25) (2)	Calculated px (3)	Lower Level (4)	Lower Level px (5)	High Level px (6)	Estimated Level (7)
5							
10	35	0.9008	0.9008	8	0.8980	0.9065	8.3283
15	40	0.9011	0.9011	13	0.8977	0.9075	13.3462
20	45	0.8355	0.8355	11	0.8298	0.8443	11.3941
25	50	0.782	0.782	11	0.7808	0.7986	11.0652
30	55	0.7032	0.7032	10	0.7020	0.7236	10.0542
35	60	0.6388	0.6388	10	0.6278	0.6527	10.4418
40	65	0.5663	0.5663	11	0.5607	0.5880	11.2050
45	70	0.4807	0.4807	12	0.4733	0.4992	12.2849
50	75	0.3967	0.3967	14	0.3857	0.4112	14.4329
55	80						

Average Level = 10.567

Table 4.6.3: Estimated mortality level for male adults using ophanhood information, 1989

Age n	1(35+n) 1(35) (1)	Calculated px (2)	Lower px Level (3)	Lower Level px (4)	High Level px (5)	Estimated Level (6)
35						
40						
45	0.9183	0.9183	16	0.9135	0.9221	16.5603
50	0.9187	0.9187	20	0.9167	0.9281	20.1747
55	0.8582	0.8582	18	0.8424	0.8584	18.9906
60	0.7826	0.7826	18	0.7797	0.7999	18.1421
65	0.6883	0.6883	17	0.6703	0.6942	17.7526
70	0.5758	0.5758	17	0.5544	0.5810	17.8036
75	0.4441	0.4441	18	0.4391	0.4679	18.1739

Average Level for x = 2, 3 and 5 is = 17.8994

Table 4.6.4: Estimated mortality level for male adults using ophanhood information, 1999

Age n	1(35+n) 1(35) (1)	Calculated px (2)	Lower Level (3)	Lower level px (4)	Upper Level px (5)	Estimated Level (6)
35						
40						
45	0.9585	0.9585	21	0.9543	0.9618	21.5605
50	0.8103	0.8103	11	0.8013	0.8168	11.5813
55	0.7452	0.7452	12	0.7448	0.7621	12.0237
60	0.6455	0.6455	11	0.6361	0.6603	11.3884
65	0.5077	0.5077	10	0.5020	0.5293	10.2079
70	0.3952	0.3952	10	0.3767	0.4047	10.6613
75	0.2688	0.2688	11	0.2668	0.2937	11.0729

Average level for X = 2, 3 and 5 is = 10.6474

It is important to note that to get the final estimated mortality level for males and females separately the average levels in both 1989 and 1999 are summed up and the final average is obtained which becomes the estimated mortality level. In addition the method depends upon the availability of a suitably mortality schedule in order to carry out the projection. and a mistake in the specification of mortality will introduce a systematic bias in the results. Change in enumeration completeness may result in large distortions of the entire age distribution if this method is applied, though such large errors are likely to be obvious. Because of these potential problems, the estimates yielded by this procedure, though useful will have to be carefully scrutinized before being adopted for future use.

Reported and Smoothed Population Distribution by Age and Sex, (CBS) , 1999.

	Reported Population			Smoothed Population		
Age	Male	Female	Total	Male	Female	Total
0-4	2208096	2172389	4380485	2136286	2096681	4232967
5-9.	2007486	1969361	3976847	2079296	2045068	4124365
10-14.	2035268	1999012	4034280	1990530	1968929	3959459
15-19	1698144	1740076	3438221	1742882	1770159	3513041
20-24	1343079	1526901	2869980	1373528	1510815	2884343
25-29	1145531	1220897	2366428	1115082	1236982	2352065
30-34	854021	862143	1716164	869577	901875	1771453
35-39	706317	738302	1444619	690761	698570	1389331
40-44	525064	527325	1052389	532380	534628	1067008
45-49	426512	426963	853476	419197	419660	838857
50-54	349902	346761	696663	323772	326211	649983
55-59	226959	240797	467755	253089	261347	514435
60-64	197170	218681	415851	191552	212832	404384
65-69	142948	163058	306006	148566	168907	317472
70-74	120281	138028	258309	113680	128745	242425
75-79	80292	83064	163356	86893	92348	179241
80+	123522	221410	97889	123522	221411	221411
	14190592	14595168	28538718	14190593	14595168	28662240

The Arriaga technique was used to generate the smoothed population.

Source; CBS (2002) Pg 4

Enumerated and projected population, 1989

	Enumerated population			Projected population		
	Males	Females	Total	Males	Females	Total population
Age	1989	1989		1989	1989	
0	1911216	1888828	3800044	3562967	3532341	7095307
5	1743647	1725292	3468939	1590683	1578638	3169321
10	1504045	1485647	2989692	1211325	1238239	2449564
15	1177983	1200712	2378695	1016660	1046937	2063597
20	889594	1013340	1902934	847360	846162	1693521
25	782474	847289	1629763	662464	637676	1300140
30	583773	575651	1159424	524097	482644	1006741
35	460949	457942	918891	418358	414884	833242
40	367933	364245	732178	294897	328824	623721
45	281126	293405	574531	298116	288036	586152
50	235906	240618	476524	223560	235606	459166
55	179017	181155	360172	182375	185785	368160
60	150496	167902	318398	131314	139923	271237
65	113689	116979	230668	92039	98621	190661
70+	231774	245758	477532	46190	52826	99016
Total	10613622	10804763	21418385	11102405	11107142	22209547

Ratio of enumerated/adjusted	0.9643774	96.44
Under coverage rate		4.56

Enumerated and projected population, 1999

	Enumerated population			Projected Population		Total population
	Males	Females	Total	Males	Females	
Age	1999	1999		1999	1999	
0	2291936	2242966	4534902	3790525	3731396	7521921
5	2000580	1962556	3963136	2158181	2102552	4260733
10	2034980	2003655	4038635	1784369	1802329	3586698
15	1681984	1721194	3403178	1414373	1441847	2856221
20	1328529	1504389	2832918	1144332	1164306	2308638
25	1094909	1164594	2259503	892596	914908	1807505
30	840692	845230	1685922	738520	775115	1513636
35	695263	723749	1419012	586214	607719	1193934
40	516502	516989	1033491	450526	460366	910891
45	419841	418987	838828	365487	376547	742034
50	344639	340167	684806	275744	301324	577068
55	223691	236325	460016	210539	234871	445410
60	194513	214715	409228	174335	193072	367408
65	140969	160364	301333	148024	150565	298588
70	118601	135524	254125	35585	35561	71146
75+	174466	202658	377124	60492	59560	120052
	14102095	14394062	28496157	14229844	14352038	28581882

Ratio of enumerated/adjusted	0.997001	99.7
under coverage rate		0.3